University of Nottingham
UK | CHINA | MALAYSIA

Advanced Computational Modelling of Metal Organic Frameworks and their Performance

Isabel Cooley

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

6th September 2023

To my family: Phoenix, Althea, Doug, Fiona and Mathew.

Abstract

The performance of metal organic frameworks (MOFs) in applications relevant to modern technological challenges is assessed using selected computational modelling methods. This class of materials has gained significant attention in recent years thanks to its ability to display advanced properties, including high surface area and tunability. This work discusses MOFs and the computational techniques which may be used to obtain useful information about them in the context of modern advanced materials and methods.

There are several areas to which MOFs may be applied, and much of the focus of this work is on their use for gas storage and separation. The ability of selected MOFs to perform the difficult separation of xenon and krypton is examined by modelling uptake of the two gases using grand canonical Monte Carlo (GCMC) simulations. Similar simulations are applied in a high-throughput manner to identify MOFs which may be promising for gas separations important to upgrading biogas fuel streams, considering both total gas uptake and appropriate selectivity. The results of these simulations are used to train machine learning models which may be used to make efficient predictions of biogas upgrading ability. Gas uptake in MOFs may be affected by a number of specifics relating to structure and conditions; an important example of this, the effect of residual solvent on uptake, is assessed via a high-throughput GCMC study. The ability to reliably obtain high-quality images of MOF structures and visualise the processes they undergo is highly desirable. Transmission electron microscopy is a route to achieve this, but can be hampered by electron beam damage. Beam damage in selected 2D MOFs is modelled and analysed using ab initio molecular dynamics. Additionally, a classical many-body potential is used to model energetic favourability of metal cluster geometries, and the performance of the potential carefully assessed. A part of the rich context of advanced materials in which MOFs sit, metal clusters are another class of materials for which behaviour under imaging electron radiation is important.

Valuable conclusions may be drawn as a result of the computational modelling applied in this work. Several conclusions are discussed, including identification of MOFs which may be useful for relevant applications, identification of relationships between performance and other MOF properties, discussion of likely pathways for damage to materials, and discussion of the quality of different methods and models for particular applications.

Acknowledgements

There are several people whose presence, input and collaboration have helped to enable me to produce the work contained herein, and to whom I extend sincere thanks. First, my supervisor Professor Elena Besley, a constant source of teaching, scientific discussion and moral support, has been a great help to me. This is in addition to others who have acted in supervisory roles before and during my PhD studies, helping to build the foundation of scientific knowledge and research experience necessary to complete this work: Professor Tony Stace, Dr Richard Wheatley, Dr Josh Baptiste, Dr Jack Broad and Dr Joseph Glover.

While working alongside other researchers towards shared goals, I am grateful for the collaboration and discussion in various capacities of Dr Samuel Boobier, Professor Claude Degueldre, Louise Efford, Supriya Bhattacharyya, Professor Nick Besley, Dr Stephen Skowron, Rachel Heelas, James Pierce, Jonathan Lister, Ian Chester, Dr Tom Irons, Professor Ute Kaiser and members of her research group, Dr Haoyuan Qi, Dr Baokun Liang and David Mücke.

I further extend thanks to the remaining past and present members of the research group of Professor Elena Besley: Dr Abigail Miller, Edward Mortimer, Dr Ivan Debernev, Dr Katherine Wickham, Maggie Stankiewicz, Connor Williamson, Dr Aleksandra Foerster, Ilya Popov and Dr Sadegh Ghaderzadeh. As a group and along with all other members of the computational chemistry department at the University of Nottingham, they have created a stimulating and enjoyable research environment and participated in many valuable discussions.

I am grateful for compute time on the University of Nottingham High Performance Computing service Augusta. I am also grateful for use of the computational chemistry cluster. I thank those who maintain it, previously Dr Steven Oatley, currently Dr David Rogers, Dr Bang Huynh and Ross Amory.

I additionally acknowledge the online provision of a template used for this thesis document by Chao Zhang.

List of Publications

1. C. Degueldre, R. Dawson, **I. Cooley**, E. Besley, Fission Gas Released from Molten Salt Reactor Fuel: The Case of Noble Gas Short Life Radioisotopes for Radiopharmaceutical Application *Medicine in Novel Technology and Devices* 10, 100057 (2021) [Chapter 3]

2. **I. Cooley**, L. Efford, E. Besley, Computational Predictions for Effective Separation of Xenon/Krypton Gas Mixtures in the MFM Family of Metal Organic Frameworks *Journal of Physical Chemistry C* 126, 11475-11486 (2022) [Chapter 3]

3. **I. Cooley**, E. Besley, Do Solvent Molecules Always Hinder Gas Sorption in Metal Organic Frameworks? *Chemistry of Materials* (Under Review) [Chapter 5]

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Metal Organic Frameworks in the Setting of Advanced Functional Materials

Functional materials with relevant architecture and behaviour at small length scales constitute a diverse and rapidly growing group of chemical structures whose members have been and continue to be applied to a wide range of pertinent problems faced in the modern world. [24–27] The chemical landscape contains several materials defined by different kinds of atomic and molecular structures facilitating novel and interesting properties and functions. Many novel functional materials are nanomaterials, generally defined as materials with at least one external dimension in the range of 1 to 100 nm. [28] Among the range of materials encompassed by this description are carbon-based nanomaterials (fullerenes, graphene, carbon nanotubes and others), [28] other 2-dimensional materials of diverse kinds, [28] and metal clusters (eg. of platinum group metals) which are famously used in the field of catalysis. [26] The field of materials with advanced functions defined by their chemical structures also includes porous materials. In a materials chemistry context porous materials are those which, although not resticted to a size scale in external dimensions, possess pores on a small scale which influence molecular-level interactions. A material

is generally considered to be nanoporous, for example, if it possesses pores with diameter less than 100 nm. [28] These nanopores, along with pores on other small size scales (eg. micropores) can be instrumental in engendering useful and interesting material properties. Such materials include activated carbon, [29] zeolites, [30] organic cages, [31] covalent organic frameworks (COFs), [32] and metal organic frameworks (MOFs). [11, 24, 33] Advanced functional materials are the subject of extensive study both experimental and computational in nature.

Among the various classes of structures, MOFs have become something of a household name in the search for and design of promising functional materials. They are a class of material composed of self-assembled periodic networks whose building blocks are metal-containing nodes and organic linkers. Crystal structures with features that fit this modern description of MOFs were published as early as the 1950s, [34] though at this point they were neither referred to by the name metal organic frameworks nor considered as their own well-defined category of structure. A subset of the family of materials was more formally proposed in 1989 by Hoskins and Robson, [35] who wrote of "a new and potentially extensive class of solid polymeric materials" with "unprecedented and possibly useful" properties. This indeed proved to be true, and the 1990s saw a great deal [36–39] of research around MOFs, with particular interest in those displaying permanent porosity and the potential uses of this feature. [38] This included the first use of the term metal organic framework by Yaghi and Li in 1995. [36] Some of the earliest studied applications of MOFs were those involving their gas adsorption properties, and since then they have attracted much interest for gas storage and separation, [24] and have also been widely studied for functionality in fields including catalysis, [40] sensing [41] and drug delivery. [42]

It is the structural features and diversity of MOFs that allow them to be so readily turned to various applications. Thanks to their propensity for porosity, MOFs are well-known for possessing exceptionally high surface areas and

volumes, [33] which can be invaluable for applications involving adsorption of guest atoms. Perhaps their most attractive attribute is their high tunability: the vast quantity of node and linker building blocks in existence and the varied structures into which they can assemble lead MOFs to occupy a large and diverse chemical space. Many tens of thousands of synthesised MOFs have been published, with structural files for almost 100,000 of them deposited in the widely used Cambridge Structural Database (CSD) as of 2020 and the number ever increasing. [43] On top of this, hundreds of thousands of hypothetical MOFs (hMOFs) have been proposed. For example, the Northwestern Hypothetical Database [33] was published with almost 140,000 structures which were generated based on algorithmic combination of common MOF building blocks and has since been supplemented with more. From tunability springs opportunity, as MOFs may be selected or designed for desired applications with high specificity. However, this is reliant on knowledge of or ability to predict relevant properties of large numbers of structures. Computational work plays a large part in this endeavour, though is not infallible; close agreement between computational and experimental work is not always attainable, with computational models necessarily making approximations, including assuming pristine crystal structure and rigidity of frameworks, as well as neglecting certain interactions and processes. These are discussed in following sections.

1.1.1 Porous Materials for Gas Storage and Separation

In today's world we face several technological challenges to which porous materials may be turned as potential solutions. Examples which occupy a large part of the focus of this work are challenges relating to storage and separation of gases. In particular, there are many cases in which efficient gas storage and separation can be environmentally beneficial. Environmental impacts can be related to the gases used, for example in storage of H_2 as clean, high energy density fuel, [44, 45] capture of the greenhouse gas CO_2, [46, 47] or separation

of CH_4 and CO_2 to improve the performance of biogas fuel. [5, 48] They can also be related to the gas storage and separation techniques which are used. An example of this is the separation of xenon (Xe) from krypton (Kr) in the stream obtained from cryogenic distillation of air. The separation is necessary as both pure Xe and pure Kr are used in industrial and medical settings involving lighting, [49] anaesthetics, [50, 51] and diagnostics and therapy of lung conditions. [52] Neither gas is primarily utilised in the sustainability sector, but the current standard for the separation is the costly and energy-intensive cryogenic distillation. [50] A more efficient technique may reduce reliance on cryogenic distillation and thus display an environmental benefit. The situation is similar for the separation of acetylene from natural gas, for which intensive distillation methods are also standard and which are further complicated by the explosive nature of acetylene above a pressure of 2 atm. [53, 54] Finding robust alternatives for both the Xe/Kr and the acetylene/natural gas separations is challenging due to the similarity of the components.

Adsorption of gases by porous materials such as MOFs is a promising avenue for separation and storage. [24] By the affinity of a porous host for the desired gas, capture may in theory be achieved efficiently and at lower pressures than otherwise possible. The origin of this affinity can be complex and is related to chemical and physical interactions within the framework involving metal sites within the host or functional groups present on host ligands. Complementary geometry often plays a part, and size and shape of pores can be strongly linked to gas uptake properties. Similarly, increased affinity of host sites for certain gas molecules compared with others may allow separation, including of gases with similar properties and size which would be difficult to separate by other means. Affinity of a host structure for one gas over another leading to a difference in adsorption between the gases is referred to as selectivity.

Various techniques are available by which advantage can be taken of the selectivity of porous materials to achieve separation. Among them are the

widely-used pressure-swing and temperature-swing adsorption (PSA and TSA) techniques. [25] PSA involves exposure of the gas mixture to the adsorbent at high pressure, facilitating adsorption. Since adsorption is primarily of the component of the mixture displaying higher affinity for the adsorbent, the component displaying lower affinity may be collected. This is followed by a reduction in pressure, collection of the higher-affinity gas, and recovery of the adsorbent. The function of TSA is similar but relies on a swing of temperature rather than pressure to promote adsorption and desorption. Meanwhile a number of similar techniques such as pressure vacuum swing adsorption and electrical swing adsorption exist. [55] A well-performing material for use with a technique within this family will be selective of one gas over another as well as displaying high uptake at high pressure (or low temperature, or some other condition favouring sorption) and low uptake at low pressure (or high temperature, or some other condition favouring desorption). An alternative to techniques based on the swing of a condition, membrane separations involve a mixed stream of gases passing through a thin film of adsorbent. They may be an more sustainable approach to gas separations by porous materials. [31]

1.1.2 Low Dimensional Materials

The bulk properties of porous materials and in particular MOFs have now been discussed and are addressed further in several chapters of this work. They are not the only kind of properties which merit consideration. Functionality of materials can be heavily influenced by reduction in dimensionality to render a system no longer bulk. This can be reduction from three dimensional (3D) materials to two dimensional (2D) sheets, or further reduction to one dimensional (1D) poylmeric chains or nanotubes, or to molecules and clusters which are essentially zero dimensional (0D). The strictest definition of low-dimensional materials requires them to be atomically thick in any of their reduced dimensions (eg. a 2D material with one atomically thick di-

mension and two dimensions on a larger scale). This definition is restrictive, and functional definitions commonly allow materials which are at least a few layers thick. [56] It is common, for example, to consider a material to be an X dimensional material if its length is on a bulk scale in X dimensions and on the nanoscale in all other dimensions, [28] classing sheets composed of stacks of layers as 2D materials, nanotubes as 1D materials and nanoclusters as 0D, and classing any low-dimensional material as a nanomaterial by the definition given in section 1.1. Materials with zero, one and two dimensions can have interesting and advanced functionality. In all of these dimensions, structures may exist which resemble the conventional 3D MOFs, and 2D MOF materials can display interesting functionality encompassing both qualities of 3D MOFs and other qualities which are less commonly seen in three dimensions. [57]

The story of 2D materials begins with graphene. This single layer of the common allotrope of carbon graphite is composed of hexagonally arranged aromatic carbon atoms and displays high tensile strength and electrical conductivity. [58] The first ideas in the scientific community of the concept of graphene can be considered to predate the creation of the periodic table: a series of experiments on graphite and oxidised graphite by Brodie in 1859 included the observation of disintegration into thin plates and the lamellar structure of graphite which would ultimately facilitate the existence of its 2D form. [59] With the structure of the carbon allotrope further examined over the following century, [60] the properties of graphene were first explored theoretically in 1947 [61] in the context of furthering understanding of the electronic properties of graphite. It would still be some time before either the first synthesis of graphene or the first use in 1985 [62] of its now-common name. In 1962 Boehm and co-workers used reduction in alkaline suspension to create "extremely thin lamellae of carbon" reaching less than tens of Ångstrom (Å) in thickness and likely including atomically thin fragments, [63] but it was not until several years later that graphene was unambigously identified. The group of Novoselov and Geim were in 2004 able to fabricate free few-layer and

single-layer graphene sheets using the method of "mechanical exfoliation" [64] - famously, by peeling off layers with scotch tape. The resulting sheets were of sufficient robustness that devices could be made from them and electron transport properties observed. [64] This has popularly been considered to be the first true discovery of graphene [65] and therefore the first true isolation of a 2D material. Geim, however, argued [66] that many pioneers [63, 67–70] merit acknowledgement in the search for increasingly thin graphitic sheets.

The existence of graphene being thus cemented, and with it the possibility of creating materials composed of one or a few layers, other 2D materials have emerged. A natural analog to graphene is hexagonal boron nitride (hBN) in nanosheet form. [71] Alternating replacement of every carbon atom in graphite with its two neighbours in the periodic table, boron and nitrogen, creates a similar hexagonal structure which can exist in single or few-layer form just as graphene can. Nanosheets of hBN possess properties different from those of graphene, meaning they can fulfil unique functions; hBN possesses high strength and thermal stability, but is an insulator whereas graphene is a conductor. [71] Synthesis of hBN nanosheets can be challenging, but a number of routes exist, including chemical vapoour deposition and exfoliation and sonication from the bulk. [71] Other highly functional 2D materials include transition metal dichalcogenides (TMDs), which are promising candidates to catalyse H_2 production from water and again can be synthesised by exfoliation or by deposition. [72]

Similar to their 3D counterparts, 2D MOFs show great promise as functional materials. They are able to demonstrate porosity and may be used for gas sorption functions in the same way as 3D MOFs. [57, 73] Meanwhile, the reduction in dimensions, as for many materials, can lead to electrical conductivity. Combined with the tunability inherent to MOFs, this can facilitate a number of desirable functions including ion adsorption, redox catalysis and energy storage and transfer. [74] Synthesis of few-layer 2D MOFs can be challenging thanks to the high strength of interactions between layers in the bulk,

but they are accessible by both top-down methods such as sonication and exfoliation and by bottom-up methods involving direct synthesis. [75] Since MOFs tend to have complex structures, chemical vapour deposition is not generally a viable route to produce 2D MOFs, but they can be grown between layers of unlike materials in interfacial synthesis processes. [76, 77]

For functional materials in general, further reduction in dimensions to 1D and 0D structures can also be an interesting endeavour. Complex structures [78] can have interesting properties in low dimensions, as can structures made up of only one or a few elements. Nanoclusters of platinum group metals, for example, which are commonly obtained by growth under the influence of a high-energy electron beam, may display significant catalytic activity [22, 27] and can represent an efficient choice compared to bulk catalysis thanks to the greater surface area available per unit volume of the material. Understanding the growth process of clusters and the relative stability of different cluster geometries is key to maximising their activity, efficiency and recyclability. The logical extreme of the reduction in dimensionality is the single-atom metal catalyst, which represents a process of ideal efficiency in terms of surface area to volume ratio, but one that is difficult to achieve thanks to the lack of control over the positions and behaviour of single atoms or even small clusters. Incorporation of catalytic single-metal sites into MOFs via open metal sites at MOF nodes is an interesting approach which can immobilise catalytic sites and afford greater control. [79] Furthering this aim requires understanding of the behaviour of MOFs and the behaviour of metal clusters, as well as the way that the two interact.

1.2 Computational Approaches to Studying Properties of Advanced Materials

Uncovering the properties and functionality of MOFs and related porous materials continues to be an active field of research, and the constant search for materials better suited to particular tasks is aided by frequent introduction of new techniques and technology. While experimental work is unquestionably vital in the study of properties and structures of porous materials, computational work is invaluable in furthering understanding as well as aiding and directing experimental efforts. Indeed, knowledge of MOF behaviour on a scale large enough to take full advantage of the high tunability is unobtainable from a purely experimental perspective. That is, it would not be possible to experimentally measure relevant properties of the thousands of possible MOF structures to select the strucutre best suited for a given application. Computational modelling methods can be used to obtain information about a range of properties, though are limited by the approximations they must employ. As a basic example from a computational standpoint, structural properties of materials and particularly MOFs such as pore size, surface area and volume are very important to their behaviour. Structural properties within the approximation of a perfect rigid framework are readily accessible computationally using methods such as Widom insertion [80] or geometrical techniques based on Voronoi networks [81] (see section 2.1.1).

Computational methods are particularly useful in measuring behaviour and interactions of a material. Considering gas sorption as a highly pertinent example of a MOF application, to experimentally measure the uptake of only a single pure gas at a single pressure for every known MOF is not feasible. Instead, computational simulations allow predictions to be made about the performance of a system at greater speed, with a lower resource cost and less safety risk than experiments. It is possible to predict gas uptake in a material, and therefore to obtain computational adsorption isotherms, us-

ing grand canonical Monte Carlo (GCMC) simulations. [82] Such simulations are described in section 2.3 and involve use of several computational energy calculations to obtain a statistical picture of a system. The ability of these simulations and their results to accurately describe experimental systems depends on the method selected. For materials modelling, energy calculations are typically done using classical force fields due to the prohibitive cost of ab initio methods (see chapter 2). The accuracy and transferability of force fields is not always sufficient to fully reproduce experimental data, but they can often offer a reasonable picture of MOF behaviour, giving general guidance in cases where quantitative agreement is not accessible. Large-scale methods of this kind have been used in this work to make predictions of the ability of MOFs to separate xenon from krypton (chapter 3), and to separate CH_4 from CO_2 under various circumstances (chapters 4 and 5). Meanwhile, energy calculations may be carried out instead by ab initio methods, increasing the chance of high accuracy. This, however, greatly increases the cost, and it is often necessary when using ab initio methods to study representative clusters of systems rather than the bulk. [83] Meanwhile, ab initio methods themselves are not infallible, and their accuracy and cost depend on the specific choice of method.

When using force field-based Monte Carlo approaches to calculate gas uptake, it is necessary for appropriate force field parameters to be applied (see section 2.1.2). General force field parameters which encompass the entire or much of the periodic table are commonly employed, particularly for high-throughput screening. These may offer useful approximations, but it is by their nature impossible to tailor them to every situation. In particular, the interaction between guest gases and open metal sites is often underestimated by general force fields. In certain cases it is appropriate to employ adjusted force field parameters for a particular kind of interaction. For example, Fischer et al developed parameters for the interaction between open copper sites and acetylene, designed for the MOF $Cu_3(btc)_2$. [84] Similarly, Haldoupis et al

developed parameters for the interaction of a number of metal sites with CO_2. [85] The sets of adjusted parameters were able to accurately describe adsorption in certain systems but no adjusted parameters are likely to translate well to all possible systems.

Use of low-cost classical methods combined with the ever-increasing computational power available to researchers is particularly useful in allowing computational studies of material properties, including MOF properties, to be conducted in a high-throughput manner. The performance of many hundreds, thousands or tens of thousands of structures can be assessed, and those displaying the most promising behaviour carried forwards for further study, including experimental analysis. High-throughput methods to study MOF properties are employed in parts of this work (chapters 4 and 5).

Making predictions of the performance of structures may also be aided by the establishment of structure-property relationships; features such as surface area, volume and pore size relate closely to properties and functions of porous materials, so can be useful in making predictions of performance. These features may be obtained experimentally, but may also be calculated readily using computational methods, as mentioned. [81] Once structural features are known, structure-property relationships may be considered case-by-case for individual structures, [1, 86] or established statistically on a larger scale. [5] Extending this latter possibility, machine learning methods may be used to establish models which predict properties and behaviour of materials using structural information, which is cheaply obtained. This approach has been taken for a range of materials property searches, [87] including searches for gas uptake properties in MOFs [88, 89] In some of these gas uptake studies, it has been observed that in addition to relationships between performance and *structural* information, relationships between performance and *chemical* information can be usefully established and utilised for prediction of properties, and can lead to improvement compared to models purely based on structure. In this work, the high throughput techniques applied to the CH_4/CO_2 separa-

tion are combined with machine learning methods to train models to cheaply predict separation properties under a range of circumstances (chapter 4).

1.2.1 Consistency with Experimental Studies and Issues Relating to Structure

The importance of making predictions about the properties and behaviour of materials, and in particular porous materials, has been established, and it is clearly necessary for any method used to replicate or predict experimental results to a useful degree of accuracy. Useful accuracy is certainly accessible in many situations, and computational work has proved instructive in identifying experimentally useful MOFs or structural features. [10, 90, 91] However, often quantitative and sometimes qualitative agreement between force field methods of accessible cost and experiment is elusive. Even when using the best available predictive force field methods there are notable cases of discrepancy, particularly when not using adjustments for accuracy which lack generality and so are limited in high throughput settings. MOFs predicted computationally to be high-performing can be experimentally observed to be less useful by some relevant metric; for example, the adsorption of methane and noble gases in manganese formate was predicted to be several factors larger than experimentally observed, and to follow qualitatively different isotherms. [92] Even considering a case in which computational predictions successfully identified a high-performing MOF, the identification of SBMOF-1 as an adsorbent for the Xe/Kr separation, the experimental selectivity of the MOF was still significantly lower than the computational prediction. [10] Meanwhile, computationally promising MOFs can also be structurally unstable. Several members of an experimental MOF database were found to collapse or experience significant structural change under DFT optimisation in a study by Nazarian et al, [93] and studies of hypothetical MOFs are limited by the question of whether promising candidates are synthesisable. [94] Similarly, it is possible

for high-throughput studies, while successfully identifying a selection of high-performing MOFs, to miss other equally promising structures which are less well-described by simulations.

One way in which computational force field methods can be limited is in overestimating the strength of attractive dispersion interactions, while another is in not accounting for the increased strength of guest interactions with open metal sites compared to with fully coordinated metal centres. [83, 85] Meanwhile, a third is in not accounting for the flexibility of MOF frameworks and the effect that it may have on gas uptake, [76] and the situation is further complicated by the dependence of experimental results on several factors, including the quality of the samples used. [95]

It has been mentioned that force field methods are generally selected as a compromise between accuracy and cost, and that density functional theory can be used to obtain higher accuracy, but is tempered by higher cost. It must, however, be noted that in addition to being limited by cost, DFT results are accompanied by their own limitations in terms of accuracy, and are highly dependent on the choice of functional, basis set and other specifics of the method, as is also discussed in section 2.2.1. For example, DFT methods based on the local density approximation (LDA) and the generalised gradient approximation (GGA) do not account well for the long-range attractive van der Waals interactions which are important to MOF/gas interactions. This can be partially addressed by the use of hybrid methods which combine DFT functionals with elements of other methods, or by use of empirical dispersion corrections to DFT calculations. [83] Local LDA and GGA functionals further have shortcomings in predicting electronic properties of materials. They give poor qualitative descriptions of open shell systems which are strongly correlated, [96] and systematically underestimate band gaps by significant margins, [97–99] a shortcoming related to the inability of the methods to account for derivative discontinuity as electrons are added or removed from a system. Again, this can be addressed by semiemprical methods employing a band gap correc-

tion, [99] or by use of more sophisticated hybrid functionals which can account for derivative discontinuity. [98] In all cases where shortcomings of DFT are addressed by use of hybrid functionals, this leads to an increase in computational cost, engendering a trade-off between cost and accuracy. Furthermore, DFT yields results whose quality depends on the basis set used within the method. In general (though not without exception), a larger basis set is likely to yield more accurate results, but comes at a higher cost, engendering further cost/accuracy trade-off. Cost/accuracy trade-off can additionally be sensitive to other aspects of the method, such as the size of the grid used for the calculation. [100] As a result of the potential for poor agreement between results of computational methods of various kinds and experiments, care must be taken when interpreting computational predictions, and care must similarly be taken when running computational simulations to ensure real systems are described as accurately as possible.

In order to make predictions of MOF behaviour by most computational methods, atomic coordinates must be known. For real structures which have been experimentally synthesised, this information is commonly obtained in crystal structure determination X-ray diffraction (XRD) processes [25, 74, 101]. As discussed, there is also power in generating coordinates for hypothetical structures which have not yet been synthesised experimentally. Coordinates of moieties of real structures can be used during algorithmic generation of hypothetical structures, in which fragments are combined according to some set of bonding rules. [33] This ultimately still relies on experimental crystal structure determination, which carries a number of uncertainties. These include difficulty in resolving hydrogen atoms [102] and uncertainty about exact atomic position. The former can result in H atoms being assigned unclear positions or left out of structural files entirely, and the latter can cause crystal structure files to be generated featuring overlapping atoms and under- or over-coordinated atoms. [102]

Further, there may be uncertainty regarding the presence of unaccounted-

for phase changes, structural collapse, missing linkers, and contaminants or defects in real structures. This includes synthesis solvent remaining within pores. With very small pores and precise interactions often determining observed behaviour, it is reasonable to expect that the presence or absence of synthesis solvent may be decisive. Several experimental procedures using MOFs habitually involve solvent removal activation steps, [103] but crystal structure determination tends to occur post-synthesis when solvent is still present. As a result, published crystal structures often include synthesis solvent. From a computational perspective, removal of solvent from structures is seen as a routine part of structure preparation to bring coordinates in line with those used in experimental application studies. There exist computational algorithms for addressing uncertainties in crystal structures and removing solvent, such as a publicly available algorithm associated with the Cambridge Structural Database (CSD), [21] and an algorithm used in generation of a Computation Ready Experimental (CoRE) MOF database which differentiates between free solvent and bound solvent. [11] It is not, however, guaranteed that the results of computational and experimental solvent removal procedures will be the same.

From an experimental perspective, one cannot always be certain that no residual solvent remains in the pores of a given structure: not all structure evacuation methods result in complete activation or are suitable for all cases. [103, 104] Furthermore, experimental solvent removal procedures may lead to partial or complete framework collapse, [103] which further affects behaviour. The results of solvent removal procedures can differ between different experimental studies, and between experimental and computational studies. If solvent is present it may not be detectable [85] and its effect on behaviour, though not always significant, can be strong, as has been shown by initial computational studies on gas uptake behaviour [11, 105] and individual experimental observations. [104, 106, 107] There is something of a paucity of literature in this area, with further investigation needed. In particular, consideration of

solvent at times assumes that effects on gas uptake, where present, will be negative, although a few studies have indicated that retaining solvent in pores of materials may be beneficial to desired processes. [11, 107] With that in mind, the importance of the presence or absence of solvent to gas separation properties of a range of MOFs with varying structures is assessed in this work (chapter 5).

Since published structures are not always stable with respect to solvent removal by computational algorithms (or indeed by experimental activation [103]), use of structures whose solvent has been algorithmically removed does not always constitute study of viable materials. Efforts have been made to computationally determine the solvent removal stability of MOFs, such as the combination of data mining and machine learning employed by Nandy et al. [108] Even if stable on solvent removal, structures can undergo phase changes, [106] and efforts have also been made to optimise computational MOF structures, such as the Density Functional Theory (DFT) structural study of Nazarian et al. [93] This study additionally compared properties of MOFs before and after structural optimisation and observed the largest differences in cases where solvent had been removed from the pores. However, it did not compare the properties of the structures before and after solvent removal.

Aside from discrepancy with experimental activation, algorithmic cleaning of structure files is neither infallible nor universal from a purely computational perspective. It has been shown that different solvent removal algorithms working on equivalent MOF structures can lead to different stripped structures that behave differently. [19] Digital solvent removal can erroneously strip charge-balancing ions or coordinated moieties (even including metal centres [19]). One unfortunate result of this is the existence of structural information files requiring atoms to occupy inaccessible oxidation states, which may also result from other experimental structural determination issues already mentioned. [102, 109] There is indication that large proportions of popular published databases may be affected by this issue, and this is supported by evidence

seen in later parts of this work (see chapters 4 and 5). The issue extends to hypothetical databases, as their constituent structures are generated from fragments of real MOFs. Issues of this kind are gaining attention as a limitation which must be addressed, and there have been attempts to automatically screen databases for problematic structure files in the hope of paving the way to cleaner databases in the future, [102, 109] while improved automation of structure cleaning is additionally to be desired.

Meanwhile, synthesisability remains an issue specific to hypothetical MOF databases. Not yet having been synthesised, a given hypothetical MOF does not come with established synthesis routes and there is no guarantee that a viable synthesis for it can be found, even if it is not affected by the previously mentioned viability issues common to real and hypothetical MOFs. This poses a problem when promising hypothetical MOFs are identified by computational simulations: the work of designing a synthesis route for such a structure remains substantial and may not be fruitful at all. [94, 110] In addition to the potential for fruitless efforts as structures which are studied further prove to be dead ends, this creates a situation in which hMOFs predicted to be promising but with unusual or unfamiliar synthesis routes are overlooked in favour of less promising MOFs which appear more amenable to synthesis. [94] Efforts to combat these issues take the form of methods to predict synthesisability of hMOFs using properties determinable from their computationally generated structures. For example, Anderson et al [94] used free energies calculated using molecular dynamics to assess synthetic likelihood, and and Park et al [110] integrated consideration of synthesisability into design protocols for MOFs with H_2 storage capacity by comparing binding energies of proposed MOFs to those of their polymorphs. Meanwhile, Luo et al [111] developed data mining and machine learning protocols to predict synthesis routes for MOF structures using structural fingerprints as input. As the ability to make predictions regarding synthesis of a hypothetical structure is highly desirable, work of this nature is important and ongoing. In general, synthesisability predictions may

give guidance about the probability of a structure being experimentally accessible, but they cannot cover all relevant factors. [94] In spite of the issues and drawbacks which characterise them, however, the very existence of both real and hypothetical databases such as the CSD MOF subset [21] and the Northwestern Hypotheticals [33] is a valuable factor in facilitating high throughput studies. Their power is such that they remain in wide use in their untreated form while the issues of problematic structures are addressed.

A further area of disconnect between computational studies of porous materials and experimental phenomena, as mentioned, is the case of flexibility of framework materials. This can affect the accuracy of computational adsorption studies, the majority of which employ the approximation of host structures as rigid frameworks which are able to interact with guest molecules but do not themselves move. In many cases, such an approximation is appropriate. However, a number of porous materials, in particular MOFs, display appreciable flexibility in response to stimuli [76, 112] which may affect their uptake properties to varying extents. This can include flexibility specific to the bonding environment of a given structure, such as breathing, swelling, and linker rotation. [76] Flexibility can be induced by heat, and in principle this applies to all structures if studied at a high enough temperature. Temperature induced flexibility is more significant for some structures than others and can sometimes involve well-defined structural changes. [76] For notably flexible MOFs, computational predictions which employ the rigid approximation clearly do not offer a full description. For example, gate opening processes which allow adsorbents to enter pores only after a system experiences a certain stimulus can lead to high selectivity. This selectivity originates from the different threshold stimuli a MOF may have for admitting molecules of different adsorbates and is seen, for example, in the selective adsorption of alkanes by ZIF-7, which experiences gate opening at a lower pressure of ethane than of ethylene. [113] It cannot be observed using the rigid approximation.

There are ways in which flexibility can be addressed in computational sim-

ulations, although they are limited, as usual, by cost and accuracy. Most obviously, the framework may be modelled as flexible instead of rigid. However, this is not straightforward, as it greatly expands the number and complexity of terms required for a force field (see section 2.1.2). A force field appropriate to the system must be used; if one does not already exist cumbersome parameterisation is necessary, and this is an area in which machine learning may be employed to efficiently produce accurate and transferable potentials. [114] Flexible simulations have been used for select MOFs, [115] and other materials [31] but it is a large undertaking. Perhaps the most useful way to employ these flexible simulations is in hybrid Monte Carlo/molecular dynamics simulations: short molecular dynamics trajectories which account for flexibility are included among the trial moves used in a Monte Carlo procedure (see section 2.3 for details on both the Monte Carlo and Molecular dynamics protocols). These hybrid methods have been used, for example, to simulate uptake of a range of noble gases, methane and carbon dioxide, in the flexible MOF MIL-53(Al), [116] though the cost of such methods would be prohibitive in a high-throughput workflow. As an alternative to building flexibility into simulations, if multiple phases of a flexible framework are known, uptake simulations may be employed for more than one phase and the resulting isotherms combined by taking relevant pressure ranges from each. This will not give a full description of the flexible system, but may be sufficient for some purposes.

1.3 Transmission Electron Microscopy for Structure Determination and Imaging

The importance of methods for structural elucidation is clear, and the established XRD methods which are often used for crystal structure determination are highly valuable. Alongside and in addition to these, methods more readily applicable to smaller crystals and non-crystalline samples, methods which

can readily image defects in samples, and methods which can image dynamic processes, are highly desirable. A promising imaging method is transmission electron microscopy (TEM). The process is similar to visible light microscopy, but involves irradiation of a sample with a high energy electron beam (e beam) rather than visible light, followed by transmission of the beam which depends on characteristics of the sample. The transmitted beam is detected and the information it carries is used to directly generate an image of the sample.

There have been several developments in imaging methods in recent years that have allowed high-resolution TEM (HRTEM) to image at the sub-Å level under the correct circumstances. Perhaps most notably, aberration correction addresses limitations inherent to all lenses and pushes resolution close to the diffraction limit. [117, 118] Meanwhile, optimisation of both electron sources and detectors can enhance information transfer. [117] Further advanced methods can further improve resolution. For example, the method of ptychography can be used in imaging, particulrly of very thin films, [119] to push beyond the diffraction limit. It uses interference patterns between scattered electron beams to obtain phase information and constitutes an important advancement in the field of imaging methods. [118]

Electron microscopy methods have several applications relating to materials imaging and to manipulation of material behaviour and properties. TEM methods can be used in determining geometries of metal clusters and study of their deposition on surfaces, aiding in characterisation of catalytic materials and their distributions on solid supports, [120] as well as in improving their catalytic activity. In this context, TEM imaging of metal nanoparticles can aid in understanding of mechanisms by which undesirable nanoparticle growth occurs. [121] The electron beam used in TEM can also be used to initiate and promote growth of metallic clusters, [122] as well as to manipulate their behaviour and dynamics. [123] In this work, modelling is applied to clusters in chapter 7.

TEM approaches are additionally beginning to be accessible for atomic

scale imaging of porous materials such as MOFs. With the use of TEM, it is possible not only to directly elucidate structures of MOFs including defects not captured by crystallographic methods, but also to image in real time phase transitions, [124] reactions, [125] and processes occurring within material pores. Additionally, it has been shown that the electron beam can act to promote chemical processes at the same time as visualising them. [125, 126] This paints a picture of a versatile and highly useful imaging tool, but in practice the application of TEM imaging to MOF structures is hindered by beam-induced damage to the integrity of the imaged framework. The usefulness of TEM to image any material depends on the structural integrity of the given material with respect to beam irradiation: if the material experiences sufficient damage to amorphise before an image is obtained, the crystalline structure cannot be elucidated. With several sources of susceptibility to damage, MOFs tend to display a low resilience to beam damage, which is enough to prevent them from being imaged in many cases. [127] This is discussed further in chapter 6, in which ab initio dynamics modelling is used to study beam damage.

1.3.1 Reduction of Beam Damage

With beam damage presenting a significant roadblock to widespread use of electron microscopy methods such as TEM in MOF imaging, there have been focused recent efforts to reduce damage under imaging conditions. Methods to reduce beam damage in MOFs are related to those used to reduce beam damage in other materials with high damage susceptibility and strong imaging motivation, such as biological materials. [128] Useful approaches can include reduction in temperature or in beam energy, which both effectively lower the energy which irradiated atoms possess and thereby reduce the chance that an atom will leave its position in the sample. [127, 129] These can reduce certain portions of damage, but the effects are dependent on the nature of the irradiated sample and reduction in beam energy comes at the expense of

resolution.

A highly significant contribution to damage reduction, which has allowed atomic-scale imaging to be achieved for some MOFs, is reduction in the electron dose applied to the sample. [129, 130] For a given beam energy, a particular material will have associated with it some critical total dose that may be delivered to a sample before amorphisation occurs. If imaging is able to proceed using some number of electrons below this critical dose, the sample will not have amorphised and its structure can be properly imaged. Use of low electron doses has been made possible by development of electron detection cameras with heightened sensitivity, and by development of methods to reduce the amount of time a sample needs to spend under a beam before imaging may proceed. [129] It can be highly successful but is not widely available thanks to the cost associated with the necessary equipment. It is useful therefore to increase understanding of beam damage processes, which can indicate which methods can be most prudently applied to image different MOF materials. This can give guidance about whether a given sample is likely to require use of high-cost and highly advanced imaging methods, as well as direct efforts towards development and synthesis of MOFs which have greater beam resilience where possible.

Computational modelling work has an important part to play in aiding beam resilience development. Modelling beam damage processes using dynamics simulations can help to reveal by which mechanisms beam damage occurs, and therefore can prompt efforts to avoid these mechanisms. This has, for example, been usefully applied in identifying the loss of hydrogen as a primary damage mechanism in organic molecules. [131] Application of similar methods could identify dominant damage mechanisms in MOFs, as well as predict whether a structure is likely to undergo significant damage under given conditions. Mechanisms of beam damage for a selection of 2D MOFs are probed in this work in chapter 6

There are limitations to the computational dynamics methods used in

beam damage modelling. Computational beam damage modelling makes use of ab initio molecular dynamics within the Born-Oppenheimer approximation (see chapter 2). [132] Using ab initio methods alongside molecular dynamics is costly and limits the timescale and size of system that can be studied, which can be particularly relevant for periodic materials like MOFs. Less expensive methods, however, are often insufficient. Meanwhile, it has been mentioned that MOFs and other materials experience different types of damage. This includes damage induced by transfer of energy to atoms, but also includes damage induced by transfer of energy to electrons. This second type cannot be readily modelled in the Born-Oppenheimer approximation. Methods exist to include electron dynamics in models which can be applied to beam damage modelling. [133] Such methods are more expensive still, meaning studies of beam damage do not always capture all damage mechanisms. Meanwhile, many DFT methods are additionally limited in their ability to model electronic properties of MOFs, as discussed in section 1.2.1. This is also discussed in chapter 6, which addresses mechanisms of damage and methods to model 6.

1.4 Overview of This Work

The remainder of this work is set out as follows. Chapter 2 gives an overview of relevant computational methods which can be used in modelling properties of MOFs and of materials more generally, with a focus on methods applied in subsequent chapters of this work. It covers methods used to determine structural and geometrical properties of materials, methods used in energetic calculations, which are divided into force field methods and ab initio methods, large scale methods used to study systems on a statistical scale, and machine learning methods which can be used in conjunction with other computational techniques.

Chapters 3, 4 and 5 discuss the setup and results of studies in which the gas uptake properties of MOFs have been computationally modelled. Various

questions relating to gas uptake are addressed in these chapters, and the methods applied in each are selected accordingly. In chapter 3, the ability to perform the important but challenging separation of xenon and krypton is measured for a small family of MOFs via GCMC simulations and force field methods in the context of computationally calculated structural information. In chapter 4, GCMC simulations and force field methods are again utilised, this time to measure biogas upgrading properties in MOFs as described by the CO_2/CH_4 separation. A large and carefully curated database of some thousands of MOFs is used, and the GCMC simulations are applied on a high throughput scale. A selection of machine learning models is trained on the resulting curated GCMC data with a view to improving efficiency of future similar situations. The performance of the models is analysed in detail. Chapter 5 addresses questions relating to the effect of residual solvent on gas sorption properties of MOFs via a high-throughput GCMC and force field screening of curated structures in which uptake of CO_2 and CH_4 is examined. Each structure included in the screening is simulated in a solvated and a desolvated form, and the difference in uptake between the two is measured under different circumstances. In addition, ab intitio calculations are applied to small model systems to further probe the effects that solvent may have on energetic interactions.

The focus of chapter 6 is using modelling to further understand the processes that occur during TEM imaging under a high-voltage electron beam. The chapter addresses the structural damage that an electron beam can cause to hinder imaging processes and the susceptibility of MOFs to these processes. Ab initio molecular dynamics is used to model electron beam damage events relevant to a series of MOFs to probe structural and chemical causes of damage.

Electron beam imaging is also relevant to chapter 7, whose focus is modelling of small metal clusters which may exist in an electron beam imaging and growth environment. Furthering understanding of metal clusters can help to widen the understanding of metal behaviour and set metal centres of MOFs, which are vital to their properties, in context, while metal clusters may also be

used in conjunction with MOFs for certain applications. In the chapter, the performance of a many-body force field to model metal clusters is examined and the potential is used to probe geometry and lability in platinum clusters.

Chapter 8 gives an overview of the questions addressed in this work, relevant findings, and future directions of the areas studied.

Chapter 2

Methodological Overview

A wide-ranging selection of computational methods has been used over the course of this work. Methods span classical force-field based approaches, ab initio methods using Density Functional Theory (DFT), and the use of machine learning (ML) to appropriately predict the results of more conventional computational methods at a reduced cost. They have been applied both on a small scale and in large-scale statistical simulations. Methods have been selected for the individual problems addressed, with classical force field approaches often favoured because of the necessity for statistical scale simulations of large systems when dealing with gas uptake in MOFs. Without attempting to present full mathematical explanations of all methods, a general introduction to each of the methods used is given in this chapter. Specific details regarding individual methods and how they have been applied are given in subsequent chapters to which they are relevant.

2.1 Classical Methods

Classical simulation methods are those which do not directly make use of quantum chemical theories. Typically classical methods are force-field based, using mathematical models which have been fit to reproduce either experimental or ab initio data. They are applied in this work as useful approximations

to predict geometrical and gas uptake material properties. They are particularly useful for MOF modelling and are certainly indispensable in any high-throughput study of MOFs, as statistical simulations of the complex structures are prohibitively expensive.

2.1.1 Geometrical Methods

Knowledge of the geometrical properties of materials is necessary for a full understanding of their behaviour. Relevant geometrical properties of porous materials relate to the size and shape of their porous networks, by which they are partially defined. This is certainly true of MOFs: strong predictive relationships have been observed between the geometry of MOFs and their gas uptake properties. From a known crystal structure, a number of important structural properties are accessible via computational methods which are available in published software packages. [81, 82] These include pore size, internal volume, and surface area, which are illustrated in Figure 2.1, using a schematic representation of a porous structure. The framework walls are shown in black and voids are represented by the white regions. Details of the geometrical properties are described below.

For a MOF material which possesses a complex pore network, its pore sizes may be defined by different measurements, in particular largest cavity diameter (LCD), pore limiting diameter (PLD) and largest free pore diameter (LFPD). They are represented by the diameters of the blue circles in Figure 2.1. The LCD is the diameter of the largest pore within the material; the circle labelled LCD in Figure 2.1 is located inside the largest pore shown, but in this case the pore is blocked from the remainder of the network by a part of the framework wall. The PLD is the largest diameter that can percolate through the structure; in Figure 2.1 the PLD is the diameter of the channel walls. The LFPD is the largest diameter located along the path which contains the PLD. These properties, in particular PLD and LCD, are highly relevant to MOF ap-

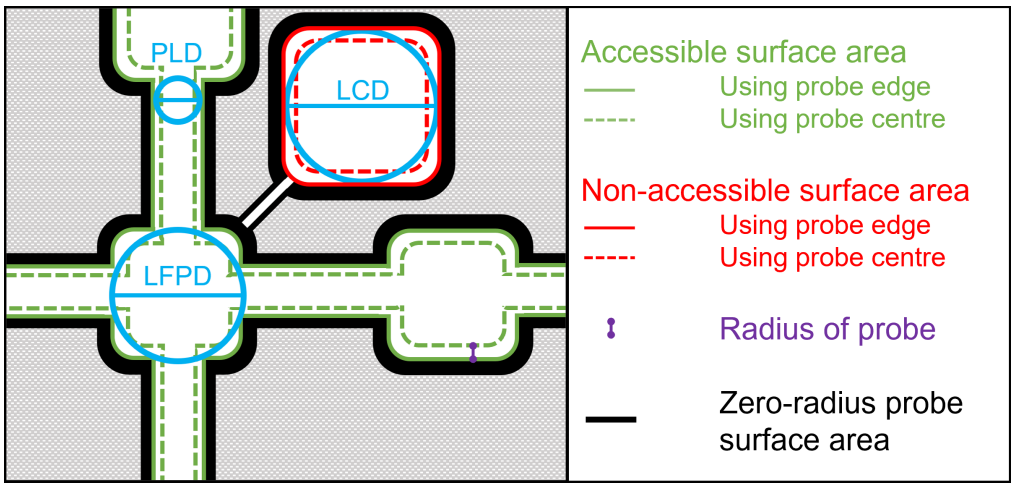

Figure 2.1: *Schematic representation of relevant geometrical features in MOFs. Pore sizes are shown by blue circles and geometrical surface areas are shown using lines. Accessible surface area is shown with green solid and dashed lines and non-accessible surface area is shown with red solid and dashed lines. Volume corresponds to the white region enclosed by either the dashed or the solid lines. Dashed lines relate to area and volume calculated using the centre of a probe of radius greater than zero, and solid lines relate to area and volume calculated using the edge of a probe or a zero radius probe.*

plications. PLD is particularly relevant to diffusivity and transport situations such as membrane separations in which motion of guests may be limited by the smallest gaps, while both PLD and LCD have relevance for separation and storage uptake applications where transport properties are not the focus. LCD is the most direct description of pore size; it describes the size of the dominant pores. Therefore it gives information about the space available and the likelihood of close geometrical fits of guest molecules within those pores. PLD, meanwhile, can also be relevant to uptake: requiring close interaction with pore walls, adsorption often takes place in the smallest available pores which fit the guest molecules, maximising interactions. [89] Therefore, in structures possessing a large LCD but other pores or channels with smaller diameters (e.g. MOFs composed of both micropores and mesopores), PLD is likely to be particularly relevant to adsorption, particularly at low pressures. The ratio of the two, which describes pore morphology, can also be relevant to gas uptake.

Internal surface area (SA) and volume (V) are relevant to a number of applications of MOFs and other porous materials. They can be split into ac-

cessible surface area and volume (ASA and AV) and non-accessible surface area and volume (NASA and NAV). ASA and AV are the surface area and volume accessible to a probe molecule of a specified size, while NASA and NAV are area and volume large enough to contain the probe but not accessible to it through the pore system. This is illustrated in Figure 2.1, in which ASA is shown using green lines and NASA is shown using red lines. AV and NAV are the volume enclosed within the relevant lines. Two sets of lines enclosing slightly different areas and volumes are shown, one solid and one dashed: which line is relevant depends on the kind of probe and algorithm used in the measurements (see below). Any area or volume not large enough to contain the probe at all is not included in the measures. Experimentally reported geometrical properties are determined using adsorption of a guest gas. [134] They are necessarily ASA and AV, and the probe is the guest gas molecule, commonly N_2 or He. Computational methods can determine NASA and NAV as well as ASA and AV; in general it is appropriate to compare ASA and AV, rather than their non-accessible counterparts, to experimental values. However, there are cases in which experimental factors such as flexibility, which was discussed in section 1.2.1, render area or volume experimentally accessible that is determined to be non-accessible by computational algorithms assuming pristine rigid structures. It is therefore useful to also make computational NASA and NAV values available for comparison to experiment where necessary. For computational area and volume calculations, the size of the probe must be selected, and affects the area and volume designated accessible. It it prudent to select a probe which reflects either the gas molecules used in experimental geometry determination or the gas molecules relevant to the study at hand. Distinction can be made between volumetric and gravimetric surface area and volume. Volumetric measurements are defined per volume of the unit cell, while gravimetric measurements are defined per mass of the unit cell. Both give related but different relevant information about a structure.

Computational values for pore sizes are commonly obtained using purely

geometrical methods, while computational volume and surface area can be determined either using force field methods (see section 2.1.2 for more details on force field methods) or by purely geometrical methods. A useful mathematical representation of the pore space is needed for geometrical calculations. In this work, geometrical calculations make use of methods based on Voronoi networks as implemented in the Zeo++ software package. [81] The Voronoi network is a graph representation composed of planes which bisect the space between pairs of atoms. Edges of the network are areas where planes overlap, and are therefore equidistant to three different atoms. Nodes are points where edges meet, and are therefore equidistant to four atoms. Each atom in a Voronoi representation is thereby enclosed in a polyhedron whose faces lie directly between that atom and one of its neighbours. [81] For each edge and node of the Voronoi network, the distance to the nearest atom is readily available, which is important for later geometry calculations. The Voronoi network can take into account atomic radius as defined by a researcher. [81]

Following construction of the Voronoi network and assignment of relevant distances, it can be used in determination of pore sizes. The LCD is given by twice the largest distance from any Voronoi node to its nearest atom. PLD and LFPD are calculated by a combination of this logic and Dijkstra's [135] lowest cost path algorithm. [81] The algorithm traverses the structure from one Voronoi node to the next, returning to the periodic image of the original node, with the nodes representing the largest pores assigned the lowest cost. The PLD is twice the smallest distance to an atom along the lowest-cost path traversed, and the LFPD is twice the largest distance to an atom along the same path. [81]

For purely geometrical calculation of surface area and volume, random Monte Carlo sampling is used. [81] The sampling is either of points on spheres representing each atom of the structure and the probe (surface area) or of points throughout the unit cell (volume). Each point is designated as viable if its position is not overlapping with any (other) atoms of the structure, or

unviable if its position does overlap. Surface area and volume are then calculated based on the fraction of points which were assigned as viable. If the radii of the spheres used to represent each atom of the structure are the van der Waals radii of those atoms and a zero radius probe is used, then the geometrical surface area is equal to the van der Waals area of that structure. Accessible and non-accessible surface area and volume can be separated by an algorithm which determines the accessibility of the Voronoi nodes associated with the sampled points. A similar method was also introduced for surface area calculations by Düren et al. [136]

For geometrical sampling methods to calculate both surface area and volume, the exact region that is sampled must be established. When SA and NASA are calculated by geometrical algorithms, it is common to consider the area which can be occupied by the centre of the relevant probe. This is represented by the dashed lines in Figure 2.1, where the distance between the dashed line and the framework wall is the radius of the probe. It was argued by Düren et al [136] that this is preferable to the surface area accessible to the edge of the probe, shown by the solid red and green lines in Figure 2.1, as it gives a better representation of the adsorption capacity.

When AV and NAV are calculated by computational geometrical algorithms, there is scope for judgement on exactly which parts of the internal volume to include. The simple geometrical volume can be considered to be the total amount of space not occupied by framework atoms, which is equivalent to the amount of space accessible to a zero Å probe. The volume occupiable by the centre of a non-zero sized probe has also commonly been used. Similar to surface area, this is the volume enclosed by the red and green dashed lines in Figure 2.1. Finally, the volume occupiable by the whole of a non-zero sized probe is represented by the solid red and green lines in Figure 2.1. Methods to compute this final type of volume were presented by Ongari et al, [137] who termed it accessible or non-accessible probe-occupiable volume. They also compared calculations this probe-occupiable voulme to calculations of simple

geometric volume and volume accessible to the centre of a probe. [137] It was found that the third definition of volume, accessible probe-occupiable volume, was likely to match experimental volume more closely than the other two methods, while volume accessible to the centre of the probe was likely to be a significant underestimation of experimental volume.

As previously described, volume may alternatively be calculated using force-field methods and Monte Carlo sampling, which are implemented in the RASPA software package. [82] The force fields and Monte Carlo methods relevant to this approach are discussed in more detail in section 2.1.2, and addressed only briefly here. The technique used is Widom particle insertion, in which a probe particle is randomly inserted at positions within the system and the energy at each insertion is calculated using a selected force field model. Volume is calculated based on integration of the exponential of the calculated energy with respect to position. [82] This method takes into account real interactions of the probe particle with the framework in addition to geometry, but it is dependent on temperature and selection of force field parameters. The assessment of methods for calculating volume by Ongari et al [137] concluded that for these reasons, as well as a comparatively poor ability to reproduce experimental volumes, the Widom inserstion method was not preferable to the probe-occupiable method.

Alternative methods for calculating computational surface area follow similar methods to those used for experimental surface area determination. That is, they involve determination of a computational isotherm using force field and statistical sampling methods (see sections 2.1.2 and 2.3) and fitting the initial linear portion of this isotherm to a model, commonly the Brunauer, Emmett and Teller (BET), [138] model in order to determine monolayer loading of a material and hence its surface area. Use of the BET method to determine a computational surface area is significantly more computationally expensive than geometrical methods, as it requires determination of a full adsorption isotherm prior to BET calculations. The values calculated by BET are also

highly dependent on the specifics of its application, as an appropriate pressure range which defines the linear portion of the isotherm must be selected, [136, 139] although this may be countered by automatic BET methods. [139] It also depends on the assumption that the adsorption mechanism for the material in question fits the BET model. [136] One way in which this may not be the case is that the material may experience pore filling contamination at the pressure selected to represent monolayer loading, something which is particularly likely to occur in cases where both micropores and mesopores are present, with a second layer of adsorption occurring in the micropores before the mesopore monolayer is complete. This can introduce error, although computational methods exist which are able to correct for pore filling contamination. [18]

The computational methods outlined in this section can be invaluable parts of investigations into the performance of MOFs, as well as of related porous materials. In addition to the predictive relationships previously mentioned, which may be vital in high-throughput studies into gas uptake properties, a computational structural analysis can give information about experimental samples and their quality. [136] It is not always known for sure when an experimental MOF is reported whether regions of its structure may be partially collapsed, its pores may retain residual synthesis solvent, or whether it has undergone a phase change. Computational calculations are done on a pristine crystal structure of the reported material. Any significant deviation between computationally calculated and experimentally obtained structural properties may be used to determine the likelihood of issues with the experimental sample that cause its structure to deviate from the pristine coordinates.

2.1.2 Classical Force Fields

For the majority of computational calculations relating to real-world systems, it is necessary to consider interaction energy between two or more particles. Classical force fields are a way to represent the interatomic (or inter-pseudo-

atomic) potential energy of a system by using some mathematical model of the relevant part of the potential energy surface comprised of parameters which have been fitted to experimental or ab initio data. Once fit, the force field described by these parameters can be used in computations of potential energy. The computational cost associated with using fitted force fields is far lower than that of using ab initio methods directly. This renders classical force fields a useful way to probe large systems or timescales which are inaccessible to more expensive methods. However, when using a classical force field, high accuracy is more elusive, the applicability of the force field to chemical environments not similar to those for which it has been fit is limited, and the fitting itself is a substantial task.

When dealing with a classical force field, it is necessary to establish the composition of the systems to which it can be applied. The vast majority of chemical systems contain multiple different chemical bodies which interact with each other. From the perspective of a real system, the bodies are atoms (composed themselves of subatomic particles) which may combine to form compound chemical entities (e.g. molecules). From the perspective of a force field, interacting bodies are commonly atoms, but force fields may also treat molecules or groups of atoms as one body (pseudo-atom).

A full description of a system comprised of more than two bodies (a many-body system) involves consideration of interactions between every individual body and every other body within the system as a collective. That is, it contains terms which depend on the positions of every combination of bodies in the system. A 2-body potential considers only the sum of interactions between each chemical body and individual other chemical bodies (within a specified cutoff radius). That is, it is composed of a sum of terms each of which depends only on the positions of two atoms. While a full many-body treatment is in general neither practical nor necessary, the 2-body treatment is not always sufficient. For example, it makes no consideration of bond angles which are instrumental to a description of molecules, and cannot correctly describe a

number of properties of atomic solids. [140] A force field which goes beyond the 2-body description may be described as a many-body force field, and those designed to treat atomic solids fit this description. One example is the family of embedded atom potentials, which include a 2-body term as well as a second term in which interaction of an atom with the rest of the electronic environment is embedded. [141–143] A second example is the Murrell-Mottram 2 + 3 body potential, which contains a 2-body term and a 3-body term. [6, 140, 144] Potentials of this kind are dealt with in more detail in Chapter 7.

Molecular systems are often treated using a combination of Taylor expansions over different contributions to the total potential energy, each truncated at a selected order. For example, one common component of a molecular potential is the bond stretching contribution, which can be expressed as a sum of Taylor expansions over deviation from the equilibrium distance between atoms I and J. Other examples of components of a molecular potential include bends, torsions, out-of-plane bends and non-bonded interactions. A number of these terms inherently include many-body contribution as they depend on the coordinates of more than two bodies, and terms depending on larger numbers of coordinates, such as bond-torsion terms, may also be included. Total energy can be expressed as a sum of individual Taylor expansions truncated at some specified order. [145] An example sum is displayed in equation 2.1. [145] A force field of this kind may be used to address, for example, flexibility in MOFs, but is not used in full in this work, thanks to the rigid approximation. In equation 2.1, each U is a Taylor expansion expressing the potential energy of a selected contribution. Within the sums, r is an interatomic distance, θ is a bend angle, ϕ is a torsion angle, and χ is an out-of-plane bending angle. The label $o.o.p$ means out-of-plane, $n.b.$ means non-bonded, and the subscripts following U are used to refer to the type of interaction described by the potential. For example, U_s refers to a stretching potential and U_{nb} refers to a non-bonded potential. The sums in each term are over all relevant groups of bodies. For example, the sum in the stretch term is over all bonded pairs of atoms and the

sum in the bend term is over all groups of three atoms connected by bonds.

$$U = \sum_{stretch} U_s(r) + \sum_{bend} U_\theta(\theta) + \sum_{torsions} U_\phi(\phi) + \sum_{\substack{o.o.p.\\ bend}} U_\chi(\chi)$$
$$+ \sum_{n.b.} U_{nb}(r) + \sum_{\substack{stretch-\\ stretch}} u_{ss'}(r, r') + \sum_{\substack{stretch-\\ bend}} u_{s\theta}(r, \theta) + \sum_{\substack{bend-\\ bend}} u_{\theta\theta'}(\theta, \theta')$$
$$+ \sum_{\substack{stretch-\\ torsion}} u_{s\phi}(r, \phi, r') + \sum_{\substack{bend-\\ torsion}} u_{\theta\phi}(\theta, \phi, \theta') \quad (2.1)$$

Within each component of equation 2.1 is a Taylor or Fourier expansion composed of a number of tunable parameters. The stretch terms may again be taken as an example. Considering only the stretch between two atoms I and J (one term of the stretch summation in equation 2.1), the stretch Taylor expansion can be given by equation 2.2, in which r refers to $r^{IJ} - r_e^{IJ}$, the difference between the distance IJ and the equilibrium distance IJ (that is, the distance at which the interaction energy is a minimum).

$$U_s(r) = \frac{1}{2}[k_2^{IJ} r^2 + k_3^{IJ} r^3 + k_4^{IJ} r^4 + ...] \quad (2.2)$$

For practical use, Taylor expansions are truncated after a selected number of terms. The coefficients k_i are parameters which must be tuned enable the force field to most accurately reproduce experimental or ab initio data for a reference system. To be useful, the parameters of a force field must also be able to produce high-quality data for systems other than the reference system to which they were fit. It is desirable for a force field to have high general applicability and therefore transferability, but transferability to systems which are chemically different from the reference is a challenging aim. Note that, in the case of the stretching potential, truncation after the first term yields the harmonic oscillator approximation, and k_2 is the harmonic force constant, commonly referred to as simply k. [100] This, in fact, is the origin of the term force field, as it refers to a mathematical model that has historically dealt in

force constants. [145] Also note that the Taylor expansion for the stretching potential in equation 2.2 does not display the correct limiting behaviour at large r, but it models stretching well in the region around the equilibrium bond length. Force fields which take the general form discussed in equations 2.1 and 2.2 are not perfect models.

Equation 2.1 takes account of many different kinds of interactions between atoms within molecules, as well as non-bonded interactions. It can also be instructive, and far less computationally demanding, to model a system using only non-bonded interactions. This approximation is commonly used, for example, in adsorption studies. [53, 84, 146] In the context of adsorption of small molecules within a porous framework (e.g. a MOF), the interactions within the system can be described as guest-guest interactions (between guest molecules), host-guest interactions (between guest molecules and the host framework) and host-host interactions (within the host framework). Guest-guest and host-guest interactions are more impactful on several relevant aspects of behaviour of the system than the bonded host-host interactions. Functionally, neglecting these host-host interactions is done hand in hand with an approximation of the framework as rigid. Consideration is of only interactions between chemical entities and requires only the non-bonded term, which can itself be considered to be composed of two terms, an electrostatic and a van der Waals term. Thus, equation 2.1 is significantly simplified to equation 2.3, where the subscript e refers to an electrostatic potential and the subscript vdW refers to a van der Waals potential.

$$U = \sum_{non-bonded} U_{nb}(r) = \sum_{non-bonded} U_e(r) + \sum_{non-bonded} U_{vdW}(r) \qquad (2.3)$$

The simplification of the non-bonded term to the form given in equation 2.3, it must be noted, is an approximation which neglects the direct contribution of polarisation to non-bonded interactions. The electrostatic part treats atoms as point charges, while the van der Waals part accounts for dispersion

interactions. However, neither account for the possibility of polarisation and induced dipoles, which may be relevant to MOF-gas interactions, particularly the interaction between open metal sites with large partial charges and polarisable guest atoms. [147] Inclusion of polarisation in non-bonded force fields is desirable for increased accuracy. Enhanced interactions between open metal sites and guest atoms can be addressed by adjustment of parameters in non-polarisable force fields, [84] but explicit consideration of polarisation has better physical justification and transferability. Polarisable force fields are currently limited by computational cost overheads involved in standard methods of calculating dipoles [148, 149] and a relative lack of widespread development, but there have been efforts to develop efficient and available implementations of polarisation. [147, 148]

The van der Waals Term

Of the two remaining terms in equation 2.3, the van der Waals term is determined by fitting to some model of inter-entity distance against potential energy. The model must display appropriate behaviour in the attractive and repulsive regions, with a steep repulsive barrier of positive interaction energy at short distances, followed by an attractive well of negative interaction energy and asymptotic behaviour tending to zero at large distances where the particles no longer interact. A commonly employed mathematical form of a curve which displays this approximate behaviour at a reasonable computational costis the 12-6 Lennard-Jones potential, given in equation 2.4.

$$U_{vdW}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{2.4}$$

Here, r is the distance between two particles and ϵ and σ are the tunable parameters. σ represents the separation at which the van der Waals potential between the two particles is zero. This can be considered to describe the size of the particle; in a hard-sphere model the particles would be touching

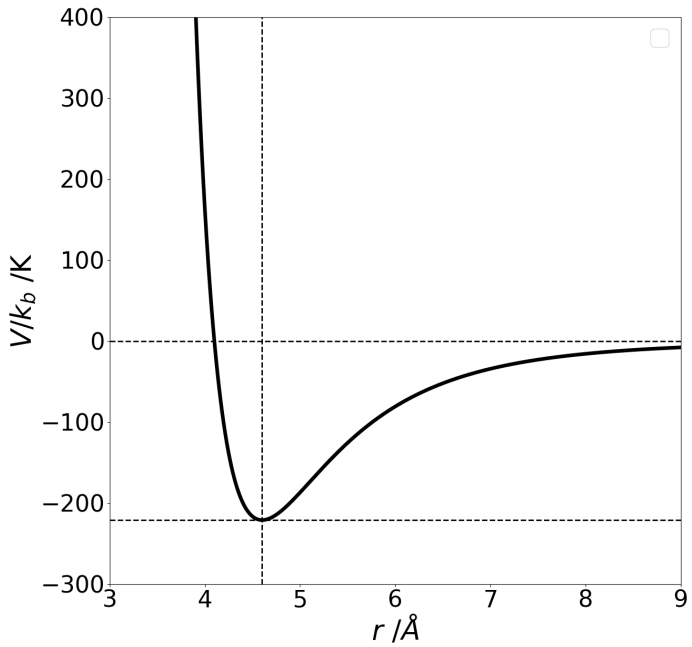

Figure 2.2: *An example of the Lennard-Jones functional form used to represent non-bonded van der Waals interactions between two particles. The plot uses the parameters* $\sigma = 4.10$ *Å and* $\epsilon/k_b = 22.10$ *K. The dashed lines are* $V/k_b = 0$, $V/k_b = \epsilon/k_b$ *and* $r = 2^{1/6}\sigma$

at this point. Meanwhile ϵ represents the well-depth, the minimum potential energy between the two particles. [145] At this point, the particles are at their equilibrium separation and the distance between them is $2^{1/6}\sigma$. To illustrate the features of the Lennard-Jones potential, its appearance is shown in Figure 2.2, with V/k_b used in place of V for convenience. This figure uses parameters for gaseous Xe taken from Hirschfelder [2] as an example. These are $\sigma = 4.10$ Å and $\epsilon/k_b = 221.0$ K. Relevant positions are indicated by horizontal and vertical dashed lines. A horizontal line of $V/k_b = 0$ shows the long-range asymptotic behaviour and crosses the curve where $r = \sigma$. A horizontal line at $V/k_b = \epsilon$ and a vertical line at $r = 2^{1/6}\sigma$ cross at the bottom of the potential well, the equilibrium distance at which the interaction energy is ϵ.

A Lennard Jones potential of this form, or indeed some other mathematical representation of non-bonded interactions, may be used to calculate van der Waals energy between pairs of chemical entities once relevant parameters are determined. Van der Waals parameters are generally determined for a given atom type. Therefore, for each pair IJ, if the two chemical entities are the

same, the values of σ_I and ϵ_I are equal to σ_J and ϵ_J, so determining U_{IJ}^{vdW} is a simple matter of substituting the relevant parameters into equation 2.4. If I and J are not the same, the Lennard-Jones parameters must be combined, which can be done using Lorentz-Berthelot mixing rules. [145] These rules are given in equations 2.5 and 2.6.

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \tag{2.5}$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{2.6}$$

The Lennard-Jones parameters thus calculated for an atom pair can be used in equation 2.4 to calculate the van der Waals energy for that pair. The success of this method relies on the fitted parameters accurately describing the situation to which they are applied. This is not always the case. If there is an interaction for which the parameters included in the force field do not provide an accurate description, then it is possible to use separate adjusted parameters for that interaction, which requires further fitting. [84, 85] In such a case the Lorentz-Berthelot mixing rules would not apply.

For large or periodic systems, it is not feasible to calculate all interactions between all atom pairs. Instead, van der Waals energy is calculated for all pairs within a certain van der Waals cutoff distance using periodic boundary conditions where appropriate. The distance must be specified for the simulation in question. Beyond this distance all $U_{vdW}(r)$ are considered to be zero. As interatomic distance increases, the actual potential approaches zero, so the approximation inherent in using a cutoff becomes less severe. At the cutoff point, the potential can either be truncated, leaving a point of discontinuity, or the entire curve can be shifted up by a small value so that it meets the x-axis at this point. [150] Researchers must judge the appropriate cutoff distance to use, considering both accuracy and computational time. A notable consideration when making this judgement is the number of framework unit

cells which must as a result be included in the calculation. Where periodic boundary conditions are used, interaction of an atom with its own periodic image must be avoided, meaning the perpendicular width of the framework cell must be more than twice the van der Waals cutoff. [145] If the unit cell does not fulfil this condition it must be duplicated to create a larger cell before boundary conditions are applied. This precludes the use of extremely large cutoffs which render very large unit cells necessary. In any case, use of very large cutoffs is redundant, resulting in calculation of very long-range van der Waals interactions which are small in magnitude and contribute little to the system. Appropriate cutoffs can be determined using benchmarking: for some system, a relevant property can be calculated using successively larger cutoffs. Very small cutoffs can be expected to give inaccurate results, while larger cutoffs converge towards some infinite-cutoff limit. A researcher can select a cutoff based on a desired degree of convergence and the observed computational cost of each test calculation.

To use a model like the Lennard-Jones potential, a researcher requires defined parameters for every atom type in a system. Fortunately, several parameterised force fields of the Lennard-Jones form exist. For example, the universal force field (UFF) [4] was fitted with van der Waals parameters for every atom in the periodic table, with some separate parameters for chemical environment, and the Dreiding force field [151] is almost as diverse. Thanks to their diversity, these force fields are commonly used to describe framework atoms in gas uptake simulations of MOFs, and are highly useful for high-throughput studies in which a wide range of atoms may be present, but they are far from universally transferable. Other parameters are required for treating atoms or molecules as guest gases in the same simulations. For example, specific gas-phase parameters [2, 3] fitted for the noble gases more accurately model their uptake than UFF or Dreiding parameters (see chapter 3). Meanwhile, the Transferable Potentials for Phase Equilibria (TraPPE) [152, 153] sets of parameters describe several organic and similar small gas molecules.

For alkanes, they make use of united-atom models in which atoms (for example a carbon and its associated hydrogen atoms) are grouped and represented with one set of parameters. For example, the TraPPE model for methane treats the molecule as a single pseudo-atom. These descriptions of guest gases combined with the UFF or Dreiding force fields remain general, and specific interactions between gas and host are not always well-described. This is sometimes addressed by fitting parameters for a specific host-guest interaction, as done by Fischer for the Cu-acetylene interaction. [84]

The Electrostatic Term

Some, but not all, descriptions of systems require consideration of partial charges of atoms and electrostatic interactions in addition to the van der Waals treatment already discussed. When modelling gas uptake in MOFs using a rigid approximation for the host structure, the necessity of this consideration depends on the nature of the guest gas molecules and the model used to represent them. With no host-host interactions considered, electrostatic terms are relevant if they describe host-guest and guest-guest interactions. A guest which is described with no partial charges such as a neutral noble gas atom does not need to be described by an electrostatic model. Similarly, the atoms of a CH_4 molecule will possess only very small partial charges, to the extent that in the TraPPE formalism it is represented as a single neutral united atom, in which it is impossible to separate charge into partial contributions. In these cases, although host frameworks could be described with partial charges, it is not necessary. Atoms of other molecules, however, have significant partial charge nature which can impact the way they interact with framework atoms. For example, the CO_2 molecule is represented by a three-site model in the TraPPE formalism, with each atom assigned a partial charge. For this and similar molecules, it is necessary to compute framework partial charges and their electrostatic interaction with the guest: the electrostatic term in equation 2.3 is non-zero.

Many methods are available for assigning point charges to atoms within molecules, including Mulliken [154] and Bader [155] partitioning or the Hirshfeld [156] and Iterated Stockholder Atom (ISA) [157] methods. The above mentioned methods involve partitioning of molecular electron density calculated using an ab initio method over the atoms of a molecule, and are not readily suited to the large periodic systems of MOFs, nor to the efficiency required from a high-throughput situation. Methods have been created which are designed to calculate charges for periodic systems, such as the density derived electrostatic and chemical (DDEC) [158] and Repeating Electrostatic Potential Extracted Atomic (REPEAT) [159] method. Both, however, still rely on high-cost ab initio calculations.

The computationally efficient charge equilibration (Qeq) method [160] is commonly used for assigning atomic partial charges within MOFs. It represents a compromise between accuracy and speed, providing a reasonable description of charges at a low computational cost. The method works by expressing the energy of an atom in terms of its charge as a Taylor expansion centred around a charge of zero and truncated at the second order. The energy can be expressed in terms of electronegativity (the derivative of energy with respect to charge) and hardness (the derivative of electronegativity with respect to charge), which come from known isolated atom ionisation potential and electron affinity, and a damped Coulomb potential term representing pairwise interaction. Energy is then minimised with respect to atomic charges to find the equilibrated partial charges of each atom. There have been several modifications to the Qeq method [161] For example, Wilmer et al [162] introduced the extended charge equilibration (eQeq) method which allows metals to be treated with a Taylor expansion centred around a formal oxidation state (as specified by the user) rather than a charge of zero. The choice of oxidation state has been seen to impact partial charges and the resulting MOF-gas simulations. [5] The eQeq method is available in the RASPA software package [82] and is widely used. It is not infallible, and even with specified oxidation states

can give unphysical charges. Care must be taken by a researcher to detect any unphysical charges following a calculation.

Once partial charges are assigned for all relevant components, Coulomb's Law [163] is used to determine the electrostatic term. The overall electrostatic energy is a sum over all electrostatic interactions between atoms I and J within a certain distance, as shown in equation 2.7, [145] where r_I and q_I denote, respectively, the position and charge of atom I, and ϵ_0 is permittivity of free space.

$$\sum_{non-bonded} U_e(r) = \frac{1}{4\pi\epsilon_0} \sum_{I<J} \frac{q_I q_J}{|r_I - r_J|} \tag{2.7}$$

Treatment of long-range electrostatic interactions must be considered. For a periodic system, the r^{-1} sum in equation 2.7 decays slowly and does not converge. Considering pairwise interactions of all atoms within a cutoff, the number of interactions calculated depends on the square of the cutoff. Combining this with the reciprocal in equation 2.7 leads to divergence. [100]. Any direct summation would require use of very large cutoffs to give meaningful results. [145, 164] Instead, Ewald summation can be used for periodic systems. It splits the divergent sum into a pair of convergent sums, a short range sum in real space and a long-range sum in Fourier space. This allows rapid, accurate convergence and thereby treatment of long-range interactions. [164]

2.2 Ab Initio Methods

Classical and empirical methods as discussed in section 2.1 are valuable for modelling large chemical systems and long time periods, both of which have computational cost as a limiting factor. However, in computational chemistry it is generally desirable where possible to determine properties of systems from first principles, though the computational cost of such an endeavour is much higher than for classical methods. That is, it is desirable to use mathematical

methods to find solutions to the Schrödinger equation (SE) as it applies to the system in question without input of experimental or fitted parameters. The SE is given in the time-independent form in equation 2.8. [165]

$$\hat{H}\Psi = E\Psi \tag{2.8}$$

In this equation, Ψ is a wavefunction, some function of the coordinates of a chemical system which is an acceptable eigenfunction solution to the SE. The product of the wavefunction with its complex conjugate describes probability density, and the function contains all information about the system. [100] Use of operators on the wavefunction may give rise to calculation of observable properties. In the SE, $\hat{H}$ is the Hamiltonian, an operator that operates on the wavefunction to give the eigenvalue E, the energy of the system. The SE is an eigenequation in which E is the expectation value of the Hamiltonian operator for the eigenfunction Ψ. The form of the Hamiltonian and its level of complexity depend on the system studied. A common pedagogical example, the 'particle in a box' - a particle moving in one dimension trapped between impenetrable (infinite potential energy) barriers - is described by the Hamiltonian given in equation 2.9, where x is the 1D positional coordinate of the particle.

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} \tag{2.9}$$

Moving to equations applicable to real chemical problems, a system of electrons and nuclei in the absence of external fields is described by the Hamiltonian given in equation 2.10. [100] In equation 2.10, atomic units are used, I and J are indices referring to the nuclei of the system, i and j refer to the electrons, M_I is the mass of nucleus I, Z_I is the charge of nucleus I, R_I and r_i are the positional coordinates of nucleus I and electron i respectively, and the Laplacians ∇_I^2 and ∇_i^2 are the second derivatives of positional coordinates of nucleus I and electron i in all dimensions. The first term of equation 2.10

describes nuclear kinetic energy, the second term describes nuclear-nuclear repulsion, the third describes electron kinetic energy, the fourth electron-nuclear attraction and the fifth electron-electron repulsion.

$$\hat{H} = -\sum_{I} \frac{1}{2M_I}\nabla_I^2 + \sum_{I}\sum_{J>I} \frac{Z_I Z_J}{|R_I - R_J|} - \sum_{i} \frac{1}{2}\nabla_i^2 - \sum_{i}\sum_{I} \frac{Z_I}{r_i - R_I} + \sum_{i}\sum_{j>i} \frac{1}{|r_i - r_j|} \quad (2.10)$$

For many chemical problems the Born-Oppenheimer (BO) [166] approximation is applied to equation 2.10. Since nuclei possess much greater mass than electrons, electron motion is much faster than nuclear motion and electrons can be treated as responding adiabatically to nuclear behaviour. Therefore, an approximation which treats the part of the Hamiltonian pertaining to nuclear motion as separate from the rest can be useful, and problems are treated using an altered form of the Hamiltonian, which is similar to that given in equation 2.10 but omits the nuclear kinetic energy term, and is instead composed of the nuclear-nuclear repulsion term and all the terms pertaining to electrons.

Acceptable wavefunction solutions to the BO-SE for a chemical system each have an associated energy E. Using a given wavefunction solution it is possible in principle to describe properties of a chemical system and thus to make accurate predictions of chemical behaviour. A Hermitian operator acting on a wavefunction represents an observable such as energy, position, or momentum, depending on the form of the operator. [100] The eigenvalue of that operator corresponds to a measurement of that observable. In practice, the SE can be solved exactly for single-electron systems. However, for most systems relevant to computational chemistry, complications are introduced by the electron-electron interaction term and approximations are required. [167] The field of developing methods to obtain high-quality approximate solutions to the SE at reasonable computational cost is a wide one, and covers Hartree-

Fock Theory (HF), Density Functional Theory (DFT) and many post-Hartree-Fock methods.

The Hartree-Fock method [168] is often considered a basic starting point in quantum chemistry. In the method, the wavefunction describing all electrons in a chemical system is constructed from a set of single-electron functions known as molecular orbitals. These orbitals consist of a spatial orbital and a spin function for each electron. The Pauli Exclusion Principle restricts the ways in which one-electron orbitals may be combined to form a many-electron wavefunction. It states that the total wavefunction of a fermionic (eg. electronic) system must be antisymmetric with respect to exchange of any pair of fermions. Therefore, the multi-electron wavefunction Φ which is used is a determinant composed from one-electron orbitals. For example, the two-electron wavefunction composed by combination of electrons x_1 and x_2 described by the orbitals χ_1 and χ_2 is shown in equation 2.11.

$$\Phi(x_1, x_2) = \chi_1(x_1)\chi_2(x_2) - \chi_1(x_2)\chi_2(x_1) \tag{2.11}$$

Equation 2.11 is equivalent to a determinant of a 2×2 matrix whose elements represent the four possible ways that one of two electrons can be placed into one of two one-electron orbitals. The same principle of total wavefunctions obtained from the matrix determinants extends to systems of more than two electrons. A determinant of this kind is known as a Slater determinant and in Hartree-Fock theory it is a Slater determinant which is used for the wavefunction of a system. With the Slater determinant form of the wavefunction established, it can be used in the SE. Within the BO approximation, the electronic Hamiltonian acts on the wavefunction, allowing the HF energy to be expressed. Since the electronic Hamiltonian is a complex operator with contributions from kinetic energy, electron-nuclear attraction, and electron-electron repulsion, there are several components to the resultant HF energy. It is composed of sums of one-electron integrals, which arise from kinetic energy

and electron-nuclear attraction, and two-electron integrals which arise from electron-electron repulsion. The two-electron contribution is itself split into two parts, a Coulomb part, which represents electrostatic repulsion between two electrons, and the exchange part, which is purely quantum mechanical and arises from antisymmetry. Once an expression for the energy arising from operating on the wavefunction of Slater determinants with the electronic Hamiltonian is obtained, Hartree-Fock solutions to the SE can be found. Since HF is a variational method, [100] they can be found by minimising the Hartree-Fock energy. In practice this is done using an iterative self-consistent field (SCF) approach in which initial guesses for the orbitals come from solutions to the one-electron SE. The multi-electron electronic repulsion term presents a problem here. Therefore, a one-electron operator, the Fock operator, is used, which treats electronic repulsion as the interaction of one electron with the average field of all other electrons in the system. [100]

Hartree-Fock solutions found in this way provide an ab initio description of a chemical system. Properties such as energy and atomic coordinates can be extracted to a greater degree of accuracy and transferability than can generally be expected of a classical method. However, the Hartree-Fock method is only an approximation to the correct solutions of the SE. Importantly, it neglects electron correlation. In the molecular Hamiltonian, the electron-electron repulsion term depends on the positions of all electrons by simultaneously making consideration of all possible pairwise interactions. The one-electron orbitals and Fock operator are unable to account for this simultaneous interaction of all pairs of electrons, instead treating each electron as interacting with the mean field of all other electrons. [100] Several post-Hartree-Fock methods have been developed to include consideration of correlation, including multiconfiguration interaction and perturbation theory. [100] The modern limit of computational chemistry is beyond the limits of HF, but with more comprehensive methods comes an increase in computational cost, so Hartree-Fock retains usefulness depending the extent of correlation in the system studied and the required

accuracy.

2.2.1 Density Functional Theory

Density functional theory (DFT) is a very widely used method in computational chemistry. The method revolves around obtaining a description of a system using a simpler representation based on electron density rather than the one-electron orbitals of Hartree-Fock. The electron density is a function of three spatial coordinates and can be obtained from a wavefunction by integrating probability density over all space. It represents the probability of finding electrons in a given location.

The concept of using electron density as a basis for determining ab initio properties of a chemical system owes its rigorous basis to the theorems of Hohenberg and Kohn. First, any two external potentials associated with the same density must be the same up to an additive constant, meaning that density can determine the Hamiltonian of a system and therefore the Hamiltonian and the energy can be written based on density. [100] Since density is itself a function of spatial coordinates, a function of it is a functional: a density functional. The energy functional of the density can be written as equation 2.12, where T is a functional for the kinetic energy part, E_{eN} is a functional for the electron-nuclear interaction energy and E_{ee} is a functional for the electronic interaction energy, a many-electron term.

$$E[\rho(r)] = T[\rho(r)] + E_{eN}[\rho(r)] + E_{ee}[\rho(r)] \tag{2.12}$$

Knowledge that electron density determines the Hamiltonian is not useful for obtaining solutions to the SE without some indication of how to predict electron density. The second theorem of Hohenberg and Kohn aids with this. It states that density obeys the variational principle: the true ground state electron density corresponds to a minimum energy, and any approximate density leads to a larger value of energy. Therefore, solutions can in principle be found

by an iterative minimisation procedure using the same ideas as those applied for HF. For this to be useful, some method to predict and optimise electron density of a general system is needed. In principle, a candidate electron density can be used to generate a candidate Hamiltonian and wavefucntion, but this on its own is not much help, and still requires solution of the many-electron SE with the problem of the many-electron term. [100] This many-electron interaction term has a non-classical contribution whose functional form is not known, and similarly the kinetic energy has contribution from electron interaction, whose functional form is not known. Both require approximation.

A practical approach is Kohn-Sham DFT. In this scheme, the physical system which contains the many-electron interactions is replaced by a fictitious system representing the same particles (so the same overall density) but without interactions. The Hamiltonian for the fictitious system can be broken down to a sum of one-electron operators and can be solved using the same methods as HF theory. The energy functional of the real system can be divided into components which include parts relating to the fictitious system, as in equation 2.13. [100]

$$E[\rho(r)] = T_s[\rho(r)] + J[\rho(r)] + \frac{v}{\rho} + E_{xc}[\rho(r)] \tag{2.13}$$

Here, T_s is a functional for non-interacting kinetic energy, J is a functional for classical Coulomb interaction energy, $\frac{v}{\rho}$ describes the interaction of the electron density with the external potential, and E_{xc} is known as the exchange correlation functional. All parts of this functional with the exception of the exchange-correlation functional can be readily obtained from the non-interacting system and a Slater determinant. [100] E_{xc} encompasses all energy contributions neglected by the non-interacting system, including correction to the non-interacting kinetic energy based on electron interaction and non-classical correction to the electron-electron energy. [100] If the exchange-correlation energy term were known and expressed in terms of a single Slater

determinant, equation 2.13 could be used with an SCF procedure to find the orbitals that minimise the energy and therefore the exact electron density of a chemical system. However, the form of E_{xc} is not known, and the search for good E_{xc} functionals is a large part of DFT research. [100]

Several functionals for the exchange correlation energy exist, with varying degrees of accuracy and cost, including purely theoretical functionals as well as semi-empirical functionals which take some parameterisation from experiment. They may additionally be augmented with empirical addition of dispersion corrections, as most DFT functionals are unable to capture the long-range interactions that depend on small changes in density. DFT functionals are often looked at in the context of a 'Jacob's ladder', in which functionals on the lowest rungs are expected generally (but not always) to perform worse and functionals on higher rungs to perform better. [169] On the lowest rung, local density approximation (LDA) functionals use only local density and are generally based on known functionals describing a uniform electron gas. Generalised gradient approximations (GGA) account for the non-uniformity of electron density by including information based on its gradient, often as a correction to an LDA. Meta generalised gradient approximations (mGGA) add more non-local information based on kinetic energy density to GGAs. Hybrid functionals are composed of some combination of a pure DFT functional and some exchange component calculated by another method (e.g. HF) in a specified ratio. Hybrid functionals can come in the form of both hybrid-GGA and hybrid-meta-GGA.

A general challenge for DFT functionals is accurate representation of metal-containing systems, [169, 170] which typically contain many energy levels close to the ground state and display significant static correlation arising from close-lying electronic states. For systems of this kind, it has been suggested that the approximation inherent to HF is more severe than the approximation inherent to DFT, and that hybrid functionals are likely to perform less well than their non-hybrid counterparts. [169, 170]

An appropriate exchange-correlation functional seeks to replicate the properties of the unknown exact exchange-correlation functional. Although it cannot be calculated, certain features of the functional can be known, and are used as constraints when constructing approximate functionals. [171] Approximate functionals can also be considered in terms of 'appropriate norms', sets of electron densities for which they can be considered exact or close to exact. Non-empirical GGAs including the Perdew-Buke-Ernzerhof (PBE) [172] and non-empirical meta-GGAs including the Tao-Perdew-Saroverov-Scuseria (TPSS) [173] functionals have been created with the aim of increasing the number of constraints which are satisfied. The Strongly Constrained and Appropriately Normed (SCAN) functional was introduced to satisfy all 17 known exact constraints, and has displayed a more accurate performance for common test molecules than other well-known functionals. [171, 174]

2.2.2 Basis Sets

Up to this point, the sets of specific functions used to construct molecular orbitals have been referred to only somewhat obliquely, but they must now be addressed. Known as basis sets, [100] they are instrumental to most practical applications of both HF and DFT. They are discussed here in the context of a one-electron orbital model.

A wavefunction, as discussed above, can be composed of orbitals. In order to use the iterative SCF procedure, some candidate orbital is needed. Orbitals are expressed as basis sets, which are linear combinations of known functions, as in equation 2.14, in which Ψ_i is the molecular orbital, ϕ_v is a basis function of known form, c_{v_i} is a parameter to be found and the sum is over all basis functions in the basis set.

$$\Psi_i = \sum_v c_{v_i} \phi_v \tag{2.14}$$

The iterative SCF process determines the c coefficients of the basis func-

tion. The functions themselves, ϕ, are specified as part of the basis set, which is selected by the researcher. Their form does not change over the course of the SCF procedure. Therefore, suitable functions which both can be computed efficiently and allow a good description of the electrons in the system are necessary, and the quality of the basis set selected is a limiting factor of any computational method which uses them. Larger basis sets (those composed of more functions) are generally, but not always, likely to offer a fuller description of a system and therefore better quality results. In the theoretical limit of an infinite basis set, the performance of a method is as high as possible for that method. [100] The extra calculations required for adding basis functions generally increase computational cost. The scaling of cost as functions are added is dependent on the method.

Different types of functions can be used as basis functions, and selected basis functions often aim to reproduce the behaviour of solutions to the 1-electron SE. Slater functions (equation 2.15, where s is a Slater function, r is positional coordinates and ζ is an arbitrary coefficient) give a good representation of the features of 1-electron MOs but it is not easy to evaluate molecular integrals over them. [100]

$$s(r) = e^{-\zeta|r|} \tag{2.15}$$

Computation of several Slater functions is expensive. Gaussian functions (equation 2.16 where g is a Gaussian function, r represents positional coordinates and ζ is an arbitrary coefficient) are much easier to calculate and can give a limited approximation of the behaviour of Slater functions. To improve the reproduction of Slater functions, particularly at short range, linear combination of multiple Gaussian functions can be used. The functions are combined in fixed ratios into one contracted basis function, and individual Gaussian functions to be combined are referred to as primitive functions. [100]

$$g(r) = e^{-\zeta(r)^2} \tag{2.16}$$

With the form of the basis function selected, contracted Gaussians being a common choice, the basis set must be constructed from them. Many families of basis sets exist. One very common example is the split-valence basis set, which contains multiple basis functions for each orbital described. A split valence set is referred to with basic notation X-Y_{1-n}G, where X and all Y_i are integers. The G stands for Gaussian, and means all primitive functions are Gaussians. The number before the hyphen, X, refers to the number of primitive functions which make up core (non-valence) orbitals. There is generally more than one number Y after the hyphen, and these refer to the functions used to compose the valence orbitals. The amount of numbers after the hyphen, n, is the number of orbitals used for each occupied valence orbital counted in the structure. Each orbital i is composed of Y_i primitive Gaussians.

With the number of standard orbitals established, it is also possible to add polarisation and diffuse functions to this setup, indicated by * or + in the notation of contracted Gaussians. Polarisation functions are basis functions corresponding to higher angular momentum which are are added to the valence orbitals. In the split valence basis set notation, they are indicated by one asterisk (*) (polarisation on all atoms other than H and He) or two asterisks (**) (polarisation on all atoms). Diffuse functions are Gaussian functions with a very small exponent which approach zero at larger r than the standard primitive Gaussians and are needed for treating atoms whose electrons display diffuse behaviour. In the notation of contracted Gaussians, a plus (+) is added if one diffuse function is included for every valence angular momentum for all atoms except H and He, and two pluses (++) are added if the same is true of all atoms. For example, the split valence basis set 6-311++G** uses a contracted function composed of 6 primitive Gaussians for all core orbitals, and three orbitals for each 'real' valence orbital, one composed of a linear combination

of three Gaussians and the others composed of one Gaussian each. It also uses polarisation and diffuse functions for all atoms.

2.3 Large Scale Simulation Methods

Different methods for determining the interaction properties, particularly the energy, of a given chemical system have now been addressed. These can take the form of classical force fields, generally parameterised to empirical data and cheap to apply, or ab initio methods, generally more accurate and more expensive. These methods alone give information about an individual snapshot of a state of the system. Calculation of energy for individual snapshots can be combined with large-scale statistical methods to describe the macroscopic behaviour of a system over time or at equilibrium under real conditions. Either classical or ab initio methods may be used to calculate energy, but it is common to use classical energy alongside large-scale methods because of its lower cost.

2.3.1 Monte Carlo Methods

Monte Carlo (MC) simulations are a common tool for determining equilibrium properties of the components of a system from statistical average values. Considering a system made up of some number of particles, there are several possible states that the system could occupy, defined by positions and momenta of particles. [100] Under real conditions, the system will occupy a probabilistic distribution of different states, by which its properties are defined. [100] All of the possible states of a system make up a very large phase space, of which each configuration of positions and momenta is one point. Monte Carlo methods sample a statistical number of configurations in phase space, enabling determination of the properties of a system. [145] They rely on the concept of ensemble averages. Within an appropriate ensemble - that is, set of conditions held constant - the average value of a property at equilibrium is defined by an integration over phase space of a function which describes

that property in terms of positions and momenta along with the probability of occupying those positions and momenta, as in equation 2.17, where A is some arbitrary property, the vectors $\mathbf{p}$ and $\mathbf{q}$ represent all position and momentum coordinates of the system, P is a function for the probability of given positions and momenta based on a Boltzmann distribution, and the denominator enforces normalisation. [100] This particular equation applies to a system in which temperature, pressure and number of particles are held constant.

$$\langle A \rangle = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}} \tag{2.17}$$

The Monte Carlo process works by sampling possible positions and momenta in phase space to obtain appropriate ensemble averages. In principle this could be done with completely random sampling, selecting states at random and calculating their $A(\mathbf{q}, \mathbf{p})$ and $P(\mathbf{q}, \mathbf{p})$. It is clear that states with a low probability of occupation will have small contributions to the total integral and the reverse for states with a higher probability. With enough sampled states, the integral would resemble that over all phase space. Practically, however, the size of phase space is far too large for this to be feasible. Instead, the Metropolis Monte Carlo [175] method can be applied to sample points in phase space with a bias towards evaluating states with a high probability of occupation. [100]

The Metropolis Monte Carlo process is based on selecting configurations to use to evaluate the ensemble average based on the probability distribution of configurations, rather than selecting them at random and subsequently applying the probability distribution. [100] The process is begun by generating an initial starting configuration of all components of the system, which may be done at random, and evaluating the energy of this configuration using some computational method. The system then undergoes an attempted Monte Carlo trial move and its energy is calculated again. Some pool of available trial moves is specified as an input to the simulation. There is a wide range of pos-

sible trial moves to use in simulations of chemical systems, including rotation, translation, removal, or reinsertion of a molecule, with any move that changes the state of the system a possible candidate. [82] After the move, the new state is either accepted or rejected based on an acceptance rule, and if accepted the resulting configuration is added to a Markov chain, a list of accepted configurations. The acceptance rule must be such that the Markov chain, under sufficient sampling, represents the probability distribution of states of the system. It is based on a Boltzmann weighting under the conditions applying to equation 2.17. The result is that the move has a probability p of being accepted, with p given by equation 2.18, where E_1 is the potential energy of the original configuration and E_2 is the potential energy of the trial configuration, k_B is the Boltzmann constant and T is the temperature. [100]

$$p = min\left[1, \frac{e^{-E_2/k_BT}}{e^{-E_1/k_BT}}\right] \tag{2.18}$$

By equation 2.18, if the potential energy is reduced, the denominator of the fraction is smaller than the numerator, so the fraction is greater than 1 and the move is accepted with a probability of 1. If the potential energy is increased, the fraction determines the probability, and causes the sampling to have the appropriate Botlzmann distribution. [100]. The process is then repeated for a number of steps. At each step the trial move is either accepted or rejected and if accepted it is added to the Markov chain. [145] This must first be done over a sufficient number of moves for the simulated system to become equilibrated, then repeated to create a Markov chain which can be used in calculating ensemble averages of the system's properties. [145] Any property which may be calculated from a system's configuration and described using the relevant probability distribution may be calculated as an average over all accepted moves. The number of Monte Carlo steps used must be sufficient for statistical significance. Depending on the ensemble used, properties such as energy, dipole moments, and number of particles can be evaluated. Over a

Monte Carlo production run composed of a series of steps, it is also possible to obtain a statistical uncertainty in the average properties calculated. This is done by splitting the simulation into blocks composed of equal amounts of steps and obtaining standard deviation from the results of each block. In this work, uncertainties are obtained using five equal blocks from production runs and 95% confidence intervals. Evaluating the number of particles in a system using Monte Carlo simulations is the basis for a number of adsorption simulations of gas uptake in MOFs (see below).

Monte Carlo calculations take place within a statistical ensemble. [145] The ensemble referred to above, in which temperature, pressure and number of particles is held constant, is called the canonical (NVT) ensemble. [100] An ensemble very widely used in gas sorption simulations, [145] including in this work, is the grand canonical (μVT) [176] ensemble in which chemical potential (a measure of the energy which addition of a particle to a system causes to be added to that system), volume and temperature are held constant. Its acceptance rules follow those for the canonical ensemble, but it allows insertion and deletion of particles, which is necessary for sorption studies. Other commonly used ensembles include the NPT ensemble, which holds number of particles, pressure and temperature constant and can be used to model changes in volume, and the Gibbs ensemble, which involves separation of a system into two different periodic boxes. Each box can undergo moves in the NVT ensemble and can undergo individual volume change moves keeping total volume constant, and particles are also able to transfer between boxes. This allows simulation of the equilibrium between two phases and so is also relevant to adsorption simulations. [145] The grand canonical ensemble is very similar to the Gibbs ensemble but has the second box infinitely populated.

Within an ensemble, Monte Carlo simulations sample points around a potential well. This can be a problem if the system contains steep energy barriers between wells, or if the aim is to model low probability regions of the energy landscape. Approaches exist to combat this. One is umbrella sampling, in

which a potential is used which is biased to the areas of the potential landscape to be studied [100] Another, Wang-Landau sampling, involves sampling energy space based on the inverse probability of the density of states. [177]

Whatever the specifics of the method, at each Monte Carlo step the energy of the system is calculated in order to apply the acceptance rule. It has already been mentioned that this can be done using different computational tools of varying accuracy and cost. Computational cost permitting, energy might be evaluated using ab initio methods, but for large systems this is often prohibitively expensive and a classical force field is used instead. Several of the simulations applied in this work employ a two-body non-bonded force field based on the Lennard-Jones potential for van der Waals interactions along with the Coulomb equation for electrostatic interactions (see section 2.1.2).

Grand Canonical Monte Carlo to Predict Gas Adsorption

Monte Carlo simulations in the grand canonical ensemble are referred to as GCMC simulations, and as already mentioned they are a common [53, 84] and pertinent choice for modelling gas sorption events in porous materials such as MOFs. Since the grand canonical ensemble allows number of particles to vary, it may be used in simulations to determine the number of particles of a guest gas in a system at equilibrium under a given set of conditions. The equilibrium number of guest gas particles is obtained from the average number of particles in the system over the accepted configurations in the Markov chain of the production run (occurring after the initial equilibration run).

It is useful to control external pressure and temperature to determine the relationship between these conditions and adsorption. Temperature is evidently controlled in the grand canonical ensemble. Although the pressure of the MOF/gas system is able to vary, pressure of the external gas reservoir is controlled by its relation to chemical potential. With both temperature and external pressure controlled, it is possible to produce a computational adsorption isotherm - that is, a plot of adsorption of a gas in a material with changing

pressure at constant temperature. For a set temperature, simulations can be carried out for a system of a MOF in equilibrium with a particular adsorbate at each of a selection of pressures, and the resulting loading values plotted. This is useful data which can be compared to equivalent experimental isotherms. It is common to generate computational isotherms when making consideration of one or a few MOF structures, although in more data-intensive high-throughput studies a researcher is more likely to calculate gas uptake at only one or a few pressures and temperatures.

Whether aiming for a full isotherm or for a measure of loading at a single pressure, the initial result of a gas uptake GCMC procedure is an absolute loading, a raw measure of the amount of molecules in the available volume. For direct comparison to experimental values, this must be converted to a value equivalent to the excess loading which can be measured using physical equipment and is commonly reported in experimental studies. Excess loading is a measure of the amount of extra molecules that are in the system compared to those that would occupy a system whose volume is equivalent to that of the pore network at standard temperature and pressure in the absence of adsorption. [90] The absolute and excess loading, n^{abs} and n^{ex}, are related by equation 2.19, where V^g and ρ^g are the pore volume of the framework and the molar density of the gas respectively.

$$n^{ex} = n^{abs} - V^g \rho^g \tag{2.19}$$

Using appropriate software such as the RASPA software package, [82] excess loading can be calculated as part of the simulation. To do this, pore volume must first be calculated and specified in the form of void fraction. This is the fraction of the unit cell that is not taken up by framework atoms, and as it is a volume measure it is subject to the volume considerations discussed in section 2.1.1. Experimental void fractions are commonly determined using a helium probe. Whether computational volume is calculated using purely

geometrical or force field methods, a probe representing a helium atom is commonly used, and the void fraction calculated is referred to as the helium void fraction.

Excess adsorption is not to be confused with net adsorption, which is a measure of the amount of extra molecules in the system compared to those that would be in a fluid occupying the same volume as both the pore network and the host framework. [178] Net adsorption calculations would consider the total volume rather than the pore volume, and net adsorption values are available from experimental measurements in a similar manner to excess adsorption values. There are advantages to using net adsorption as opposed to excess adsorption, such as the absence of ambiguity over the definition of pore volume and a direct usefulness to gas storage applications. [178] However, excess adsorption is more commonly measured by experiment and computational values of excess adsorption are calculated here where relevant for comparison to experimental results.

To run an adsorption study of a gas in a framework, a number of researcher-specified inputs are required. Perhaps most obviously, definitions and coordinates for the host and guest atoms are required. For a classical adsorption study, host and guest atoms are defined using a set of charges and force field parameters as laid out in section 2.1.2; an ab initio simulation would define atoms and molecules using atomic numbers, charges and spins. The coordinates of host atoms can be specified in a crystallographic information file (cif). They commonly come directly from experimentally resolved X-ray crystallography of an observed structure, [25] but can also be crystallographic coordinates which have been relaxed by a computational method, [93] or can be fully computationally generated, as for hypothetical MOFs, [33] which are composed from combinations of fragments of real MOFs. Coordinates of a unit cell are generally taken, and periodic boundary conditions applied to replicate the bulk material. Once the coordinates are defined, it is common for them to be held fixed throughout a simulation, approximating the framework as rigid.

Although this is not always a completely accurate representation, [76, 112] it is often sufficient to give useful information. [53, 84, 146] A specified number of guest atoms or molecules are also assigned coordinates at the beginning of a GCMC calculation. This generates the initial configuration for which energy is first calculated, but guest coordinates change throughout a simulation.

There are several other inputs required in a GCMC simulation, which represent researcher-specified choices regarding computational setup. These include the number of Monte Carlo steps at each stage, the force field used, the size and nature of the van der Waals cutoff, the number of replica unit cells to use when implementing periodic boundary conditions (which must be such that all perpendicular distances are greater than two times the van der Waals cutoff), whether any parameters are to be adjusted, and, if applicable, the method used for framework charge calculations. Once a setup is established with a given combination of inputs and conditions, it is pertinent to test its ability to model systems similar to those of interest, to confirm its applicability to the problem at hand.

2.3.2 Molecular Dynamics Methods

Monte Carlo techniques are certainly useful for obtaining information about average properties of a system at equilibrium (or beyond with advanced methods). However, they do not give information about the time evolution of that system or its properties. Each trial configuration tested is not linked in time to the previous trial configuration; it is simply a snapshot of some other possible state. To take account of time evolution, a second family of methods, molecular dynamics (MD) is used. Like Monte Carlo, MD takes place within a specified ensemble

Molecular dynamics involves examining instantaneous states of a system which is evolving over time. It is possible to determine properties at each state, and one of the uses of this is calculation of average equilibrium properties as

a time average over all states visited, as in equation 2.20, where A is some arbitrary property, M is the number of points in time (timesteps) at which the property is to be sampled and t_i is time at a given timestep. [100]

$$\langle A \rangle = \frac{1}{M} \sum_{i}^{M} A(t_i) \tag{2.20}$$

Any practical sampling happens discretely and finitely, as in equation 2.20. In the hypothetical limit of continuous infinite sampling over a time t, equation 2.20 becomes equivalent to an integration rather than a sum over t, as in equation 2.21. [100]

$$\langle A \rangle = lim_{t \to \infty} \frac{1}{t} \int_{t_0}^{t_0+t} A(t) dt \tag{2.21}$$

By the principle of ergodicity, if sampling occurs over a large enough time, equation 2.21 is equivalent to the ensemble average given in equation 2.17 and used in Monte Carlo simulations, independent of the choice of initial configuration. This means that MD can be used in a similar way to Monte Carlo. MD typically takes longer to sample sufficient points to reach equilibrium than MC, so its primary use is not in seeking information that MC can obtain. It can, however, be used to access ensemble average information about properties that depend on time and cannot be accessed via Monte Carlo simulations, such as displacement properties. It also allows a researcher to watch evolution of a system in real time and make conclusions about the processes happening within that system.

The possible utilities of MD established, it is clear that to be able to use it practically, either the integral in equation 2.21 must be evaluated analytically or it must be possible to generate states of the system at selected times based on some initial starting state. In real systems $A(t)$ is generally not a known analytical expression, but it is possible to generate approximations of future states of the system from a starting state. This relies on the fact that points in phase space are defined by position and momentum, and the fact the trajectory

of future states of a system is defined by a combination of initial positions and momenta and the forces acting on the particles. Force is the negative of the derivative of potential energy with respect to position, and in a chemical system the potential energy experienced by one particle results from the presence of all other particles, and can be modelled by force field or ab initio methods depending on the problem at hand.

The task, therefore, is with known positions and momenta of all particles at time t, to determine positions and momenta at some slightly more advanced future time defined by the sum of the original time and a small timestep, $t + \Delta t$. This is an approximation as the true trajectory $A(t)$ is continuous, but with very small timestep Δt the approximation can be reasonable. The positions and momenta of all particles at time $t + \Delta t$ are given by the sum of the position or momentum at time t and the product of Δt and the derivative of the position or momentum. This is shown in equations 2.22 (position) and 2.23 (momentum), where $\mathbf{q}$ is position, m is mass, and $\mathbf{p}$ is momentum. The equations use the facts that (i) the derivative of position with respect to time is velocity, (ii) the derivative of velocity with respect to time is acceleration, and (iii) velocity equals momentum over mass. [100]

$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \frac{\mathbf{p}(t)}{m}\Delta t \tag{2.22}$$

$$\mathbf{p}(t + \Delta t) = \mathbf{p}(t) + m\mathbf{a}(t)\Delta t \tag{2.23}$$

Positions and momenta are known and $m\mathbf{a}(t)$ is equal to force, which can be obtained from the modelled chemical potential. From this point it is a straightforward matter to, for some timestep, determine updated values of position and momentum for all particles, and continue the simulation with the specified timestep until satisfied.

However, the method of dealing with the discrete timestep in equations 2.22 and 2.23 is not ideal as it is based on the tangential gradient and so will

not find a point exactly on the trajectory curve. Equations 2.22 and 2.23 are equivalent to Taylor expansions of position and momentum with respect to time about time t truncated at the first order. To obtain a better approximation of continuity, it is preferable to use similar equations of Taylor expansions truncated at some higher order. A second order Taylor expansion brings an extra derivative into each equation, which complicates the solution. Algorithms for solving molecular dynamics simulations at higher orders are readily available and routinely in use. [100] With an algorithm of this kind, it is possible to use MD to obtain good approximations of continuous trajectories.

It has already been mentioned that any MD algorithm involves some evaluation of force as in equation 2.23, and that force is calculated from potential energy, which can be modelled by classical or ab initio techniques. In modelling of uptake situations similar to those already described in the context of Monte Carlo techniques, classical force fields are generally preferable due to their lower cost. However, there are some situations for which classical force fields are insufficient. For example, when modelling the damage done to materials during imaging under an electron beam, it is common to use molecular dynamics simulations to observe damage which may occur. Under such circumstances, flexibility of the material and bond breaking are instrumental to the processes observed. A non-bonded Lennard-Jones force field is clearly inappropriate, and a force field parameterised for flexibility is unlikely to retain sufficient transferable accuracy to model the required situations. Therefore, in such situations, ab initio methods may be used to determine forces at each timestep. [131, 132] The combination of ab initio methods and molecular dynamics is referred to as ab initio molecular dynamics (AIMD) (see Chapter 6).

2.4 Machine Learning Methods

The methods addressed up to this point in this chapter cover many popular ways to determine information about individual chemical systems. To varying degrees and as a result of varying computational cost, they can be used in parallel or sequence to model several chemical systems and obtain information on a high-throughput scale. For instance, GCMC simulations can be used to model thousands of MOF systems to make predictions about gas uptake trends and high performing materials, though the scale at which this can be done is ultimately limited by the cost of GCMC, with uptake information taking up to days to obtain. It has been discussed in chapter 1 that machine learning (ML) methods may complement more historically conventional computational chemistry techniques to allow information to be obtained about chemical systems with increased efficiency.

Machine learning refers to a family of statistical methods by which a computer learns from some experience to perform some task without the specifics of the task being explicitly programmed. [179] Early exploration into machine learning applications involved computer programs learning to perform well in playing board games, [179] and the field has since seen great expansion to be employed in researching many problems throughout almost all areas of science and technology.

Machine learning algorithms involve detection of patterns between or within sets of data which can then be applied to new data outside of that on which they were trained. In the context of computational materials modelling, this can be useful in a number of ways. One example is that machine learning can aid in accurate and efficient parameterisation of force fields. Using ab initio data corresponding to only a few points on a potential energy surface, an algorithm can effectively learn unseen parts of the potential surface, which can be used to develop a force field for use in modelling with the aim of producing ab initio quality results at force field operation cost. [114] Further, ML can be

used to directly predict material properties. Models may be trained to use data which is cheap to calculate to predict the outcome of more expensive property calculations. This kind of model is used in MOF research to predict GCMC-style results of uptake properties based on some selection of cheap data, which is usually at least partially composed of geometrical data. This can be either in the form of a regression model, which predicts the value of some continuous MOF performance metric, or a classification model, which separates candidate MOFs into classes and can be used to predict whether they have good or bad performance as defined by a relevant performance metric. An appropriately trained model of this kind can be used to reduce the cost of high throughput screening by reducing the need for its time-limiting step, GCMC simulations. [88, 180] Machine learning is employed in this manner in chapter 4, and the remainder of this section focuses on machine learning methods in this context.

2.4.1 Selection, Preparation and Exploratory Analysis of Data

The starting point for a machine learning property prediction model is data, and there are several useful considerations to make and processes to follow when it comes to ensuring that the data used is likely produce appropriate models. A model must be able to calculate some expensive property or properties of a material (herein referred to as labels) from some set of cheaper properties of the same material (referred to here as features). The starting data is some set of material structures about which the label(s) and features are known. As large a dataset as possible is required at this stage to maximise the chance of the model being well-performing. The selection of this dataset is important, and can have a bigger impact on the quality of a trained model than the nature of the model itself. [181] The full dataset is split into data which will be used in training the model and data which will not be seen by the model in training, but which will be used in testing its performance. For

a MOF study, the dataset is a set of MOF structures, with either known features and labels (which can come from experimental or computational results) or known structural coordinates which facilitate determination of features and labels by classical or ab initio methods.

The particular data that is used to train a machine learning model must be selected and can have a significant impact on its performance. Often the first stage of building a model is selection of materials whose properties are to make up the dataset. For any high-throughput MOF study, structures come from some MOF database, of which a number are published with open access. As discussed in chapter 1, these include databases of structures which have been experimentally synthesised [11, 21] and databases of hypothetical MOFs algorithmically generated from known structural moieties and rules. [33] Selection of the database from which structures will be taken is a vital part of the machine learning training process, and both real and hypothetical MOF databases are accompanied by advantages as well as issues which must be considered. Hypothetical MOF databases are more extensive than databases of real MOFs, containing up to hundreds of thousands of structures compared to tens of thousands. However, as discussed in chapter 1, it is by no means guaranteed that every structure in a hypothetical database is practically synthesisable, while all structures in a real database have been synthesised in some form. Common databases of both forms, meanwhile, contain a significant proportion of cifs corresponding to unviable structures as a result of issues with structural determination or computational structure preparation, which was also addressed in chapter 1. Use of such structures in a machine learning algorithm is certainly problematic. If a portion of the training data fed to a model is not relevant, the ability to make high quality predictions will be reduced: [182] 'garbage in, garbage out' is an epigram used to express this issue. Structure curation procedures therefore ought to be an early step in the training of any machine learning model of this kind, and it is fortunate that steps are being taken among the high-thoughput MOF modelling community

to prepare databases with lower populations of problematic structures. [109]

A second feature of structural databases which must be addressed when obtaining data from them for use in machine learning models is bias. A given database may by chance be unevenly weighted towards structures possessing particular categories of features or performance indicators. For example, one MOF database may contain a particularly high proportion of MOFs with pores of a particular size, which also have a high uptake of some target gas. In a hypothetical database, this may result from several structures all being made from the same limited selection of building blocks. A model trained on such a database may have particularly strong predictive performance for MOFs with this popular pore size, but perform less well for MOFs with differently sized pores. When applied to the structures of some other differently populated database, such a model may also disproportionately predict high uptake of the target gas. The transferability of a machine learning model is an important consideration, and it can be useful to test models on data taken from sources different from that of the training data. The extreme of the problem of bias is a dataset containing several copies of the same structure, and this can be an issue for MOF databases. When a MOF is synthesised experimentally, it is deposited in a database, and in some cases different groups may deposit structures representing the same MOF synthesised using some alteration in conditions. This results in structures which differ only by a very slight change in coordinates, to the extent that there are more than 50 copies of the famous MOF CuBTC in the CSD MOF database. [183] This can lead to clearly visible issues in ML models: a model trained on a database containing several copies of the same MOF will obtain near-perfect predictive power for that MOF, but its bias towards that MOF is likely to reduce its predictive power for other structures. When developing machine learning models for MOF databases, it is prudent to check for exact duplicates and remove them. In terms of wider bias reduction, there have been recent efforts to facilitate the training of transferable machine learning models by publication of balanced, unbiased

datasets. [109, 184]

With a dataset of materials selected, features and labels for each member of the dataset are needed. The nature of the labels is determined by the problem to be addressed: if seeking a material with good gas uptake properties, some label which defines the gas uptake properties of a material is needed. As a basis, gas sorption labels involve calculated or measured loading of a gas at a given pressure, but the precise nature of the label depends on the desired properties. Selectivity, working capacity, and labels which address the trade-off between different desired properties can be selected. [88, 180] When it comes to features, there is fairly large scope for selection of descriptors which are most useful. Features must be cheaper to obtain than the labels: otherwise there is little utility in the model. To obtain a high quality model, they must additionally display a statistical relationship with the labels which they are to be trained to predict. Features of multiple kinds may be pursued. For example, it is possible to define features of frameworks based purely on a fingerprint of atom composition and connectivity [185] or string representation similar to the Simplified Molecular Input Line Entry (SMILES) strings which are commonly used to define organic molecules, [186] without further modelling or calculation. [181] Resulting features are large and high-dimensional. The alternative, commonly used in MOF studies, is use of features based on some cheap modelling calculation. Geometrical features are a popular choice, as MOF geometry is commonly seen to be closely related to behaviour, and geometrical properties can be calculated very readily without the use of force fields (see section 2.1.1). Several geometrical features exist, including those describing pore sizes, accessible, non-accessible and total surface area and volume in both gravimetric and volumetric form, and density. Features based on energetic interactions have been seen to improve models [187] where cheap enough to be practical.

With the range of available features established, it may seem an attractive option to apply a large volume of them to a model in order to benefit from

several statistical relationships with the labels. However, adding additional features does not always improve a model and can be detrimental to its performance. Addition of a feature which does not have a particularly strong statistical relationship with the label, for example, may add more noise than useful information to a model. Further, adding each new feature adds to the dimensionality of feature space, adding at a fast rate (eg. exponentially) to the number of parameters that must be fitted, an issue referred to as the 'curse of dimensionality' [188] If the set of training data is not large enough, it may not be sufficient to describe such a high-dimensional system and overfitting may result. [181] As a general rule, if the number of features is more than 10% of the number of members of the training set, dimensionality problems may be expected. This makes the use of fingerprint-based features difficult except for very large datasets, and is also a cause for caution in cases where features are taken from among the several available calculated descriptors. Further, with every additional feature that must be calculated when training and applying a ML model comes additional computational cost. Since the aim of the endeavour is to reduce computational cost associated with chemical calculations, reduction of the number of features is desirable from an efficiency perspective.

When building a machine learning model, a researcher may have tens of calculated features or higher-dimensional fingerprint features available to them. It is prudent to select some smaller number of features to use in a ML model. This can be done in one of two ways: feature extraction or feature selection. [189] Feature extraction methods involve using some large number of features to generate some smaller number of features. These include principal component analysis (PCA), in which some new list of features is generated, beginning with the feature that best captures the variance in the training dataset, followed by the feature orthogonal to it which captures the next most variance, and so on. [188] Feature selection is more conceptually simple: it involves examining a set of available features and selecting a subset of those features which is expected to best describe the behaviour of the label. The process of

feature selection tends to involve seeking features which display a high correlation with the label according to the training set, but avoiding high correlation among features. High correlation among features would render some of them redundant, and can be considered in the form of individual correlations between features, or of multicolinearity, which defines correlation between one feature and all other features. Feature selection can involve various steps, such as calculating Pearson's or Spearman's correlation coefficients between features and labels and selecting those for which correlation is highest, or calculating correlation coefficients between pairs of features and removing one of a highly correlated pair. Several in-depth approaches exist, including those which make consideration of measures of multicollinearity between features (eg. variance inflation factor (VIF)) and those which fit models using all features and measure the contribution of each to the model (eg. variable importance in the projection (VIP)). [190] Feature selection methods can be useful guides, and can be used in combination with each other and with domain knowledge to determine which features are likely to perform well.

There are a number of other considerations about data which may be made and which may inform choices about machine learning models. For example, one can calculate how far the distribution of the label differs from normal, using the measures skewness, which describes the asymmetry of the distribution, and kurtosis, which describes its sharpness. [191] A label whose distribution displays high skewness and kurtosis is less likely to be modelled well using parametric machine learning methods which assume a normal distribution compared to a label whose distribution displays low skewness and kurtosis. It can nevertheless be useful to train simple parametric models as a starting point, which can then be used as a baseline against which to compare more complex models if necessary. If the data is skewed it can also indicate that models may be less likely to predict labels well in the less populated region of the distribution, as they will have received less training data close to this region. In such cases it may be pertinent when fitting models to apply

data manipulation approaches such as oversampling or undersampling which involve artificially adding data to under-populated regions or removing it from over-populated ones to better balance the distribution. [191] Such approaches may rely on access to sufficient total training data.

A further necessary step in the data preparation process is data transformation. With data in a raw form, different features may have very different scales, and those with higher scales will be artificially weighted more strongly in models than those with smaller scales. Routine data transformations include standardisation and normalisation. Standardisation centres all features about the same mean value, typically zero, with the same standard deviation, and is achieved for each feature by subtracting the mean value in the dataset from every individual value, then dividing by standard deviation. [190] Normalisation scales the values of every feature to lie within a set range, generally between 0 and 1 or between -1 and 1. This is commonly a simple linear scaling, but does not have to be. [191] Less routine feature transformations can also be applied to improve the distributions of poorly-behaved features, an example being logarithmic transformations.

2.4.2 Machine Learning Models

With a set of chemical objects selected, and appropriate features and labels established and measured for each, the process of training a machine learning model can begin. It has already been mentioned that the data must be split into sets for training and testing. Some subset of the data which is to be used for training is fed into the selected machine learning algorithm during training, and the model's parameters are fitted to the data. Some other subset of this data is used to tune hyperparameters, and can be referred to as a validation set to distinguish it from the remainder of the training data, the training set. Hyperparameters are parameters which can be adjusted to affect the specifics of a ML model, but which are not determined automatically as

part of a fitting. In general, a model is trained using the training set for some selection of values of the hyperparameters, and the performance of the model on the validation set determines which hyperparmaeters are carried forwards into the final model. The test set is used to measure the performance of the final model. It is important that the test set does not contain any data from the training or validation sets which has been used in creating the model. [182] If it does, the true transferability of the model to unseen data cannot be measured. There are different approaches to separating training, test, and validation data. Holdout testing involves simply dividing the dataset into three subsets with different percentage populations. [189] Division of data can be done randomly or by an algorithm which attempts to make each subset of data diverse and representative of the full dataset. [187] An alternative method for defining training and validation sets is k-fold cross validation. [189] This involves division of the data that is to be used for both training and validation into some integer k 'folds'. In turn, each fold is removed from the set and the other k−1 folds are used as a training set, with the removed fold used for validation. The test set must still be removed completely from the training and validation sets to ensure it is unseen.

In general, machine learning methods involve some mathematical model which must be parameterised. During the training process, the model is optimised by minimising some loss function with respect to adjustable parameters. Several models exist which can be used for classification or regression problems, or for both. Those designed for regression can also be applied to classification tasks as they can predict labels using regression which can then be classified based on some defined threshold. It can be useful as part of a single study to train multiple models and compare their performance. More complex models may often detect patterns in data which are inaccessible to simpler models, but it is generally prudent to test both complex and simple models. [181] One of the simplest techniques, which can be used to illustrate the concepts behind training a machine learning model, is multiple linear regression (MLR). This

can be used as a starting point in training models, and as a baseline against which to compare the performance of other models. MLR is explained below, followed by descriptions of other machine learning models which are relevant to this work.

The simplest form of linear regression is two dimensional and univariate: fitting a single linear function $y = mx + c$ to some set of points $(\mathbf{x}, \mathbf{y})$ in two dimensions, where $\mathbf{x}$ is a vector of a single feature and $\mathbf{y}$ is a vector of the label. The function can be fitted by minimising some measure of the distance from each point to the line the function defines. The measure of distance is a loss function which is commonly the squared distance. This simple regression model is not generally considered to constitute machine learning, but multiple (multivariate) linear regression can be viewed as an extension of the same concept. When extending the concept to multiple linear regression which involves learning relationships between several features and the label, there are two important points to make. The first is that the function that must be fitted now involves not one x variable, but a variable x_i for every feature used. The second is that MLR models are so called because they are linear functions of the adjustable parameters, not because they are linear functions of the features x. [188] Although linear functions of x are valid within the framework of linear regression, models composed of linear combinations of non linear functions of x can give better fits. Therefore, a general form for MLR is that given in equation 2.24. [188]

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\mathbf{x}) \tag{2.24}$$

Here, $\mathbf{x}$ is a vector of all features, $\mathbf{w}$ is a vector of all weights, ϕ_j are functions of $\mathbf{x}$ with pre-defined form, w_j are parameters which are found during the fitting, known as weights (they determine the weight of each function), and M is the total number of parameters. The simple $y = mx + c$ univariate linear regression can be fit into this definition, where there are only two w and one

x, w_0 is c, w_1 is m and ϕ is simply the function $1x$. With the form of the function established, the MLR model can be trained by minimisation of a loss function such as the mean squared prediction error.

A second conceptually simple model, this one most naturally applied to classification problems, is k-nearest neighbour (kNN). In this approach, each point among the training data has a class associated with it and is represented in multi-dimensional feature space. As new unseen points points are added, they are classified according to the majority class of the k points closest to them in feature space, where k is a hyperparameter to be tuned. [188] One extreme of this method is the case where k= 1 and new points are simply assigned to the class of their single nearest neighbour. The other is the case where k is equal to the number of training points and new points are all assigned to the majority class in the training set. The kNN method can be applied to regression problems by assigning new points a label which is the average of their k nearest neighbours rather than a class which is the majority.

Random Forest (RF) is a machine learning model which is an extension of another model, the decision tree (DT). Traditionally a classification model, a DT partitions multidimensional feature space into a number of regions defined by hyperplanes, with each region corresponding to a prediction about points whose feature coordinates cause them to fall into that hyperplane. [188] It does this by a sequence of decisions regarding features. For example, one decision may split the tree into two nodes based on whether a feature x_1 is greater than a certain value. At each resulting node, another decision is applied which further partitions the feature space. The process is continued until some threshold condition is met, and predictions are made about points based on the hypercube into which they fall. In its most basic form, a decision tree classifies new points according to the majority population (classification) or mean value (regression) in their hypercube. However, they can also be more complex, and can include different ML models within each hypercube for classification or regression. Training a decision tree involves learning its structure, which is

defined by the conditions applied at each node. [188] Hyperparameters include the threshold used to finish growing the tree. A single decision tree alone is prone to overfitting: it can give a very accurate description of its training data but does not transfer well to new data. This can be combated by fitting a random forest of several decision trees and combining their predictions. In this way, each individual decision tree may overfit, but the overfitting of all can be mitigated by the others in a wisdom of the crowds effect. [187]

Support vector machines (SVMs) are a class of models which can be applied to both classification and regression problems. They are generally examples of kernel methods: they use some kernel function which allows non-linear functions of parameters to be encompassed into otherwise linear models, thereby allowing use of non-linear space without fully parameterising it. [188] A regression model using SVM looks similar to a MLR regression model, but it makes use of kernels in the context of improving the loss function, and can achieve a better fit to the form of data than MLR. [188] A classification SVM model separates feature space into hyperplanes to distinguish between features, and uses kernels in cases where data cannot be separated by linear hyperplanes. [181]

Once a model has been fitted using training data, its quality must be assessed using the test set. This involves comparing known values of the labels in the test set (calculated or measured according to the relevant computational or experimental technique) to values of the same labels as predicted using the trained machine learning model. There are multiple ways that the performance of a regression model can be assessed, and it is useful to use these in combination as they can give different information. Some measures of the quality of a model involve calculating a single number, such as root mean squared error (RMSE), which is similar to the loss function, the correlation coefficient R and the coefficient of determination R^2. [189] In addition to these measures, it is useful to examine the distribution of model performance over the range of values the label can take. This can be done by plotting predicted value

against true value, which gives a visual representation of those areas in which the model performs poorly and those in which it performs well. From here it is also possible to examine the details of individual points which are predicted particularly poorly, and possibly to draw conclusions about which kinds of data the model can predict less well and why. When it comes to classification, model performance is generally assessed by the two measures precision and recall. For a particular class, precision is the fraction of points predicted to be in that class that were correctly predicted, and recall is the fraction of points whose true values indicate they should be in that class that were placed in that class.

2.5 Summary

Over the course of this chapter, computational chemistry methods of several kinds have been introduced. Approaches from each family are applied to MOF and materials modelling problems throughout subsequent chapters, selected based on the requirements and constraints of the problems at hand.

The first category is classical force field and geometry-based methods, which represent the physical and chemical properties of a system using parameterised models and without direct input from quantum chemistry. They are generally relatively fast to apply, but this can come at the expense of accuracy. They are favoured when modelling large systems, large numbers of systems, or long time periods. Problems of these kinds are addressed frequently throughout this work, so classical methods are often selected.

The second category is ab initio methods, which make direct quantum chemical considerations and can take into account electronic structure where classical calculations cannot. They are therefore necessary in situations where high accuracy or knowledge of electronic structure are required, or situations which are not well covered by force fields, but are also characterised by a cost-accuracy trade-off of their own. They come at a higher cost than classical force

fields and cannot easily be turned to large systems.

The third category is large scale simulation methods, which are used to obtain information about the statistical nature of a system or its propagation in time, and must be used in conjunction with classical or ab initio methods which determine the energy of individual states of a system. They include molecular dynamics and Monte Carlo methods, and are important for application of computational chemistry to situations approximating real conditions.

The final category is machine learning methods. These can be used to establish relationships between cheap data and expensive data of interest. In this way, they can efficiently make predictions about the behaviour of a system, and can improve the speed with which computational calculations can be done. They can work well in combination with other computational chemistry methods: indeed, the results of some other method often make up the training data for a machine learning model.

Chapter 3

Gas Uptake Properties of Metal Organic Frameworks Modelled Using Monte Carlo Methods: The Xe/Kr Separation

As discussed in Chapters 1 and 2, computational methods may be used to make predictions of properties of MOFs, with an important example being predictions of gas uptake properties. MOFs, along with other porous materials, are attractive for a range of gas storage and separation cases.

One such case is the separation of xenon (Xe) and krypton (Kr). Noble gases are widely used in industrial and medical applications involving lighting, [49] anaesthetics, [50, 51] and diagnostics and therapy of lung conditions, [52] and their separation is pertinent in many chemical contexts. Xenon and krypton are present in low concentrations in air amounting to 0.086 ppmv and 1.14 ppmv, respectively. [50, 192] A 20/80 xenon/krypton stream is typically obtained as a by-product of cryogenic distillation of air, and is separated into its constituent components by further cryogenic distillation. The process is costly and energy-intensive, particularly due to the very similar chemical properties of the two noble gases. It also has associated fire risks due to ozone build-up

by radiolysis. [10, 193]

Another source of xenon and krypton is as by-products of nuclear fission, and is pertinent both for production of xenon gas, and for safe sequestration of radioactive krypton isotopes formed during fission. Used nuclear fuel (UNF) contains both ^{85}Kr (half-life 10.8 years) and Xe isotopes with relatively short half-lives, such as ^{127}Xe (half-life 36.3 days), among other radioactive elements, notably uranium and plutonium. [10] The noble gases may be stored together as a 10:1 Xe/Kr mixture, [50] but effective separation of the two components is more attractive. Once separated, storage of ^{85}Kr becomes easier due to the lower total quantity of gas, while ^{127}Xe, with its shorter half-life, may be processed and used. Cryogenic distillation is again a possibility for this separation, but has the same associated risks and drawbacks. Trace levels of radioactive Kr can also be left in the resulting Xe stream. [194]

Xenon and krypton may additionally be obtained as by-products of a molten salt reactor (MSR), and produced in this way may have specific applications in medicine. [52] Such uses apply to the radioactive isotopes in their own right, before the radioactivity is spent. For example, ^{135}Xe and ^{133}Xe may be used as imaging agents in lung diagnostics, and ^{133}Xe in lung therapy to attack suitably-sized viruses using beta-radiation. [52] Since the half-lives of useful MSR by-products are short, often on the order of hours, [52] it is important that any separation technique used in this context is fast enough that the resulting radioisotopes can be used before they are spent. The operating temperatures of an MSR are high (600-800 °C), so a separation that can be achieved at high temperature is desirable. This is likely to be a particular sticking point in the emerging use of MSR by-products, [52, 86] as many porous materials which may be used in separations under ambient conditions are not stable up to such temperatures. Additionally, the success of separation methods depends heavily on the conditions under which separations are carried out, highlighted best by the popular temperature swing adsorption (TSA) method. [25]

In each relevant case, whether for an Xe/Kr stream obtained from air, a conventional nuclear reactor, or an MSR, an efficient and appropriate method is required for the separation. An alternative to cryogenic distillation is widely sought, with porous materials for adsorptive separation being a promising avenue. [31, 50, 195, 196] The use of porous materials applies to pressure swing and temperature swing adsorption and to membrane-based separations. With full valence shells, no dipole or quadrupole moments and minimal reactivity, neither noble gas has strong chemical or Coulomb interactions with pore walls to promote adsorption. Xenon is the larger and more polarisable of the two, with a kinetic diameter of 3.96 Å, [17] compared to the kinetic diameter of krypton, 3.80 Å. [17] The higher polarisability of xenon leads to many more cases of Xe than Kr selectivity through van der Waals interactions, although krypton-selective materials have been observed under certain conditions. [197, 198] In such cases, extremely small pore size leading to lack of accessibility to xenon accounts for the separation behaviour.

Experimental and computational studies have previously examined Xe/Kr selectivity in various porous materials. These include high-throughput screenings [10, 187] for promising structures and properties, high-throughput screenings for thermodynamic trends, [199] and studies of individual materials or groups of materials. [198, 200, 201] Early approaches considered activated carbon as a high-surface area material (reaching over 2,000 $m^2\ g^{-1}$) [29] with impressive gas uptake properties. [193, 195, 202] However, not having uniform pore characteristics, activated carbon lacks specificity and thus strong selectivity. In any case, it bears a fire risk due to reaction with nitrogen oxides when heated, which precludes its use for at least temperature swing adsorption methods in nuclear reprocessing. [50]

Porous alternatives to activated carbon with greater selectivity and reduced fire risk have been widely proposed, with a classic example being zeolites. [193, 194, 198, 203, 204] Zeolites can be robust and highly porous, and can reach high thermal stability (200-1000°C). [205] The narrow-pored (3.8

Å) zeolite SAPO-34 was investigated for membrane separations by molecular sieving, diffusing Kr before Xe. [194] It was additionally found to have Xe selectivity of 7 for nuclear reprocessing concentrations calculated from single-component isotherms. [194] The zeolite NaA, in which Xe is adsorbed by interaction with Na^+, has a predicted selectivity of 4. [203] The possibility to modify zeolite materials and thereby improve separation properties has also been investigated. Silver exchanged ETS-10 has improved xenon adsorption properties compared to its non-exchanged counterpart. [204] Reasonable selectivity may clearly be observed in zeolites, but they can suffer from low total capacities, lowering separation efficiency. Additionally, zeolites lack tunability compared to the more recently popular metal organic frameworks (MOFs), and thus cannot be as extensively screened for particular applications.

A less widely studied class of separation candidates, nevertheless demonstrating promising performance, is organic cages. [31, 201, 206] The cage molecule CC3 packs into a 3D structure and has very narrow pores, with a largest cavity diameter (LCD) of 4.4 Å, and a pore limiting diameter (PLD) of 3.6 Å. It is found to have very impressive xenon selectivity (20.4 at ambient concentrations) and heat of adsorption (31.3 kJ mol^{-1}), along with reasonable uptake capacity (2.69 mol kg^{-1} at 1 bar) and fast kinetics. [31] The strong selectivity is ascribed to the close geometrical fit of xenon in the pores of the structure, with no pores large or small enough to adsorb different-sized molecules preferentially.

MOFs make up a class of material that has experienced a great deal of attention for gas storage and separation applications due to their porous nature, high surface areas and tunability. [36, 45, 89, 90, 187] It has already been mentioned that MOFs can have exceptional gas uptake properties covering many different purposes and that many thousands of real and hypothetical MOFs have been published. Properties may be sought to match almost any relevant separation application. The xenon/krypton separation is no exception and a number of promising MOFs have been found. [50] Until recently,

benchmark MOFs reached selectivities of less than 20, [50] but in the last few years there has been an increase in predicted and measured selectivities of the most impressive materials.

The large and diverse chemical space occupied by MOFs makes them a prime target for high-throughput computational screenings, which have facilitated the increase in benchmark selectivity. Computational screenings enable structures to be tested on a scale simply unreachable by experiment. Sikora et al [89] screened over 137,000 hypothetical MOFs, and Simon et al [187] used machine learning methods to reduce cost further and screen over 670,000 MOFs from the Nanoporous Materials Genome. Both screenings considered a 20/80 Xe/Kr gas stream. This composition has the most widespread relevance in industry, but it could also be instructive to consider other proportions, such as those relevant to nuclear reprocessing, or a 50/50 mixture which would elucidate the behaviour when both gases are allowed to compete equally. From both the screening of Simon et al [187] and that of Sikora et al, [89] top perfoming hypothetical MOFs have Xe/Kr selectivity exceeding 70. Both studies also examined structure-property relationships with a view to aid future design and selection of MOFs for Xe/Kr separation. This yielded useful guidance, although showed that various diverse structures can behave well and poorly, proving that there is no straightforward catch-all formula for Xe/Kr separation behaviour.

Since the advent of computational screenings into the field, promising MOFs identified therein have been synthesised and tested, and the structural guidance they provide has been used to inform MOF selection. A third screening by Banerjee et al [10] resulted in computational identification of a MOF predicted to be one of the top-performing structures for Xe/Kr separation. It followed this with experimental tests of the performance of the MOF. The MOF in question was SBMOF-1 (also a stand-out candidate in the Simon et al [187] study). It is predicted from the computational screening to have an incredible selectivity at infinite dilution (determined from Henry coefficients)

of 70.6. Infinite dilution selectivity calculated using the same metric but experimental isotherms was found to be 16, which, although a large drop from the computational prediction, was the highest observed selectivity for this concentration at the time. In another effort to use computational data to guide experimental study, Li et al [91] synthesised a squarate-based, polar hydroxyl-functionalised MOF targeted as having pores of perfect size based on guidance from previous screenings. [89] The authors refer to the MOF as MOF 1a in the dehydrated form, and as MOF 1 in the hydrated form, and it is referred to in the same way here. It has the best discoverable recorded experimental Xe/Kr selectivity of any MOF, the selectivity of MOF 1a being 69.7 at 1 bar as calculated from single-component isotherms using ideal adsorbed solution theory (IAST).

An overwhelming common thread among Xe/Kr separation studies, confirmed on a statistical scale by high-throughput screenings [89, 187] and used in MOF design by Li et al, [91] is the benefit of narrow channels within which Xe interactions with the framework may be maximised. If a pore is only very slightly larger than a Xe atom (it has a PLD slightly larger than 3.96 Å, [17] which is the kinetic diameter of Xe), Xe can experience strong interactions with the framework on multiple sides at once. Pore morphology is also shown to be important in this process. [89] Pore morphology may be described by the LCD/PLD ratio: a MOF with a low LCD/PLD ratio (close to 1) is mostly uniformly tubular, whereas a higher ratio indicates large cavities connected by narrow channels. This is illustrated in Figure 3.1, with SBMOF-1 as an example of a MOF with LCD/PLD close to 1, and a MOF from the CoRE MOF database [11] with identifying code LIPQIL, as an example of a MOF with much larger LCD than PLD. The LCD/PLD ratio of LIPQIL is 4.4. Sikora et al [89] determined that MOFs for which LCD/PLD lies between 1 and 2 are more likely to have the best Xe selectivity than any other LCD/PLD ratio. They also determined, ostensibly somewhat contrarily, that for Xe selectivity, spherical cavities are theoretically better than tubular ones. This need

not be at odds with their morphology observations, as true spheres would be impractical, having no way for guests to enter.

The particular chemical groups on the pore walls are also relevant, and different types of functionalisation have been shown to be effective Xe binding sites. In strongly adsorbing MOFs such as MOF 1a reported by Li et al, [91] polar OH groups form the binding sites. Meanwhile, SBMOF-1 is shown to adsorb Xe in the centre of a channel composed of aromatic rings on four sides, with a dense wall of chemical interactions. This aromatic channel is visible in Figure 3.1c. Here, the noble gases interact with the π cloud of the aromatic rings. Xe and Kr interaction with π clouds is also observed in formates, [196] notably $Co(HCOO)_2$ and $Ni(HCOO)_2$, [207] whose π faces orient into pores to allow guest interaction. Following their screening, Simon et al [187] carried out binding site analysis on the most promising identified structures. In the analysis, channels of aromatic rings, as in SBMOF-1, were observed as binding sites, along with rings containing benzonitrile monomers. The commonly observed small 2D channels are prevalent, although small 3D cages were also identified.

Based on this culmination of literature, it is judged that, when examining MOFs previously not considered for the Xe/Kr separation, small pores and tight channels comparable in size with an Xe atom stand out as promising geometrical features. Within this, channels populated by either polar groups or aromatic rings (or other groups featuring π clouds) display favourable functionalisation. A series of MOFs which bears resemblance to this description while also having the benefit of being previously experimentally synthesised, robust, and proven in other gas sorption situations is the MFM series (MFM 126-128 and MFM 136-138). These MOFs were synthesised by Humby et al [1] and tested for CO_2/CH_4 and CO_2/N_2 separations. The series is based on MFM-136, [208] a copper paddlewheel-based MOF, with ligands composed of amide, phenyl, pyrimidine, and isophthalate functionalities. For MFM-137 and MFM-138, the amide group of MFM-136 is replaced with an alkyne and

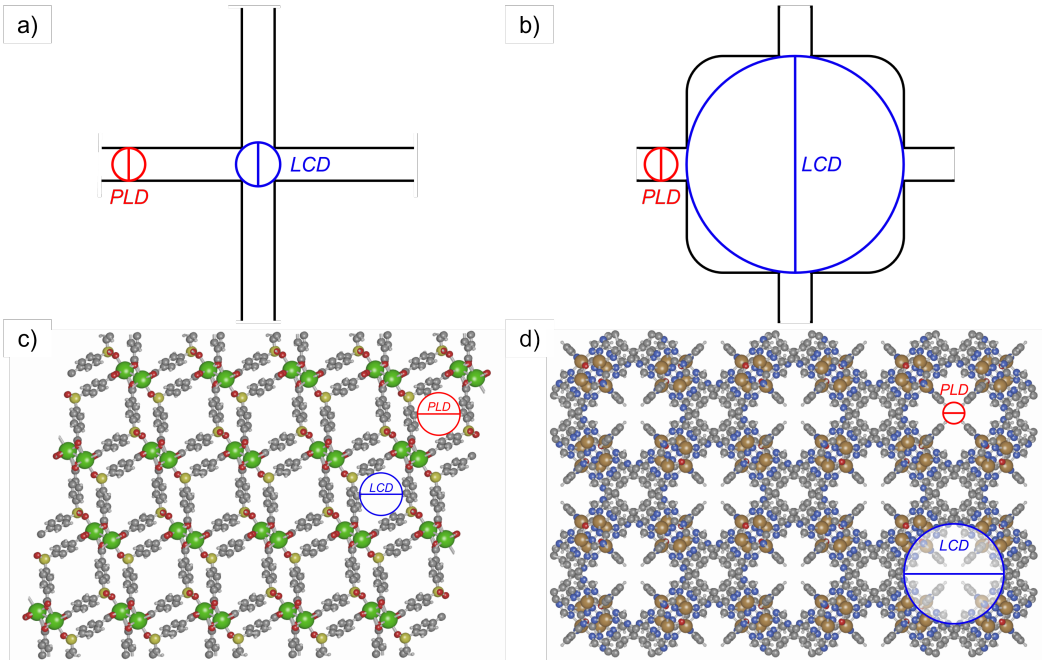

Figure 3.1: *Top: diagram of pore morphology showing pore limiting diameter (PLD) in red and largest cavity diameter (LCD) in blue: a) small value of LCD/PLD ratio; b) large value of LCD/PLD ratio. Bottom: The periodic structure of c) SBMOF-1 which performs well for Xe/Kr separation [10] (LCD/PLD = 1.36),[11] and d) LIPQIL MOF not known for good Xe/Kr separation performance (LCD/PLD = 4.4).*

a further phenyl ring, respectively. MFM 126-128 are analogous to MFM-136-138, but with the ligand shortened by removal of a phenyl ring. The series is illustrated in Figure 3.2. For CO_2 adsorption in MFM-136, binding sites defined by multiple interactions with aromatic rings have previously been shown to dominate. [208] Interactions with aromatic rings also contribute to guest binding in the rest of the series.

With these MOFs already established for gas sorption, identification of further separations which they may perform as multifunctional materials is desirable. It is seen that the MFM MOFs possess thin, tubular pore structure and multiple host-guest interactions involving aromatic rings. They might therefore be expected to be useful for the Xe/Kr separation. In this chapter, computational predictions of binary mixture gas uptake and heat of adsorption are predicted for this MOF family and performance is shown to be promising. This work therefore extends the range of potential applications of the experimentally accessible and robust MFM series. The series contains no open metal

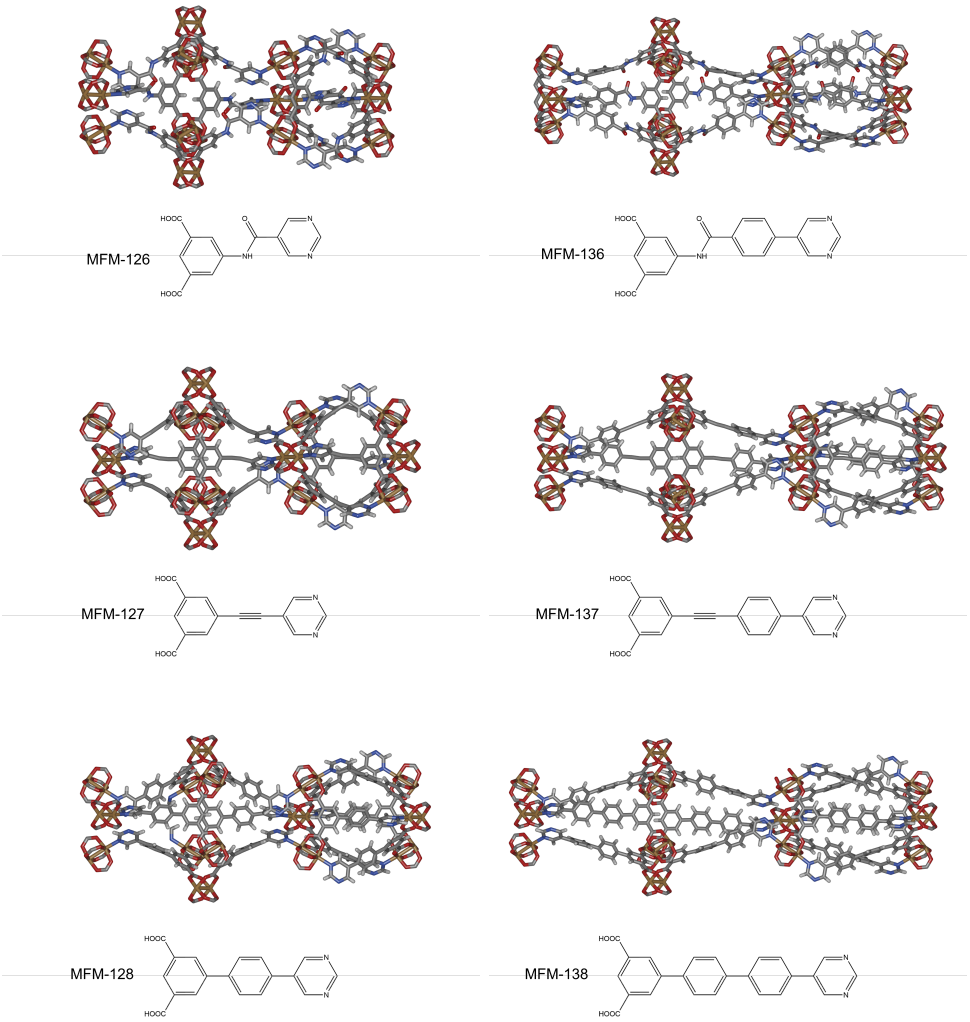

Figure 3.2: *Illustrations of the structures of the pores and ligands of the MFM MOFs family.*

sites: although their nodes are based on copper paddlewheels, in which copper atoms may commonly be undercoordinated, the copper atoms in these particular paddlewheels are capped by a part of the ligand framework, meaning the sites in this case are not open. This is somewhat analogous to cases in which capping solvent molecules such as water render copper paddlewheels no longer open. The lack of open metal sites is a positive feature in terms of water stability, which may be compromised by the presence of undercoordinated metal atoms.

3.1 Computational Setup used to Model Xe and Kr Uptake and Validation of its Applicability

In this chapter, a selection of geometrical and Monte Carlo methods have been employed to computationally calculate properties of the MFM MOFs. It is important to select methods which are suitable for the situations to which they are applied and which describe relevant properties of systems to a useful degree of accuracy. As discussed in section 2.1.2, the usefulness of a force field method depends on the fit of the parameters used to the system and conditions studied. This similarly relates to geometrical methods and their parameters, such as probe sizes. It is important to know exactly what the results of structural calculations describe and how they may be interpreted.

PLD, LCD, surface area, and volume are here calculated using the Zeo++ software package, [81] as described in section 2.1.1, with high accuracy settings used in all cases. For surface area calculations, a probe the size of a nitrogen atom (kinetic diameter 3.64 Å [209]) is used and both accessible surface area (ASA) and non-accessible surface area, NASA are calculated. The nitrogen-size probe facilitates comparison with experimental BET surface area values presented by Humby et al for the MFM MOFs and obtained from N_2 adsorption

isotherms. [1] Volume is presented in the form of both void fraction (necessary for uptake calculations, see section 2.1.1) and gravimetric accessible volume (in $cm^3\ g^{-1}$). Accessible volume was calculated using a zero-radius probe. Void fraction was calculated as probe-occupiable volume following the conclusions of Ongari et al [137] (see 2.1.1). Unlike in the study of Ongari et al, [137] a probe with the radius of He (1.3 Å [209]) was used for consistency with conventional He void fraction values. The void fraction presented is a total volume, the sum of accessible volume (AV) and non-accessible volume (NAV). This gives consistency with uptake simulations in which inaccessible pores are not blocked, and accounts for flexibility of the MOFs which may permit adsorption in supposedly non-accessible areas.

Turning to binding properties, heat of adsorption in the limit of infinite dilution was determined using Widom insertion [80] (see section 2.1.1). The RASPA software package [82] was employed for the task using 100,000 Monte Carlo cycles. Grand Canonical Monte Carlo (GCMC) methods, also using the RASPA software package, [82] were used to predict Xe and Kr uptake, and thus Xe selectivity, in the MOF family. For both GCMC and Widom insertion, the commonly applied [5, 50, 187] cost-reducing rigid approximation (see section 1.2.1) was used, which was judged to be appropriate to a useful degree of accuracy, as adsorption simulations are produced at 273 K and 298 K. This is unlikely to be appropriate at elevated temperatures, and even at room temperature appreciable flexibility is possible, as discussed in section 1.2.1.

Classical force fields considering Coulomb and van der Waals forces are applied as introduced in section 2.1.2. Interactions between unlike atoms are calculated using Lorentz-Berthelot mixing rules. [210] Potentials are truncated with a van der Waals cutoff of 12.0 Å. For gas uptake simulations, a total of 25,000 initialisation cycles and 25,000 production cycles were used. For framework atoms, van der Waals interactions are described using Lennard-Jones parameters from the Dreiding forcefield, [151] except in the case of copper, for

which Dreiding forcefield parameters are unavailable and parameters from the universal forcefield (UFF) [4] are used instead. Similar combination of UFF and Dreiding parameters has previously been used to good effect in terms of reproducing experimental data. [16, 211] Guest atoms are described using parameters taken from Hirschfelder et al (Xe) [2] and from Talu and Myers (Kr). [3] These parameters were designed for modelling transport properties of Xe and Kr.

Several of the force field parameters used have been designed or previously demonstrated to perform well under circumstances of adsorption and transport. It is useful to carry out further tests using systems as close as possible to those that are studied. Since there is no current experimental data for adsorption of Xe and Kr in the MFM series it is not yet possible to verify the method directly for the systems in question. Instead it is verified separately for the hosts and the guests. For the MFM MOFs, the available experimental data for gas uptake processes is CO_2 and CH_4 adsorption data. [1] Although the gases here are different from Xe and Kr, this is the only experimental data available to validate the choice of force field parameters for modelling the MOFs. Therefore, computational CH_4 and CO_2 uptake data for the MFM MOFs was calculated and compared against the experimental data. [1] For the guests, it was similarly necessary to select data of Xe and Kr adsorption in MOFs other than the MFM series. Therefore, computational Xe and Kr uptake data in MOF-505 and IRMOF-1 was calculated and compared to existing experimental data. [12–15] MOF-505, like the MFM MOFs, possesses a copper paddlewheel metal centre, while MOF-505, IRMOF-1, and all of the MFM MOFs have ligands composed of aromatic rings.

For adsorption of CO_2 and CH_4 in the MFM family of MOFs, the guests are described using the TraPPE forcefield, which is commonly used and well-validated. [152, 153] The combination of Dreiding and UFF framework parameters already mentioned was used, and initial CO_2 and CH_4 tests on MFM-136 revealed that these parameters performed slightly better than purely UFF

Table 3.1: Comparison of experimental data and computational predictions for CH_4 and CO_2 uptake at 1 bar and 298 K in the MFM MOF series, for validation of the force field parameters used to model the MFM MOFs. Experimental data is taken from Humby et al [1] and computational predictions are determined using single component GCMC simulations.

MOF	CH_4 uptake, mmol g^{-1}		CO_2 uptake, mmol g^{-1}	
	This work	Experiment[1]	This work	Experiment [1]
MFM-126	1.37	0.90	4.88	4.63
MFM-127	1.42	0.99	4.16	2.97
MFM-128	1.46	0.95	4.21	3.20
MFM-136	1.55	1.64	4.57	4.28
MFM-137	1.33	0.87	3.84	2.92
MFM-138	1.84	1.02	4.86	2.89

parameters. The CO_2 isotherms obtained for this method verification require consideration of Coulomb interactions, as CO_2 is modelled with partial charges. For these purposes, framework charges for each MFM MOF were calculated using the extended charge equilibration (eQeq) method (see section 2.1.2). [162] Long-range charges were computed using Ewald summation. Partial charges on CO_2 are defined within the TraPPE formalism. [153]

For this indirect validation, numerical experimental uptake data for CO_2 and CH_4 in all MFM MOFs at 1 bar and 298 K is taken from Humby et al, [1] and compared to equivalent computational uptakes in Table 3.1. Computational isotherms up to 1 bar at 298 K were also obtained, and data was extracted from the figures of the full experimental isotherms of Humby et al [1] for comparison where possible. The resulting computational and experimental isotherms are plotted alongside each other in Figure 3.3. MFM-126 and MFM-127 are omitted from these plots as it was not possible to obtain sufficient data from the figures of Humby et al for these MOFs. In general, the computational isotherms display similar shape and order to their experimental counterparts, and the magnitude of uptake, as well as the difference between CO_2 and CH_4 loading, is broadly comparable between computation and experiment, although with some notable variation. For both gases in MFM-136 and for CO_2 in MFM-126, computational prediction is close to the experimental

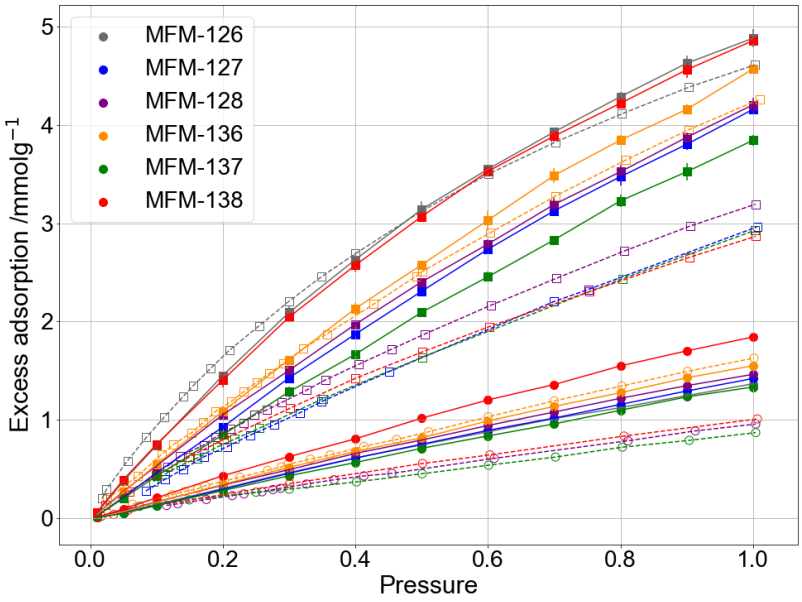

Figure 3.3: *Comparison of experimental and computational adsorption isotherms for CH_4 and CO_2 single component uptake in the MFM MOFs at 298 K. Circles: CO_2 uptake; squares: CH_4 uptake. Solid lines and filled points: computational uptake calculated using GCMC simulations; dashed lines and empty points: experimental uptake taken from the plots of Humby et al, [1] shown for MOFs for which data was obtainable (MFM-128, MFM-136, MFM-137, MFM-138). For the computational points, error bars shown are statistical uncertainties based on standard deviation in the GCMC process.*

values, differing by less than 10% at 1 bar and with similar isotherms. A particularly large difference is observed for MFM-138, which is computationally predicted to have particularly high CO_2 adsorption but was experimentally observed to display uptake properties more similar to the weaker adsorbers of the family. Differences for the remaining MOFs are largely consistent, falling between the MFM-136 (or MFM-126/CO_2) and MFM-138 differences. CH_4, with lower total uptake, displays generally larger relative bu smaller absolute differences than CO_2.

In general, computational values overpredict experimental uptakes, as seen in the reported numbers at 1 bar as well as the isotherms. Obtaining larger computational uptakes compared to experimental values is not uncommon for adsorption data and may occur for a variety of reasons. One possibility is the presence of defects in experimental structures which are overlooked by pristine computational structures. These can take the form of partially collapsed regions leading to inaccessibility of pores and adsorption sites, generally reducing uptake. They can also take the form of missing nodes, linkers, or fragments thereof. This second kind of defect can either enhance or reduce uptake, as missing fragments can amount to missing adsorption sites, reveal otherwise inaccessible sites, or have either positive or negative effect on geometrical compatibility of guests and hosts. A further possible explanation for discrepancy between computational and experimental uptake values is the inappropriate treatment by computational simulations of the effect of flexibility. The small pore sizes of the MFM MOFs combined with their proven CO_2 and CH_4 uptake (see subsequent section 3.2.1) indicate that they are likely to display flexibility, so this may play a part here. A MOF which is flexible under working conditions may have altered geometry and adsorption sites compared to the rigid structure which is modelled, having either a positive or a negative effect on gas uptake. A final possible contributor to the discrepancy is deficiencies in the modelling parameters, which may be unavoidable using the general force fields Dreiding and UFF. Perhaps the most common

deficiency in these general force fields is an inability to properly account for the particularly strong interactions between open metal sites and guest atoms. It is unlikely that this particular deficiency is seen here, as it would lead to an underprediction rather than an overprediction and the MFM MOFs (at least in their pristine form) do not contain open metal sites. Importantly, the MOF series follows broadly the same order of uptake and shape of isotherm computationally as experimentally, indicating good qualitative agreement. An exception is that MFM-138 has particularly low experimental uptake compared to computational prediction.

Computational isotherms for Xe and Kr adsorption in both IRMOF-1 and MOF-505 are shown in Figure 3.4. These may be compared to experimental and computational reference data which is also shown in Figure 3.4 as extracted from relevant literature figures. [12–16] The MOF-505 data is compared to experimental data from Bae et al [12]. Limited experimental data points were given by Bae et al, but it is clear from Figure 3.4 that the magnitude of the experimental uptake of both Xe and Kr at the pressures for which experimental data is available is well-reproduced using the selected computational models. The trends implied by the experimental data of a sharp loading increase between 0 and 1 bar, particularly for Xe, additionally appear to be reproduced in the pressure region for which data is available. The IRMOF-1 data is compared to experimental data from Mueller et al [13] and from Pawsey et al. [14, 15] The computational data reproduces the shape and magnitude of the experimental isotherm of Pawsey et al reasonably well, though with some deviation as pressure increases. In terms of the data of Mueller et al, [13], the computational model provides reasonable qualitative agreement, Xe experiencing a steep loading increase and subsequent levelling out over the pressure range considered while Kr experiences a slower and flatter increase. There is not good quantitative agreement, but the values presented are not of identical kind, with the experimental isotherm being given per litre of the container filled with the MOF sample, and the computational value being per litre of

framework. The manner of filling is likely to have affected the values: the volume used to determine the experimental value will inevitably include some amount of empty space not accounted for in the computational value. Under such circumstances, it would be expected that the two isotherms would display the same shape subject to some scaling constant determined by a ratio of the real volumes used. It may therefore be possible to address the quantitative disagreement. The true experimental framework volume is not known: for Xe, scaling was attempted by multiplying all computational values by the ratio of the highest-pressure experimental Xe loading to the highest-pressure computational Xe loading, at which point both isotherms have reached saturation. For Kr, neither isotherm reaches saturation in the pressure range considered, so the ratio by which all computational values were multiplied for scaling was different. Th extracted experimental loading whose pressure is closest to the highest-pressure computational loading was identified, and the ratio of this value to the highest-pressure computational value was used. The resulting isotherms are plotted in light green and light purple on the corresponding part of Figure 3.4. For both gases this results in very good quantitative agreement, confirming the similarity of the shapes of the isotherms and supporting the suggestion that the volume consideration is the cause of the original quantitative disagreement. The IRMOF-1 data is additionally compared to computational reference data extracted from Ryan et al [16], and reasonably good agreement is seen. As a result of the overall good agreement seen throughout the discussed comparisons, it was determined that the selected Xe and Kr parameters were appropriate.

These results may additionally be used to illustrate the importance of selecting appropriate force field parameters by comparison to equivalent isotherms obtained using less suited parameters. For example, isotherms obtained using parameters from the UFF [4] for the guests Xe and Kr in the hosts MOF-505 and IRMOF-1 are shown in Figure 3.5 along with the reference data which was included in Figure 3.4. Since the UFF is a general force field covering the

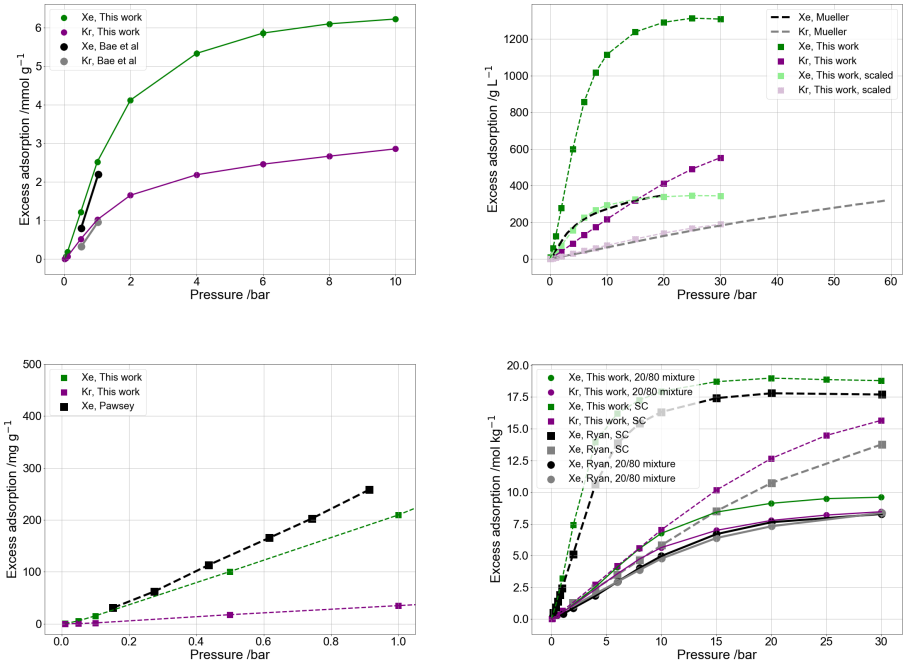

Figure 3.4: *Comparison of computational Xe and Kr adsorption isotherms in MOF-505 and IRMOF-1 calculated in this work with isotherms extracted from experimental and computational reference data. Top left: computational 20/80 mixture isotherms for MOF-505 at 298 K, along with experimental 298 K 20/80 mixture data extracted from Bae et al [12]. Top right: computational 298 K single component isotherms in IRMOF-1 with adsorption in g $L^{-1}_{framework}$ along with room temperature experimental data in g $L^{-1}_{container}$ extracted from Mueller et al. [13] Also shown in lighter colours are the computational isotherms scaled by the ratio of the experimental to the computational maximum, to address the difference in measurement units. Bottom left: computational single component isotherms in IRMOF-1 at 298 K along with experimental 292 K data from Pawsey et al [14] extracted from the figure of Greathouse et al [15]. Bottom right: Computational isotherms in IRMOF-1 at 273 K, along with computational reference data for IRMOF-1 at 273 K taken from Ryan et al [16]. Formatting throughout is as follows. Green: Xe uptake, this work; purple: Kr uptake, this work; black: Xe uptake, reference data; Grey: Kr uptake, reference data. Single component uptake: dashed lines and squares; 20/80 mixture uptake: solid lines and circles.*

Table 3.2: The Lennard Jones parameters used to model Xe and Kr in this study, [2, 3] along with UFF Lennard Jones parameters for Xe and Kr. [4]

		ϵ/k_b(K)	σ(Å)
UFF	Xe	167.04	3.92
	Kr	110.69	3.69
Hirschfelder [2]	Xe	221.0	4.10
Talu and Myers [3]	Kr	166.4	3.636

entire periodic table and aiming to reproduce molecular properties, its parameters are not tuned to transport properties of Xe and Kr in the gas phase. The isotherms obtained from UFF parameters are strikingly different to those obtained using parameters tuned to Xe and Kr transport properties. In general, the UFF isotherms both qualitatively and quantitatively match experimental data from Bae et al and Pawsey et al [12, 14, 15] less well than the isotherms obtained using the parameters from Hirschfelder et al (Xe) [2] and from Talu and Myers (Kr). [3] Meanwhile, the greater discrepancy between the scaled computational isotherms and the data of Mueller et al [13] shows that the shape of the UFF isotherms is a worse match in this regard also. Use of the UFF Lennard-Jones parameters for Xe and Kr leads to inability to well reproduce the initial sharp shoulder of adsorption at low pressures characteristic of Type I isotherms [212] and observed experimentally. [12–15] UFF parameters and the parameters employed in this study for Xe and Kr are shown in Table 3.2: the UFF ϵ parameters, which correspond to the Lennard-Jones well-depth (see section 2.1.2), are notably smaller than the Hirschfelder [2] or the Talu and Myers [3] ϵ parameters which are able to reproduce the Type I shoulder. Its reduction by the UFF parameters is a result of underestimating the strength of attractive interaction that the noble gas atoms experience.

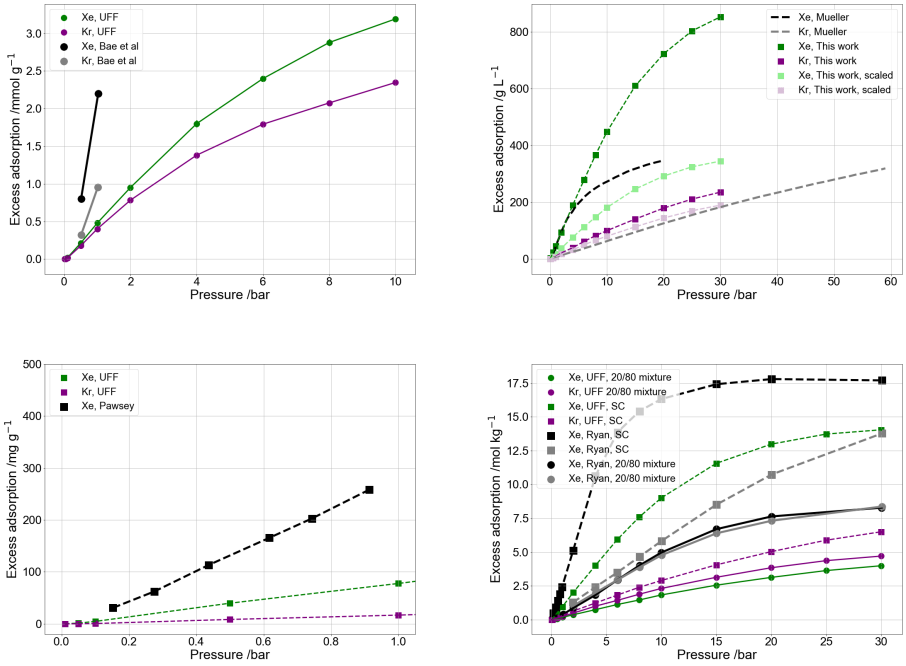

Figure 3.5: *As Figure 3.4, but with data obtained from this work, which is shown in green and purple, calculated using the UFF force field parameters for Xe and Kr.*

3.2 Results and Discussion

3.2.1 Structural Analysis

Inseparable from the study of adsorptive separation of Xe and Kr is the common conclusion among separation studies [31, 50, 89, 187, 194] of the strong links between structural properties of porous materials and their ability to separate Xe/Kr mixtures. It has already been mentioned that beneficial structural properties may optimise Xe-framework interactions, and such properties come in the form of narrow pores, low LCD/PLD ratios and aromatic or polar functionality on ligands.

Any new separation candidates must be considered in the context of their structural properties, and this consideration is made here for the MFM family. As discussed, the MFM MOFs appear to possess a number of beneficial features, being composed of 2D channels walled with aromatic rings and other functional groups which may be favourable for Xe adsorption. Computational structural data is here obtained for the MFM MOFs according to the methods

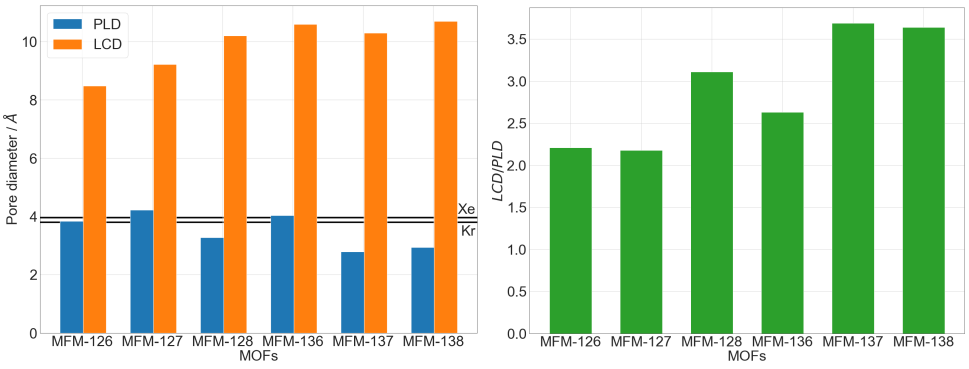

Figure 3.6: *Bar charts showing computational pore diameters of the MFM MOFs. Left: Pore limiting diameter (blue) and largest cavity diameter (orange), along with horizontal lines showing the kinetic diameters of Xe and Kr.[17]. Right: LCD/PLD ratio (green).*

laid out in sections 3.1 and 2.1.1.

PLD and LCD are calculated, along with the ratio LCD/PLD (Figure 3.6). Low PLD is observed, with LCD/PLD ratio which is fairly low, but not below 2: the series contains MOFs which possess some very narrow channels, where Xe adsorption may be achieved very effectively. Other parts of the channels are not as strictly narrow, potentially reducing the strength of Xe adsorption and allowing adsorption of Kr to compete in some areas of the MOF pores.

It is evident that four MOFs, MFM-126, MFM-128, MFM-137 and MFM-138, possess PLD smaller than the kinetic diameter of Xe. This does not prevent *simulation* of adsorption as long as at least part of the pore is large enough to admit the guest, as is the case here, since the GCMC simulations take no account of a guest's journey to an adsorption site. This also does not preclude Xe or Kr adsorption experimentally if the assumption is made that the MOFs display some flexibility which allows guests into the pores. This possibility is taken into account by the definition of void fraction as the sum AV + NAV (see section 3.1). All of the MOFs have previously been shown experimentally to adsorb CO_2 and CH_4 and N_2, [1] which are also larger than some of the smaller values of PLD. A similar effect has previously been observed and explained using MOF flexibility by Chen et al [31] for Xe in the organic cage CC3. In this case, the authors used molecular dynamics

simulations of the framework to define a pore-limiting envelope which took into account vibrational motion of the cage and was large enough to admit the relevant gases with windows 'open' for only a small fraction of the time.

Geometrical surface areas calculated using a N_2 radius probe are compared to published experimental N_2 Brunauer-Emmett-Teller (BET) [138] surface areas. [1] Both accessible surface area (ASA) and non-accessible surface area (NASA) were initially calculated, and are given in Table 3.3. In general, ASA with a N_2 probe is the surface area value directly comparable to experimental N_2 BET surface areas, as discussed in Chapter 1. However, several MFM MOFs possess zero ASA as calculated computationally. Since experimental surface area measurements (and CO_2/CH_4 adsorption) are possible, [1] it is clear that experimental samples of all of the MOFs do in fact possess area accessible to N_2. Again, it is likely that flexibility of the MOFs plays a part here, in allowing small-pored structures to take up gas molecules somewhat larger than their pores. However, it is not clear from the outset just how much ostensibly non-accessible area flexibility may reveal. Therefore, in Figure 3.7, two different values of computational surfacea area are plotted against experimental BET surface area. [1] These are ASA and total area, where toal area is the sum of ASA and NASA. For the majority of the MOFs (particularly those with zero ASA), total surface area values are very close to experimental values, while ASA alone does not capture the experimentally measured surface area well. In these cases, it is likely that the majority of area within the MOF, whether termed accessible or inaccessible for a rigid structure, may be accessed. This is with the exception of MFM-126, whose experimental surface area is much lower than its total surface area, as well as being much lower than the other surface areas of the series. Instead, its experimental surface area is closer to its ASA, and it is therefore likely that MFM-126 does not display sufficient flexibility under the conditions applied by Humby et al [1] to reveal the non-accessible area to an N_2 probe. It is alternatively possible that defects such as structural collapse reduced the available area in the experimental MFM-126

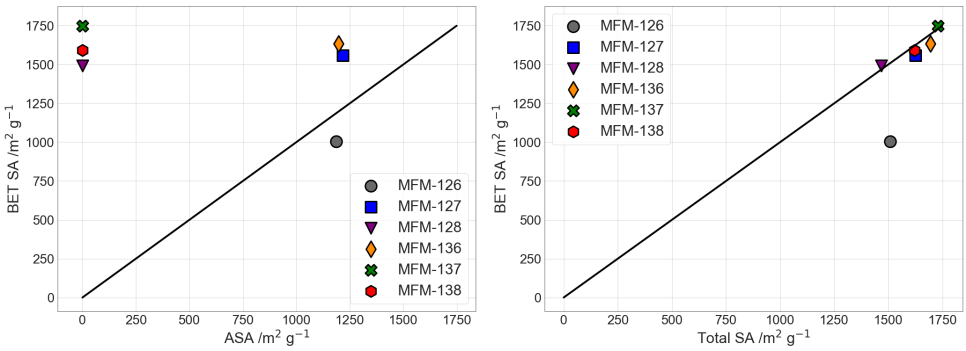

Figure 3.7: *Plot of computational geometrical surface areas (this work) against experimental BET surface areas [1] for the MFM MOFs. Left: computational values are accessible surface area. Right: computational values are total surface area. Also displayed in each plot is the line $y = x$ (black).*

Table 3.3: Accessible surface area (ASA), non-accessible surface area (NASA), total surface area and experimental BET surface area [1] of the MFM MOFs.

MOF MOF	ASA /m^2 g^{-1}	NASA /m^2 g^{-1}	Total SA / /m^2 g^{-1}	BET SA[1] /m^2 g^{-1}
MFM-126	1,186	323	1,509	1,004
MFM-127	1,216	408	1,624	1,557
MFM-128	0	1,467	1,467	1,491
MFM-136	1,197	498	1,695	1,634
MFM-137	0	1,729	1,729	1,749
MFM-138	0	1,620	1,620	1,590

sample.

High surface area in a MOF is a celebrated property for storage applications, and can be instrumental in maximising uptake capacities. Increasing surface area up to an optimal point (2,500-3,000 m^2 g^{-1} for methane) [33] maximises storage capacity, before capacity begins to reduce as surface area increases further. For separation applications, however, as the metric selectivity becomes important, surface area becomes less decisive. Where it does affect selectivity, surface area commonly leads to a decrease, not an increase. [213] For practical separation applications, while high selectivity is clearly desirable, it is beneficial for this to coexist with high capacity. This can allow the separation process, which can involve multiple adsorption and desorption events, to proceed efficiently. The competing influence of surface area here, in

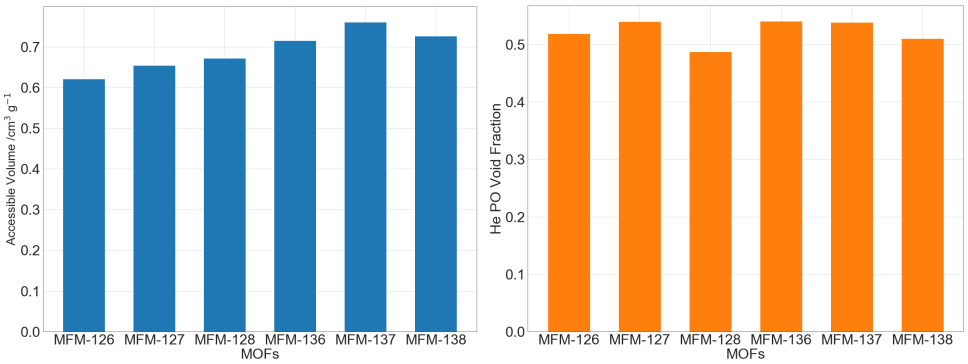

Figure 3.8: *Bar charts showing volume of the MFM MOFs. Left: accessible volume calculated using a zero-radius probe (blue). Right: total probe-occupiable void fraction calculated using a helium-radius probe.*

increasing capacity while decreasing selectivity can pose a problem. Indeed, a trade-off relationship is often seen between uptake and selectivity [214, 215] as surface area and other structural features display this contrasting effect.

High volume in a MOF can allow high uptake at high pressure, although at low pressure can go hand in hand with larger pores and reduced surface interactions, reducing uptake. The larger pores often associated with high volume can also be detrimental to selectivity, although the effect on behaviour of the interplay between different structural properties is case dependent. Geometrical accessible volumes, calculated using Monte Carlo sampling of a probe on a Voronoi network, are shown in Figure 3.8. A zero radius probe is used, so all of the volume is accessible. Volume is larger for those MOFs in the series which contain longer ligands. Of particular note is the volume of MFM-137, the highest in the series.

Figure 3.8 also shows total probe-occupiable helium void fraction for the series. This is the sum of probe-occupiable AV and probe-occupiable NAV using a He-sized probe. These are the void fraction values which are used to calculate excess uptake in subsequent loading calculations. The presented uptakes were compared to uptakes resulting from force-field based void fraction methods, yielding negligible difference.

3.2.2 Xe and Kr Uptake, Selectivity and Binding Affinity

Adsorption isotherms, heats of adsorption (Q_{st}), and Henry constants (K_H) were calculated for the MFM family of MOFs using Grand Canonical Monte Carlo (GCMC) and Widom insertion [80] methods available in the RASPA software package [82] as described in section 3.1.

Xe and Kr uptake isotherms up to 10 bar at 273 K and 298 K are shown in Figure 3.9. Isotherms obtained from single component simulations are displayed, along with those from binary mixture simulations with Xe/Kr ratios of both 50/50 and 20/80. The 20/80 ratio was selected for its relevance to industrial separations, and the 50/50 ratio for ease of comparison to other selectivity measures. Further gas ratios could also be instructive to consider, such as the 10:1 Kr:Xe ratio of UNF, but the 50/50 and 20/80 ratios are judged to be sufficiently representative for the purposes of this work. Xenon is adsorbed more readily than krypton for the whole series. When guests are allowed to compete for binding sites, a higher uptake of Xe than Kr is observed at all pressures considered. This remains the case even with significantly more Kr than Xe in the gas mixture (the 20/80 case). The single component isotherms show Kr adsorption overtaking Xe at sufficiently high pressure and low temperature. Here, saturation is approached and the limitations of volume become more relevant, the smaller size of Kr allowing higher molar quantities to be adsorbed. The difference in behaviour between the two temperatures considered is familiar, uptake increasing as guest atoms lose kinetic energy and interactions with pore walls dominate.

Using binary mixture isotherms, preferential adsorption can be quantified by the metric selectivity, $S_{i/j}$, calculated as [216, 217]

$$S_{i/j} = \frac{q_i y_j}{q_j y_i} \tag{3.1}$$

where q_i is the quantity of component i in the adsorbed phase and y_i is the

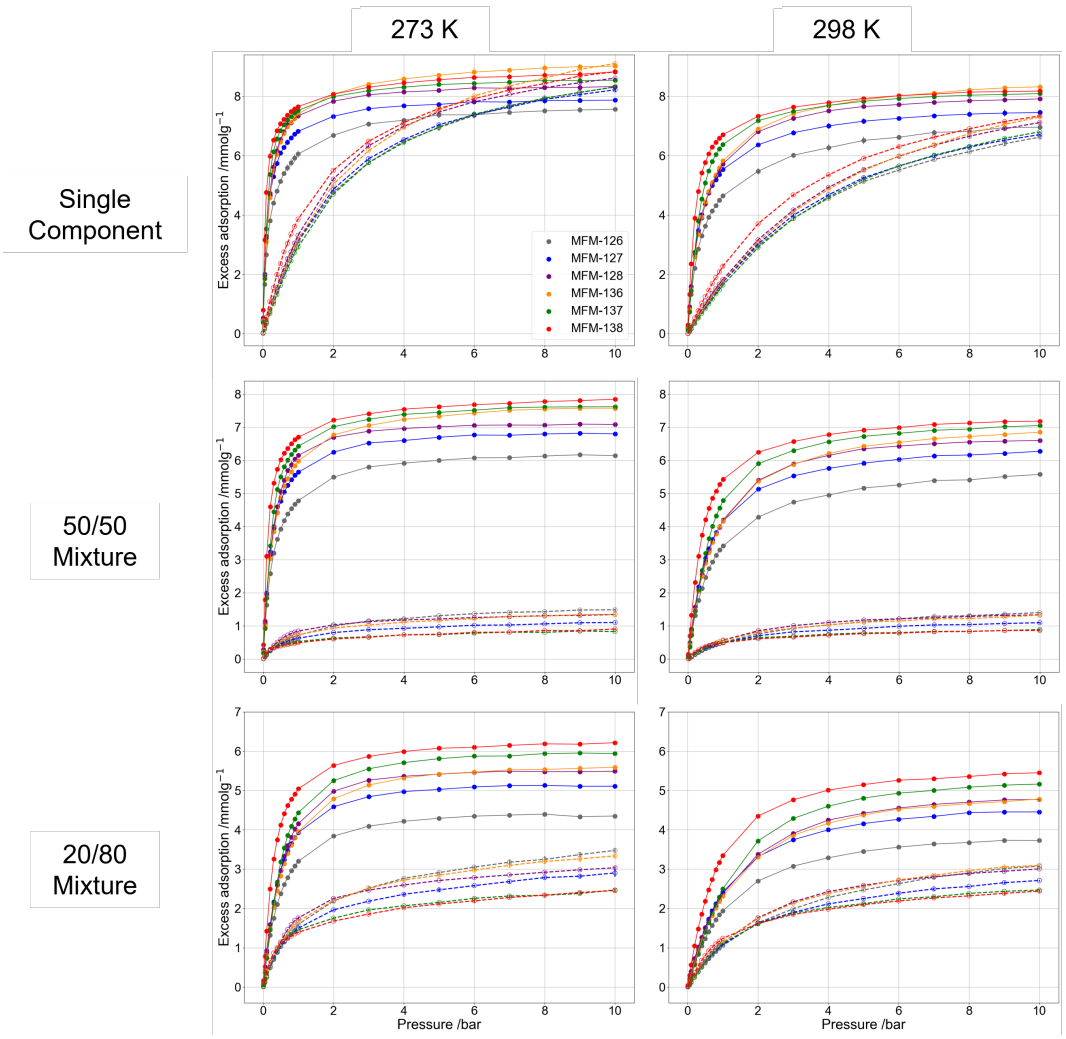

Figure 3.9: *Xe and Kr adsorption isotherms of the MFM MOFs calculated up to 10 bar pressure at 273 K (left) and 298 K (right). Top: single component adsorption isotherms; centre: 50/50 mixture adsorption isotherms; bottom: 20/80 mixture adsorption isotherms. Xe uptake: closed circles, solid lines; Kr uptake: open circles, dashed lines. Error bars are based on statistical uncertainties from standard deviation in Monte Carlo production runs (section 2.3.1).*

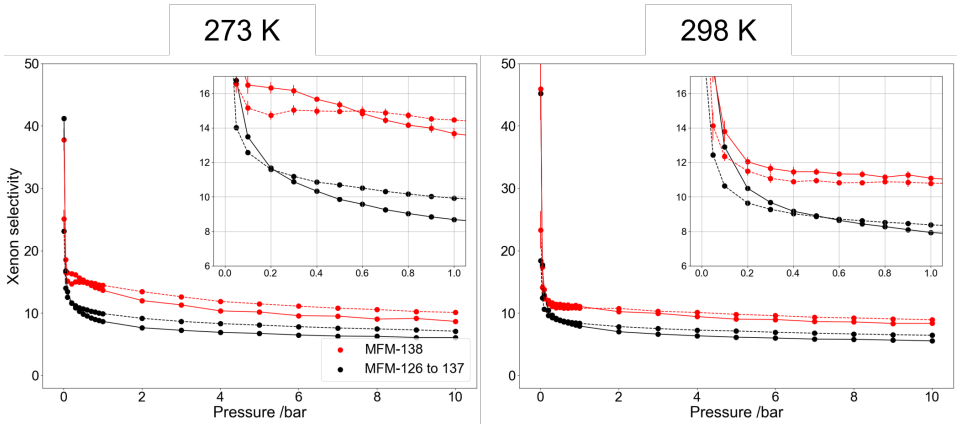

Figure 3.10: *Xe selectivity of MFM-138 (red), and average Xe selectivity of the remaining MFM MOFs (black) calculated up to 10 bar pressure at 273 K (left) and 298 K (right) Dashed line: 20/80 mixture; Solid line: 50/50 mixture. Error bars for MFM-138 selectivity based on statistical uncertainties from standard deviation in the Monte Carlo production runs (section 2.3.1) are given.*

mole fraction of component i in the gas phase. Xenon selectivity for each MFM MOF was calculated using the 20/80 and the 50/50 Xe/Kr mixture. In each case, MFM-138 displays the highest or very close to the highest Xe selectivity over the full pressure range, while MFM-137 is its closest contender, with very similar or slightly higher selectivity at high pressure. At 273 K and 298 K, the Xe selectivity of MFM-138 in equilibrium with both gas mixtures is plotted against pressure in Figure 3.10. Figure 3.10 also includes the average selectivity of the other five MFM MOFs. Individual selectivity plots for the whole series are given in Figure 3.11. All structures display higher selectivity at very low pressure, dropping rapidly before 1 bar is reached, and levelling off as pressure further increases. MFM-137 and MFM-138 experience a selectivity minimum between 0 and 1 bar in some cases.

To illustrate the range of selectivity values observed, the two extreme pressure points, 0.01 bar and 10 bar, are highlighted at both temperatures in Table 3.4 for the 50/50 gas mixture and in Table 3.5 for the 20/80 gas mixture. The highest and the lowest selectivity displayed by an individual MOF in the series in each case are given. These values do not compete with the best literature benchmarks [10, 91] for the highest observed or predicted Xe/Kr selectivity of a material. However, they show high selectivity at low pressure for the whole

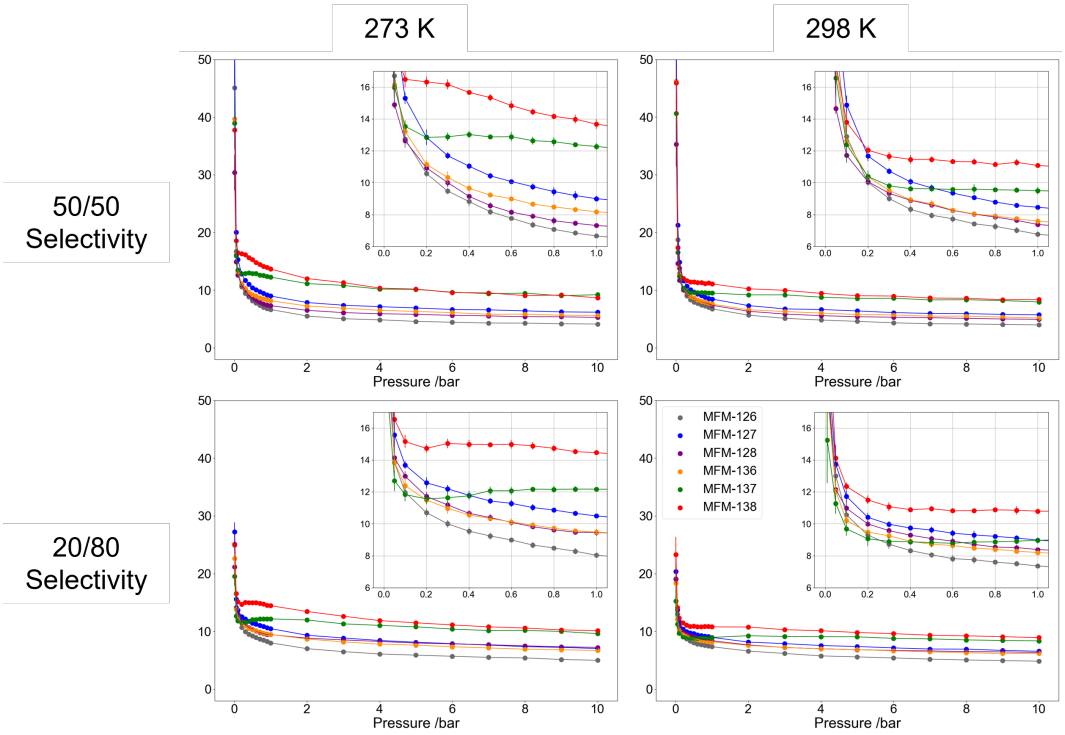

Figure 3.11: *Xe selectivity of all of the MFM MOFs calculated up to 10 bar pressure. Top left: a 50/50 mixture at 273 K; Top right: a 50/50 mixture at 298 K; Bottom left: a 20/80 mixture at 273 K; Bottom right: a 20/80 mixture at 298 K. Error bars are based on statistical uncertainties from standard deviation in the Monte Carlo production runs (section 2.3.1).*

series, and appreciable and practical behaviour at higher pressures. This may be added to the already established separation applications of the family. [1] In particular, MFM-138 stands out for the best Xe-selective behaviour predicted. On the other hand, MFM-126 has low Xe uptake compared to the rest of the series, translating to low selectivity, being the least selective MOF at 10 bar for every case.

In addition to selectivity, it is instructive to consider the magnitude of guest uptake. Although this metric is not as important for separation applications as for storage, it can be decisive in determining the efficiency of a separation and whether use of a given framework is practically viable. Total Xe uptake in the MFM series is predicted to compare well with many previously identified promising materials. From single component simulations (Figure 3.9), Xe uptake at saturation is predicted to be between 7.56 and 9.01 mmol g^{-1} at 273 K, and between 6.96 and 8.32 mmol g^{-1} at 298 K. This range competes well with

Table 3.4: The highest and the lowest values of computational Xe/Kr selectivity for a 50/50 mixture in the MFM MOF series at the temperature points T = 273 K and T = 298 K.

	0.01 bar		10 bar	
	$S_{Xe/Kr}$	MOF	$S_{Xe/Kr}$	MOF
T = 273 K				
highest $S_{Xe/Kr}$	51.98 ± 6	MFM-127	9.22 ± 0.2	MFM-137
lowest $S_{Xe/Kr}$	30.39 ± 3	MFM-128	4.14 ± 0.1	MFM-126
T = 298 K				
highest $S_{Xe/Kr}$	52.16 ± 10	MFM-127	8.37 ± 0.08	MFM-138
lowest $S_{Xe/Kr}$	35.29 ± 4	MFM-128	3.99 ± 0.07	MFM-126

Table 3.5: The highest and the lowest values of computational Xe/Kr selectivity for a 20/80 mixture in the MFM MOF series at the temperature points T = 273 K and T = 298 K.

	0.01 bar		10 bar	
	$S_{Xe/Kr}$	MOF	$S_{Xe/Kr}$	MOF
T = 273 K				
highest $S_{Xe/Kr}$	27.26 ± 2	MFM-127	10.12 ± 0.2	MFM-138
lowest $S_{Xe/Kr}$	19.56 ± 2	MFM-137	5.01 ± 0.08	MFM-126
T = 298 K				
highest $S_{Xe/Kr}$	23.33 ± 3	MFM-138	8.93 ± 0.1	MFM-138
lowest $S_{Xe/Kr}$	15.24 ± 3	MFM-137	4.85 ± 0.06	MFM-126

Xe uptakes of previous relevant MOFs, including those recently identified as having exceptional selectivities. [10, 91] Of the two MOFs with published selectivity values close to 70, the first, SBMOF-1, has an experimental Xe uptake around 1.4 mmol g^{-1} at 298 K and 1 bar, approaching saturation. [10] Under the same conditions, the second, the squarate-based MOF 1a published by Li et al [91] has a reported Xe uptake of 66.1 cm^3 cm^{-3}, which is equivalent to 1.35 mmol g^{-1}. In general, high literature Xe uptakes [50] are between 4 and 6 mmol g^{-1}. Note, however, that computational values may be often behave as an upper bound for experimental performance, and that the predicted binary mixture uptakes for the MFM MOFs are reduced compared to the single component values. Although in practical separation applications Xe uptake is likely to be lower than the computational single component uptakes given, the MFM MOF series competes very well with leading structures for total Xe uptake.

For a separation application, uniting this high uptake with high selectivity is desired. It has already been established that the MFM series displays Xe/Kr selectivity, and to probe further the relationship between uptake and selectivity within the family, uptake at 10 bar is plotted against selectivity at 10 bar for the two gas compositions in each of the MOFs at 273 K and 298 K in Figure 3.12. The relationship among the series is similar regardless of the conditions selected. The trade-off relationship which may often be seen between uptake and selectivity [214, 215] is not observed here, MFM-138 consistently displaying both the highest uptake among the series and the highest selectivity by a clear margin and MFM-126 consistently displaying the lowest performance by both metrics. Among the remaining MOFs, a general, but not strict, increase in selectivity with uptake is seen. The notable exception to this trend is MFM-136, which displays particularly high uptake but low selectivity among the series.

Consideration of the magnitude of guest uptake extends not only to total available capacity but to working capacity. As discussed in chapter 1, common

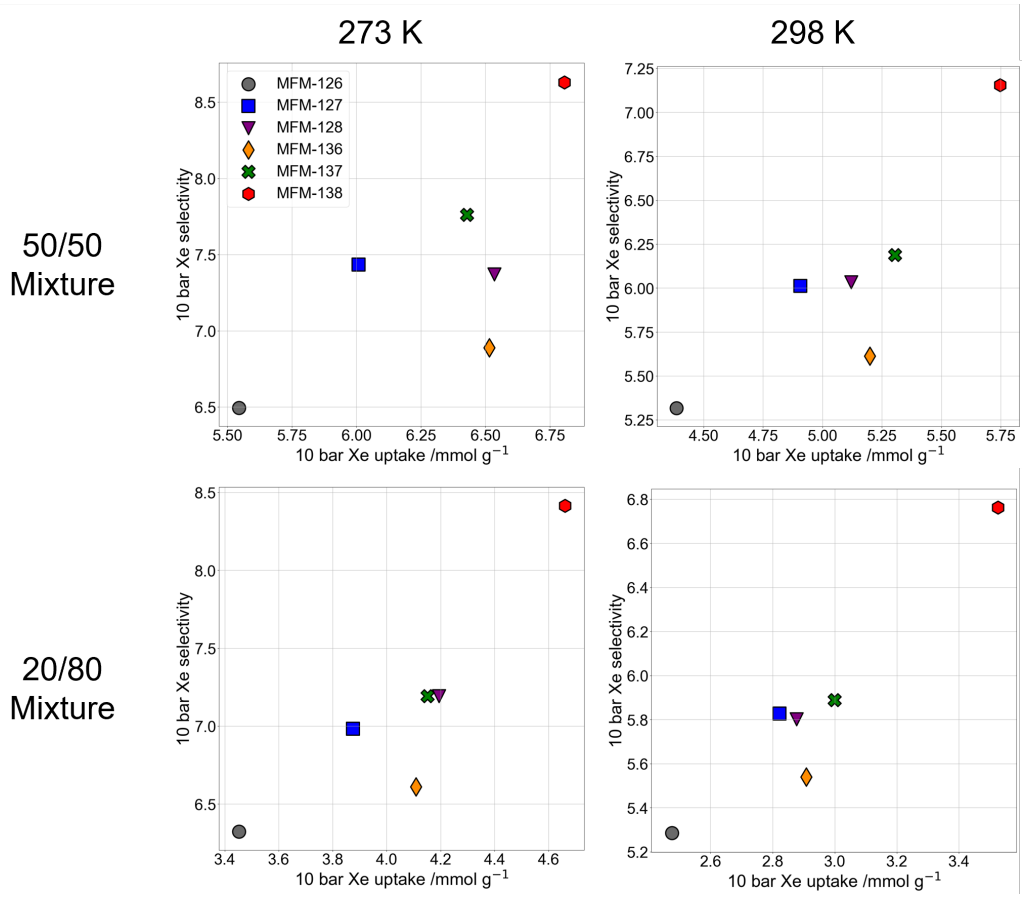

Figure 3.12: *Plots of uptake of Xe at 10 bar against Xe selectivity at 10 bar as predicted by GCMC simulations for each of the MFM MOFs under all sets of conditions studied: a 50/50 Xe/Kr mixture at 273 K (top left), a 50/50 Xe/Kr mixture at 298 K (top right), a 20/80 Xe/Kr mixture at 273 K (bottom left) and a 20/80 Xe/Kr mixture at 298 K (bottom right).*

separation processes such as PSA and TSA which rely on the swing of a condition involve an adsorption step under some condition (e.g. high pressure or low temperature) and a desorption step under another (e.g. low pressure or high temperature). Considering only total capacity assumes that the desorption step fully evacuates the MOF, but practically this is not the case. With PSA as an example, gas separation processes are carried out at a set adsorption and desorption pressure depending on the system at hand. Typically, adsorption steps are at or above 10 bar [218, 219] (though lower pressures are possible [220]) and desportion steps are close to 1 bar. [88, 220] It is also possible to use very low pressures close to zero (less than 0.01 bar) for desorption in the case of pressure vacuum swing adsorption (PVSA). [218] Working capacity and its impact on choices about operating conditions has often not been directly considered in searches for promising Xe/Kr separation materials. For the MFM MOFs, working capacities based on selected adsorption and desorption pressures are given in Table 3.6. The adsorption pressure used is 10 bar, and desorption pressures of both 1 bar, reflecting standard PSA, and 0.1 bar, reflecting PVSA, are used. The magnitude of the working capacities reflects the shape of the isotherms in Figure 3.9, particularly at 273 K. The isotherms have reached their plateau by 10 bar and approach maximum loading by around 2 bar: the MFM MOFs display no problems in reaching the maximum possible loading for the adsorption step. However, with a sharp loading increase in the low-pressure range, a very low desorption pressure reflecting PVSA conditions is necessary to reach the near-zero desorption loading which allows high working capacity. Working capacity decreases rapidly as desorption pressure increases, and an intermediate value of desorption pressure leads to predicted working capacities which are fairly low, but not entirely unusable. Isotherms which reach high total loading at an easily accessible elevated pressure but remain at lower loading over a larger part of the low-pressure range are preferable over very steep Type I isotherms from a practical perspective to allow for a more easily accessible desorption pressure. For the MFM MOFs, the isotherms

Table 3.6: Working capacity of each of the MFM MOFs with 50/50 and 20/80 Xe/Kr gas mixtures at 273 K and 298 K. An adsorption pressure of 10 bar and two different desorption pressures are used. One desorption pressure corresponds to a standard PSA situation (labelled PSA, desorption pressure of 1 bar) and the other to a PVSA situation (labelled PVSA, desorption pressure of 0.1 bar).

	MOF	Working Capacity /mol kg $^{-1}$			
		50/50		20/80	
		273 K	298 K	273 K	298 K
PSA	MFM-126	1.36	2.16	1.15	1.79
	MFM-127	1.16	2.10	1.18	2.04
	MFM-128	0.93	2.40	1.33	2.40
	MFM-136	1.59	2.68	1.63	2.47
	MFM-137	1.19	2.26	1.50	2.66
	MFM-138	1.15	1.76	1.17	2.11
PVSA	MFM-126	5.91	5.50	4.28	3.71
	MFM-127	6.59	6.21	5.04	4.44
	MFM-128	6.80	6.50	5.38	4.74
	MFM-136	7.36	6.78	5.52	4.77
	MFM-137	7.44	7.00	5.89	5.15
	MFM-138	7.42	7.04	6.05	5.41

at 298 K display a shallower loading increase over the low-pressure range than the 273 K isotherms (Figure 3.9), and, as seen in Table 3.6, throughout the standard PSA conditions case, working capacity is higher at 298 K than at 273 K, whereas when using PVSA conditions, the higher total loading at 273 K allows higher working capacity. The results indicate that room temperature conditions are more optimal than reduced temperature for the Xe/Kr separation by the MFM MOFs if PVSA is not attempted, but that PVSA is expected to lead to a much more efficient separation than standard PSA. Among the MFM series, MFM-138 remains the best or very close to the best performer as measured by working capactiy under PVSA conditions, but under PSA conditions its performance compares less favourably to the rest of the series, as its higher desorption loading counteracts its higher adsorption loading.

Examining both single component and binary mixture data, insight is available into the usefulness of single component isotherms as a cheaper approximation for the behaviour of porous materials and gas mixtures. The single

component approximation is particularly pertinent for experimental studies, where mixture isotherms are less accessible. The single component isotherms calculated here accurately predict higher xenon than krypton uptake at low to medium pressures, and thus that the MOFs are xenon-selective. They additionally prove to be a good qualitative approximation for uptake of xenon, the dominant guest, in a 50/50 mixture. For Xe, the single component and the 50/50 isotherms follow a very similar shape, and the single component isotherms overestimate total Xe uptake by only around 1 mmol g^{-1} compared to the 50/50 mixture data. From the 50/50 mixture simulations, it is apparent that a large amount of krypton adsorption that would occur in equilibrium with pure krypton gas is prevented by the presence of xenon. Krypton adsorption in the 50/50 mixture fails to exceed 1.5 mmol g^{-1}, and single component krypton adsorption exceeds this by 3-5 times. The unequal 20/80 mixture isotherms can be less directly compared to single component results, though for both gases they too follow a broadly similar shape to the single component isotherms. Both sets of mixture isotherms predict markedly lower total uptake than the simple sum of xenon and krypton single component uptake, particularly the 20/80 mixture in which the less readily adsorbed krypton is the more prevalent element in the gas phase. This is a natural consequence of the competition of the two species for the same adsorption sites, but could not be quantified from single component isotherms. The single component approximations are thus demonstrated to have qualitative predictive power for binary mixture calculations for these two gases before saturation is reached, but any quantitative predictions made from them are limited.

Heat of adsorption, Q_{st}, is a useful measure of strength of binding between host and guest. In the limit of infinite dilution, heat of adsorption can be determined using Widom insertion. [80] Considering each gas individually, infinite dilution Q_{st} values (Table 3.7) do not vary very widely among the members of the MFM series, with the highest heat of adsorption for xenon being 28.40 ± 0.1 kJ mol^{-1} for MFM-138. This fits in favourably with typical literature values

Table 3.7: Computational heat of adsorption, Q_{st}, values calculated at infinite dilution for Xe and Kr at 298 K using Widom insertion.

MOF	Q_{st}(Xe) kJ mol^{-1}	Q_{st}(Kr) kJ mol^{-1}
MFM-126	26.31 ± 0.06	18.95 ± 0.02
MFM-127	25.29 ± 0.02	18.22 ± 0.2
MFM-128	26.44 ± 0.01	19.03 ± 0.04
MFM-136	25.74 ± 0.1	18.62 ± 0.1
MFM-137	25.08 ± 0.01	18.54 ± 0.1
MFM-138	28.40 ± 0.1	20.34 ± 0.01

of the heat of adsorption for Xe in other materials, although values above 30 kJ mol^{-1} have been observed for materials with particular Xe affinity.[50, 91, 221] It is thus clear that the MOFs in the series contain favourable binding sites for xenon. Heat of adsorption values for Kr are markedly lower, in line with the observed low pressure selectivity. Although comparatively low, these remain fairly high in absolute terms showing that the adsorption sites in the series also possess appreciable affinity towards krypton.

A second measure of binding affinity, the Henry constant for a host/guest system is equal to the gradient of the adsorption isotherm at infinite dilution, as shown by Equation 3.2, where X is the gas uptake, P is pressure, and K_H is the Henry constant:

$$X = K_H P \tag{3.2}$$

For two gases i and j, the ratio of the Henry constants $K_H(i)/K_H(j)$ can be used as a somewhat limited measure of selectivity for gas i. Henry constants and infinite dilution Henry selectivities were calculated for each MOF using Widom insertion methods at 298 K and are displayed in Table 3.8. The selectivity values may be compared to those obtained using binary mixture isotherms, which are displayed for comparison at 298 K in Table 3.9. At low loading, a generally similar trend is followed between Henry and 20/80 mixture selectivities, with MFM-138 standing out as having notably the largest

Table 3.8: Computational Henry constants and Henry Xe selectivity at infinite dilution at 298 K of the MFM MOFs calculated using Widom insertion methods.

MOF	K_H(Xe) mol kg^{-1} Pa^{-1} $\times 10^{-5}$	K_H(Kr) mol kg^{-1} Pa^{-1} $\times 10^{-5}$	K_H(Xe)/K_H(Kr)
MFM-126	22.2 ± 0.1	2.23 ± 0.009	9.96 ± 0.0008
MFM-127	20.8 ± 0.04	2.08 ± 0.006	9.98 ± 0.0003
MFM-128	24.2 ± 0.1	2.27 ± 0.005	10.66 ± 0.0005
MFM-136	19.9 ± 0.1	2.15 ± 0.01	9.34 ± 0.001
MFM-137	16.5 ± 0.1	1.95 ± 0.01	8.46 ± 0.002
MFM-138	36.0 ± 0.1	3.08 ± 0.02	11.68 ± 0.0008

selectivity in both cases. The remaining MOFs display only small variation, reducing in Henry selectivity in the order MFM-128 > MFM-127 ∼ MFM-126 > MFM-136 > MFM-137, and reducing in 20/80 mixture selectivity in the order MFM-127 > MFM-126 > MFM-128 > MFM-136 > MFM-137. The values of Henry selectivity underestimate those of 20/80 mixture selectivity, but reproduce the trends well, having only a small discrepancy in the order of the MOFs; the presence of some discrepancy is unsurprising over such a narrow range. At 0.01 bar, both Henry selectivity and 20/80 mixture selectivity are significantly lower than 50/50 mixture selectivity. For a 50/50 gas mixture at very low pressure, the amount of krypton adsorbed is very low, which drives selectivity up. A different order is also observed, being MFM-127 > MFM-126 > MFM-136 > MFM-138 > MFM-137 > MFM-128. The very low krypton uptakes observed lead to high errors in selectivity for the 50/50 mixture at 0.01 bar. However, even if the selectivity values were the lowest allowed within their errors, 50/50 low pressure selectivities would be notably higher than infinite dilution Henry selectivities and indeed 20/80 low pressure selectivities.

As pressure increases, the effect causing unusually high selectivity for a 50/50 mixture is no longer seen. Table 3.9 shows this result for both compositions of binary mixture at 10 bar, with MFM-138 having the highest selectivity. For the remaining MOFs at 10 bar, the order of selectivity values is altered compared to low pressure selectivity for both 50/50 and 20/80 mixtures. The

Table 3.9: Computational selectivity values for the MFM MOFs with a 50/50 and 20/80 binary Xe/Kr mixture at 298 K and pressures of 0.01 bar and 10 bar calculated from GCMC simulations.

MOF	$S_{Xe/Kr}$			
	50/50		20/80	
	0.01 bar	10 bar	0.01 bar	10 bar
MFM-126	51.65 ± 9	3.99 ± 0.07	19.15 ± 3	4.85 ± 0.06
MFM-127	52.16 ± 10	5.73 ± 0.1	20.36 ± 2	6.57 ± 0.06
MFM-128	35.29 ± 4	4.96 ± 0.05	19.01 ± 0.9	6.35 ± 0.05
MFM-136	46.21 ± 7	5.19 ± 0.08	18.34 ± 3	6.19 ± 0.03
MFM-137	40.64 ± 2	7.96 ± 0.1	15.24 ± 3	8.37 ± 0.09
MFM-138	45.91 ± 5	8.37 ± 0.08	23.33 ± 3	8.93 ± 0.1

resulting order of selectivity, however, is very similar for the two gas mixtures. Most notably, in both cases MFM-137 goes from being the least or almost the least selective MOF at low pressure to the second most selective at higher pressure. At higher pressure, available volume becomes more important and MFM-137, with the highest available volume, becomes relatively more selective.

To this point, adsorption in the MFM MOFs at commonly studied temperatures applicable to the majority of relevant use cases has been considered. Under select circumstances, it may be desired to pursue gas separations under extreme conditions, an example being to make a process amenable to the high operating temperatures of an MSR, which can reach 600-800 °C. [52] Such an aim is not without significant challenges, [222] not least that stability, uptake and often selectivity of MOFs degrade with temperature. The idea is impossible in the case of TSA, as operation at only elevated temperatures is antithetical to the basis of the process. However, PSA and membrane separations, and related methods, may be carried out at high temperature using the correct materials and circumstances. [223–226] No MOF has been discovered with thermal stability reaching 800 °C, but the usefulness of the MFM MOFs at some intermediate temperature is not inconceivable. With consideration for cases in which chemists may wish to conduct Xe/Kr separations under

extreme conditions, an initial approximate quantification of selectivity trends in the MFM series based on Henry constants is provided below. However, quantifying adsorption properties at elevated temperatures beyond a basic approximation is an involved task which is not attempted here.

Henry selectivity is examined as temperature increases above the two temperature points for which adsorption isotherms were obtained. The ratio of Xe to Kr Henry constants at infinite dilution is plotted in Figure 3.13 for the MFM MOFs over a range of elevated temperatures up to 900 K. As temperature increases, the stability of the MOFs is not guaranteed, and at the highest temperatures used here it is very unlikely. However, temperatures up to 900 K are included to demonstrate the predicted trends in behaviour as temperature approaches this value. Additionally, inaccuracies in predictions associated with the rigid framework approximation are likely to become more pronounced with increasing temperature. Full GCMC simulations without extra consideration of framework flexibility are unlikely to be accurate at extreme temperatures and have not been attempted above 298 K, but the qualitative theoretical trend in affinity, as measured by Henry constants, can still be instructive. Among the series, selectivity is shown to follow a very similar trend, reducing with increasing temperature, as the affinity of both guests for the MOF decreases. Greater variation among the series is displayed at low and ambient temperatures than at high temperatures. At low temperature MFM-138 has the highest Henry selectivity by some margin, and the variation decreases as temperature becomes higher. The drop levels out to a large extent over the temperature range relevant to MSR product processing.[52] By the highest temperature modelled, Henry constants predict a selectivity below two. The MOFs remain theoretically Xe-selective at all temperatures considered, with Henry selectivity at no point dropping below one, though above around 400 K selectivity is notably low throughout the series and unlikely to have significant practical use.

In practical sorption-based MOF applications, there are relevant thermal

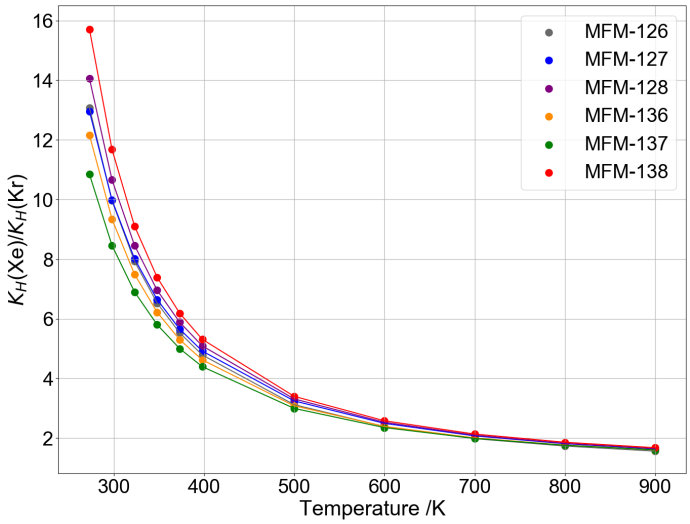

Figure 3.13: *The ratio of computational Henry constants at infinite dilution for Xe and Kr, $K_H(Xe)/K_H(Kr)$, in the range of 273 K to 900 K.*

properties and phenomena in addition to the temperature imposed by external conditions. Adsorption and desorption are accompanied by energy fluctuations, [222] and the porosity of MOFs causes a challenge by limiting dissipation of additional heat. Thermal behaviour thus further complicates efficient MOF use in practical applications compared to theoretical isothermal conditions. This must be further examined in later stages of development of potentially useful MOFs, or indeed other porous materials. It is particularly true in cases where high temperature conditions are considered, such as for MSR product separation. [52]

The locations of adsorption within a MOF framework can give information on the nature of binding and site affinity for guests. Such information can determine whether binding behaviour follows similar trends to those previously observed in well-performing structures. To this end, the probability density of the guests within the framework over the duration of the Monte Carlo simulations may be calculated using kernel density estimation in two dimensions. Probability density data is presented for MFM-126 as an example. For the case of a 20/80 Xe/Kr mixture at 10 bar and 298 K, probability density distributions of Xe and Kr locations are plotted alongside framework atom positions and shown in Figure 3.14. Viewed from the z-direction and looking into the

pores, both gases appear to display adsorption primarily in pore centres, in line with literature findings of noble gas adsorption involving interaction with all sides of the pore. [89, 187] Viewed from the x-direction and towards the side of the pores, it is clear that there are particular preferential adsorption sites. These do not all fall directly in the centre, but many appear to facilitate interaction with multiple parts of the framework at once. The probability density distributions of Kr and Xe are very similar: the two gases adsorb in the same sites, so it is only the strength of this adsorption that causes the preferential binding of Xe observed in uptake magnitudes. Also in Figure 3.14 is a snapshot of uptake in MFM-126 at one Monte Carlo simulation step for the 20/80 mixture at 10 bar and 298 K. Here, both Xe (green) and Kr (purple) are observed throughout the pores.

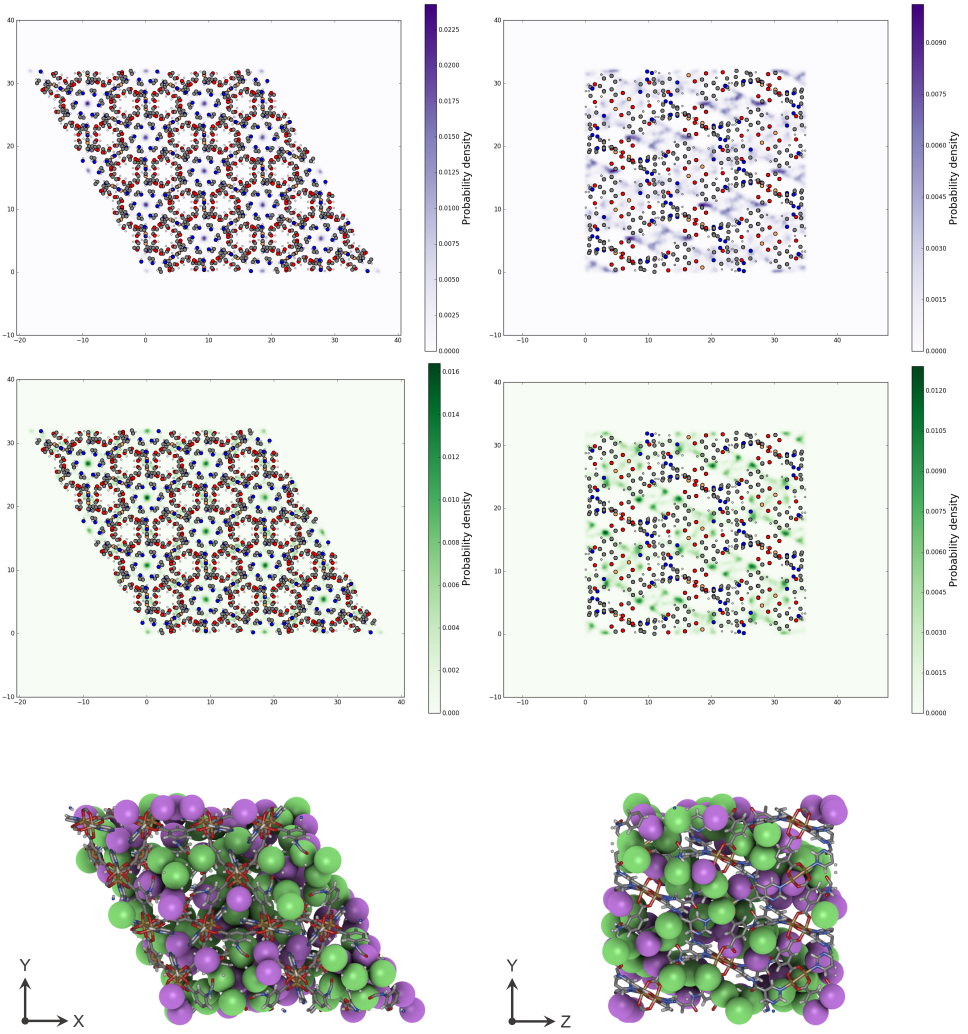

Figure 3.14: *Visualisation of Xe and Kr binding for a 20/80 Xe/Kr mixture in MFM-126 at 10 bar and 298 K. Top and centre rows: probability density distributions of Kr (purple) and Xe (green), viewed along the z-direction (left) and x-direction (right), with axis dimensions in Å. Bottom row: snapshot of loading of Kr (purple) and Xe (green) at one Monte Carlo step during the simulation.*

3.3 Conclusions

The MFM series of six robust, experimentally synthesised Cu-based MOFs has been computationally assessed for its Xe/Kr uptake and separation properties. These MOFs have already been synthesised and have shown uses for other environmentally and industrially relevant separations, so finding additional applications for them as multifunctional materials is desirable. The series has been shown to display fairly substantial selectivity for Xe over Kr. Their particularly impressive combination of selectivity and high total uptakes makes them promising separation candidates. Comparison to literature has allowed the MFM MOF series as a whole to be established within currently observed trends for Xe/Kr separation. It has been possible to identify MFM-138 as the most promising structure by several measures. The best-performing MOF of the series computationally possesses the longest, thinnest pores, decorated with the most aromatic rings, in line with similar previous observations of other structures. Additionally, visualisation of locations of guest adsorption has shown appreciable Xe and Kr adsorption at the centres of pores and interacting on many sides with the frameworks.

Further experimental study is required to establish the Xe/Kr separation properties of the MFM MOFs. Computational uptake predictions are likely to be a theoretical upper bound for experimentally observed behaviour. Although experimental behaviour has been demonstrated for the MOFs for other separations, their experimental Xe/Kr separation behaviour remains to be seen.

This study has additionally demonstrated some of the usefulness of computational work in materials discovery, particularly in the context of porous materials and gas separation. Careful selection of computational methods and of the parameters used within them has been employed in order to build up a rich picture of the structure of a promising family of MOFs, and of the adsorption and binding behaviour of those MOFs with Xe and Kr. Various important methods have been used and the process of applying them demonstrated. The

importance of judicious method selection has been illustrated, in particular regarding selection of appropriate force field parameters.

Chapter 4

Gas Uptake Properties of Metal Organic Frameworks Modelled Using Grand Canonical Monte Carlo and Machine Learning Methods: Biogas Upgrading

This chapter reports the results of collaborative work in which some of the machine learning aspects were carried out by Dr Samuel Boobier (PDRA of Professor Jonathan Hirst, School of Chemistry, University of Nottingham). This work is acknowledged in captions of the corresponding figures and tables.

The use of structural calculations, classical force fields and Monte Carlo methods in predicting gas sorption behaviour for small groups of MOFs has been seen in chapter 3. [86] With these methods, a great deal of information can be obtained about a material before any experimental work is carried out. As discussed in chapter 1, the power of computational methods in this context increases when they are applied in a high-throughput manner. Facilitated by the existence of large material databases, high-throughput methods enable prediction of behaviour of vast numbers of porous materials, helping to focus

experimental efforts and to elucidate structure-property relationships. [5, 50] Prediction power can increase further by incorporating machine learning methods into high throughput studies. With well-trained machine learning models, results of gas uptake calculations at working pressures can be predicted from more cheaply calculated data. [88, 180, 187] Machine learning calculations can also further understanding of structure-property relationships.

There is value in applying such methods to environmentally and industrially important processes in order to promote progress by efficiently suggesting promising uptake and separation materials. A number of uptake situations have previously been explored using machine learning techniques, including the Xe/Kr separation discussed in chapter 3 [89, 187], H_2 storage [227] and situations involving sorption of CH_4 and CO_2, which are relevant in several environmentally important contexts, both separately for storage and together for separation. [88, 180, 228]

Purification of the biogas stream obtained from decomposition of agricultural and industrial waste is one such environmentally pertinent aim relating to CO_2 and CH_4. It remains a challenging but promising goal towards a renewable [229] source of biomethane fuel and a more sustainable alternative to fossil fuels. [48] Biogas is composed predominantly of a CH_4/CO_2 mixture along with trace contaminants including water and H_2S. CH_4 must be separated from the rest to obtain biomethane of increased purity sufficient for use in internal combustion engines. [48] Multiple established approaches [48, 230, 231] are routinely used for upgrading the calorific content of biogas by removal of CO_2, although these technologies can be costly and energy-intensive. Among them, adsorptive [232] and membrane [48] separations by porous materials are attractive options subject to improvement in yield, efficiency and sustainability. [48, 230, 231] Operating conditions for the separation depend on the process and materials used, and can be optimised for a given setup. The common PSA separation tends to require adsorption at high pressures at or above 10 bar [218, 219] and desorption at pressures of 1 bar, or much lower for PVSA,

as discussed in chapter 3. [88, 218, 220] Meanwhile, TSA can be carried out at ambient pressures [218, 233] with adsorption temperatures between 273 K and room temperature and desorption temperatures elevated by a margin on the order of 100 K. [234] Membrane separation uses pressures of a few bar with no demand for elevated temperature. [48] Discovery of optimal porous materials is a route towards improved performance; such a material must be selective of CO_2 over CH_4 while also exhibiting sufficiently high CO_2 uptake for practical use. Many gas separations are characterised by a trade-off between selectivity and uptake, [214, 215] rendering search for high-performing materials challenging. Meanwhile, if membrane separation is to be used it is further complicated by additional reliance on effective diffusivity. [48]

Among prominent candidate materials, as for many gas separations, are porous MOFs. [35, 37, 38, 235] It has already been discussed in chapter 1 and chapter 3 that these high surface area complexes of metal-containing nodes and organic linkers have become well-known as highly promising when applied to a range of processes [24, 40–42, 236] including several gas separations. [24] As discussed, structural variety of MOFs occupies a vast and diverse chemical space, with around 100,000 reported synthesised MOFs [43] and many more proposed hypothetical structures. [33, 237] With such a range of MOFs comes opportunity and it is not unreasonable to hope that some MOF may exist displaying any realistic uptake property. However, this is also accompanied by the challenge of finding that desired structure. Experimental prediction and tuning of relevant properties for the entire chemical space is inaccessible. High-throughput use of computational force field methods provides reasonable uptake and selectivity predictions and fundamental insight into structure-property relationships. Screenings of thousands of MOFs for separation of gas mixtures are readily available in the literature, [89, 238] including a recent search of nearly 7,000 MOFs for biogas upgrading properties with a focus on membrane separation. [18]

The advancement of the computational MOF screening by development of

ML models suitable for predicting gas sorption properties has expanded the size of database that can be screened and further reduced computational cost. Application of ML has become prominent in many fields including prediction of material properties [87] and solubility, [239] and the methods which are followed in the area were introduced in section 2.4. While not yet dominating all gas separation studies of MOFs, the use of ML models in the area is rising, and they have been applied on several occasions, [180, 187, 228, 240, 241] including for CO_2/CH_4 separation. [88] Among these studies, some have considered the relative performance of different ML methods in ways that can inform future studies. As might be expected for machine learning algorithms attempting to model a complex space, favourability of non-linear over linear methods has been observed. [180] When it comes to feature selection, structural descriptors which are readily and cheaply calculated are favoured, as in the CO_2/CH_4 study of Aghaji et al [88] However, features which capture chemical information such as Henry constants, [242] binding energy or the Voronoi energy introduced by Simon et al [187] can be seen to improve the quality of ML models. Though all this has been observed, the specifics of a machine learning model are complex and models developed by Suyetin [240] to predict CH_4 uptake and working capacity appear to perform well using multiple linear regression and only three features, all of them structural. This study demonstrates the prudence of attempting to fit simple models and not dismissing them out of hand. By Occam's Razor, a simple model which performs well is preferable to an equivalent more complex model.

Selection of appropriate models and features is important, but as discussed in section 2.4 that is only part of the story. The quality of a ML model is as much or more decided by the data selected to train it as by the nature of the model itself. Most ML studies of MOFs obtain their structures from hypothetical databases [33] as a direct result of the very usefully large volume of data they can contain. The significantly smaller training sets that can be made from only real MOFs can reduce the statistical performance of a machine

learning model and restrict the methods that can be readily applied. However, since the structural viability and experimental accessibility of hypothetical MOFs is not guaranteed, hypothetical databases may contain fictitious training data that does not add to the quality of models for predicting viable MOFs. It is instructive, therefore, to analyse the quality of high-throughput methods and machine learning models which use only structures that are confirmed to exist. Furthermore, it has already been discussed in this work that significant levels of identifiable structural issues can contaminate training data and may exist in databases of both kinds as a result of structure processing.

In this chapter, high throughput calculation of biogas upgrading performance is made for a dataset of real MOFs, which in itself provides insight into the structures most likely to display strong performance combining both high uptake and high selectivity. ML models are then developed to efficiently and accurately predict performance of MOFs for biogas upgrading using a small training and validation set containing only real structures which have been curated to minimise the presence of structural problems. Features are chosen from among a selection covering both structural and chemical information expected to be relevant to gas adsorption processes. Calculated biogas separation data pertaining to an external set of hypothetical MOFs is used as a test set. This allows assessment of the ability of the models to make predictions on data from other sources, and allows comparison between models trained on real data and literature models trained on hypothetical data. Models capable of making strong predictions of the properties of realistic MOF structures using a few well-selected features are crucial to unlocking biogas separation potential.

4.1 Dataset Preparation, Cleaning and Curation

The starting point is selection of structures whose biogas upgrading properties are to be described. The prevalence of problematic structures in commonly-used databases of both real and hypothetical MOFs was referred to in the previous section and discussed in Section 1.2.1, along with the problems that this can cause for MOF studies. With that in mind, MOFs were selected for this study by a careful pruning process which minimises the presence of structures featuring the issues which can arise from structural determination methods and render published coordinates chemically unfeasible.

MOFs were initially taken from the Cambridge Structural Database (CSD) MOF subset. [21] The subset contains structures selected from the CSD on the basis of a number of criteria which indicate that a structure is a MOF. In the previous high-throughput search of Glover and Besley for MOFs with biogas upgrading properties, [18] the members of this subset were algorithmically stripped of solvent using the python script published alongside the CSD MOF subset. [21] They were also at that stage filtered according to geometric criteria, excluding frameworks with zero ASA and frameworks with PLD less than 3.80 Å, which is the kinetic diameter of CH_4, the largest biogas component considered. [18] This left a total of 6,768 structures which were used in the previous study and taken as a starting point in this work.

Close examination revealed problems associated with a number of structures. Some contained either no metal or no carbon, which would prevent them from being considered as metal organic frameworks, while others contained only one or two elements in total. These problems may be the result of over-zealous solvent removal algorithms. Meanwhile, some contained overlapping atoms or lacked H atoms, which may be due to problems in initial crystal structure determination. All structures that contained no metal, no carbon, or only one or two elements were first identified and removed from the dataset,

leaving a total of 6,664 structures.

Overlapping atoms were next combated. Any structures with any atom-atom distances less than 0.5 Å were removed from the dataset. This is a reasonably lenient bond length criterion, as 0.5 Å is substantially shorter than the H_2 bond (0.74 Å) the shortest common chemical bond. Therefore, any MOFs that are removed at this stage can be confidently considered to be problematic structures. Following application of this criterion, 6,639 structures remained.

The structural information files of some of the MOFs contained references to D atoms in place of H in some places. While D atoms exist in nature and there is no reason to believe that the MOF structures did not include them, computational force fields do not in general distinguish between isotopes and a D atom is not by default recognised by a force field algorithm. For simplicity, all D atoms were replaced with H. Since this did not involve removal of any MOFs from the database, 6,639 structures still remained. It was then possible to remove all structures which still contained no H atom, following which 6,630 structures remained.

In addition to this collection of observed symptoms of problematic structures, structure determination methods and solvent removal algorithms can result in representations of materials with atoms in unviable oxidation states (see section 1.2.1). [109] A published MOF Oxidation State And Electron Count (MOSAEC) code was used to identify offending MOFs and assess the prevalence of this issue. [109] This code is designed to read a crystallographic information file (cif) of a MOF and identify metal atoms with the following problematic features: impossible valency (that is, valency larger than the number of valence electrons on the metal atom); valency that has not been observed for the metal atom in nature; zero valency; non-integer valency; valency that has a low occurrence for the metal in nature; unusually high electron count (above 20 for non-f-block metals and above 32 for f-block metals) or unusually low electron count (below 14). The MOSAEC code was applied to all 6,639

structures remaining in the dataset and any structures that were flagged as having any one or more of these problematic features were removed. A total of 3,553 structures were flagged as problematic, leaving only 3,086 MOFs in the dataset.

A filter for dimensionality was then applied. The raw database contained MOFs that were 3D, 2D, and 1D, as well as 0D structures which were not periodic in any direction and so would be better described as molecules or metal complexes than as MOFs. The dimensionality of each of the remaining 3,086 structures was determined algorithmically using the Zeo++ software package. [81] Of these, 1,715 3D structures, 686 2D structures, 611 1D structures and 74 0D structures were identified. It has already been discussed in chapter 1 that both 3D and 2D structures can constitute MOFs, and can have advanced and valuable properties, including in gas separation. 1D structures may be considered to fit the definition of MOFs: they can contain metal nodes joined into periodic structures by organic linkers, and classification of an object that is periodic in one dimension as a framework, though questionable, is not wholly to be dismissed. However, 1D structures may possess different structural properties to more conventional MOFs, and 1D structures may also arise from erroneous removal of connecting struts during solvent stripping. Therefore, all 1D and 0D structures were removed from the database, leaving a total of 2,401 MOFs. Of these 2,401 MOFs, later charge equilibration calculations were unable to complete for 20, leaving a total of 2,381.

In the CSD, the refcode of a MOF is a string of six letters which identifies it from among other MOFs. The six letters may have a two-digit number appended, and materials with the same string of six letters with different digits appended are based on the same MOF structure, but may differ by small coordinate translations, inclusion of solvent, or similar. This can lead to duplicates of highly similar structures in the database. The databae of 2,381 MOFs contained several near-identical copies of the MOF Cu-BTC (refcode DOTSOV) and initial model fitting suggested that this would negatively affect

the statistics of the dataset. Therefore, for any set of MOFs with the same 6-letter refcode, only the first appearance in the curated dataset was kept. This may have removed some MOFs which differed substantially from their partners, but in general will have removed duplicates and improved the diversity of the database. Following this, 2,169 MOFs remained in the dataset. This is a small dataset compared to those used in some previous high-throughput and machine learning studies of MOFs, but it contains only structures which have been synthesised experimentally and have passed through a stringent filtration process.

The removal of MOF structures at each stage of this workflow highlights the prevalence of problems in MOF databases which cannot be ignored. Prior to the oxidation counting stage, problems obvious to the eye but obscured when encountered within a large volume of data were addressed. Nearly 10% of the data was filtered out. This amounts to a substantial proportion of calculations being carried out on structures with low relevance in cases where filters are not applied. The proportion increases dramatically when oxidation counting is taken into account and more than half of the remaining structures are flagged as problematic. The CSD MOF subset used here is a well-known dataset commonly utilised in MOF studies, and is far from the only database which suffers from the issues encountered. Thus, the importance of checking databases for the viability of their MOFs must be emphasised, while it is also necessary for researchers to ascertain that the structures they use possess the dimensionality they intend to study.

4.2 High-Throughput Prediction of Biogas Separation Properties Using Classical Simulations

With the 2,169 curated MOFs identified, the first stage of modelling involves assessment of biogas separation potential by calculating a performance metric using force fields combined with GCMC methods, similar to the processes used in chapter 3. This alone is valuable, giving insights into which structures are likely to be useful for biogas upgrading and more generally for uptake and separation applications of CH_4 and CO_2, its major components. Further, the results of these calculations may be used as input data for a machine learning model which predicts the measured performance metric for new MOFs (section 4.3).

4.2.1 Performance Metrics, Geometrical and Infinite Dilution Data

Several uptake and separation properties of MOFs can be identified, and several may be and previously have been used as performance metrics for gas uptake and separation properties. [88, 180] Some of the most relevant are uptake of the dominant gas and selectivity, a measure of the affinity of a material for the dominant gas over its competitors (see chapter 3). When seeking gas separation properties, uptake and selectivity are generally forefront, but other metrics relevant to the practical uses of a given material such as working capacity [88] have previously been put to use, while permeability and permeability selectivity are highly relevant to membrane separations. [5, 48]

Here, the focus is on separation of CO_2 and CH_4 as the major components of biogas: absolute uptake of CO_2, N_{CO_2}, and selectivity of CO_2 over CH_4, S_{CO_2/CH_4}, have the most relevance. The selectivity metric was additionally used in chapter 3, and is defined by equation 3.1 in that chapter. As discussed

previously, it is a measure of the preferential adsorption of one gas (in this case CO_2) over another (CH_4), and a porous material will only be useful for biogas separation if it displays sufficiently preferential adsorption, while a strong candidate will also necessarily display high uptake of CO_2 in absolute terms, maximising efficiency. Both uptake and selectivity of small-molecule gases in MOFs may be calculated using classical force field methods and GCMC simulations, as in chapter 3. However, a trade-off relationship may exist between uptake and selectivity, with materials displaying high uptake often exhibiting low selectivity and vice versa. [214, 215]

The uptake-selectivity trade-off complicates the search for separation candidate materials. This complication merits the use of a metric that encompasses both uptake and selectivity. To that effect, a metric known as trade off between selectivity and uptake (TSN) has previously proved useful in describing separation by MOFs of natural gas components [215] and in describing separation by zeolites of H_2S from alkane gases. [214] A version of TSN, TSN_{CO_2/CH_4}, may be applied to the CO_2/CH_4 separation as in Equation 4.1.

$$TSN_{CO_2/CH_4} = N_{CO_2} log_{10}(S_{CO_2/CH_4}) \tag{4.1}$$

Computational predictions for the metric TSN_{CO_2/CH_4} were here calculated for working conditions of 10 bar and 298 K for the members of the curated MOF dataset using GCMC simulations similar to those employed in Chapter 3. Uptakes of each gas N_{CO_2} and N_{CH_4} for the MOFs in the presence of a 50/50 binary mixture of the two gases were determined, and hence S_{CO_2/CH_4} was determined under the same conditions. Thus, TSN_{CO_2/CH_4}, as well as the individual binary mixture values of N_{CO_2}, N_{CH_4} and S_{CO_2/CH_4} can be used in evaluating the performance of members of the dataset and proposing useful MOFs from it for the separation. Alongside this, values of N_{CO_2} and N_{CH_4} were calculated for the MOFs in equilibrium with pure reservoirs of each single component at 10 bar. These may be compared to the binary mixture values

and may be indicative of gas storage potential in addition to the biogas separation ability, although further investigation under storage conditions would be necessary for confirmation. Following analysis of the performance of the curated MOF dataset as quantified by the calculated TSN_{CO_2/CH_4} and loading data, the same data are valuable combined with machine learning methods, which is done in section 4.3.

In addition to quantification of uptake, selectivity and the trade-off between the two, it is instructive to consider other properties of MOFs including geometrical and infinite dilution energy descriptors. These can be useful in elucidating structure-property relationships and identifying features which unite well-performing MOFs and so should be promoted. Meanwhile, establishment of a relationship between infinite dilution data and more practical data applicable at working pressures is an important part of understanding separation materials. Thanks to their low cost, it is also common to use geometrical and infinite dilution descriptors in early stages of high throughput screenings to filter out materials likely to perform poorly and reduce the number of structures on which expensive calculations are performed. It is no coincidence that features of the same kind are used as the training input for machine learning models (see section 4.3.2 for more details on selection of features for the machine learning part of this chapter). Throughout this chapter, structural and infinite dilution properties for the curated dataset are taken from the data of Glover and Besley. [5]

4.2.2 Computational Setup Used to Calculate Biogas Upgrading Metrics

To determine the absolute uptake values N_{CO_2} and N_{CH_4} used to acquire the other biogas upgrading metics, GCMC simulations were employed using the RASPA software package. [82] A similar process was followed here to that used in Chapter 3, but now in a high throughput manner, being applied to

thousands of MOF structures rather than six. Computational details of the Monte Carlo setup are outlined below.

Interactions between components of the system were modelled using Lennard-Jones potentials, with host-guest interactions and guest-guest interactions considered, but no host-host interactions. The guest molecules CO_2 and CH_4 were modelled using the TraPPE formalism, with CO_2 treated using a 3-site model in which each site has a partial charge and van der Waals parameters associated with it, and CH_4 treated using a single site model with no charge but with van der Waals parameters.

Van der Waals parameters for frameworks were taken from the UFF, [4] which is a useful general force field for high-throughput screening, but is as a consequence is not necessarily well-tuned to specific systems (see section 2.1.2). Framework partial charges are necessary to model interactions of framework atoms with CO_2. These were determined prior to uptake simulations using the extended charge equilibration (eQeq) method available within the RASPA software package. [82]

During GCMC simulations, van der Waals interactions were calculated with a truncated cutoff of 12.8 Å, with sufficient replicas of the unit cell of each MOF used so that perpendicular distances were at least 25.6 Å, two times the cutoff. Pairs of unlike atoms were treated using Lorentz-Berthelot mixing rules. Electrostatic interactions where reqired were calculated using the Ewald method with precision of 10^{-6}. The MOF system was in equilibrium with an imaginary gas reservoir whose composition depended on the task at hand, either pure CO_2, pure CH_4 or a CO_2/CH_4 mixture where each gas has a mole fraction of 0.5. The temperature was 298 K and the external pressure was 10 bar. Monte Carlo moves available to the system were translation of a guest, rotation of a guest, insertion or deletion of a guest (grouped together by the software under a move called swap) and in the binary mixture case changing identity of a guest from CO_2 to CH_4 or the reverse. Simulations were run for 10,000 equilibration cycles followed by 10,000 production cycles, where the

number of steps in a cycle is equal to the number of molecules in the system, or to 20 if there are fewer than 20 molecules in the system.

4.2.3 Results of Classical Simulations: Identification of Promising MOFs

Gas Uptake

Uptake of CO_2 (N_{CO_2}) and CH_4 (N_{CH_4}) at 10 bar and 298 K were calculated for each MOF in the curated dataset using the methods described in section 4.2.2. Both single component uptakes and binary 50/50 CO_2/CH_4 mixture uptakes were calculated, where in the single component case the MOF can be considered to be in equilibrium with a reservoir of a pure gas, and in the binary mixture case it can be considered to be in equilibrium with a gas mixture reservoir. Single component uptakes will be referred to as $N_{CO_2}(sc)$ and $N_{CH_4}(sc)$ for the purposes of this discussion, and binary mixture uptakes will be referred to as $N_{CO_2}(bm)$ and $N_{CH_4}(bm)$. Where no suffix in brackets specifies whether an uptake is of a single component or a gas within a binary mixture, it can be assumed to be a binary mixture uptake.

In figure 4.1, uptake of CO_2 is compared to uptake of CH_4 for both cases: $N_{CH_4}(sc)$ is plotted against $N_{CO_2}(sc)$ and $N_{CH_4}(bm)$ is plotted against $N_{CO_2}(bm)$. To allow identification and analysis of structure-property relationships, four plots are shown for each case, with the points in each plot coloured according to the values of a structural feature, void fraction (VF), volumetric surface area (VSA), pore limiting diameter (PLD) and largest cavity diamter (LCD). Values for the structural features are taken from the work of Glover and Besley. [18] Single component and binary mixture plots use the same axis scales to facilitate comparison between the two cases.

Examining the single component case can give a general indication of sorption performance. It can also give some indication of whether the MOFs in the curated dataset are likely to be effective for caputre and storage of CH_4 and

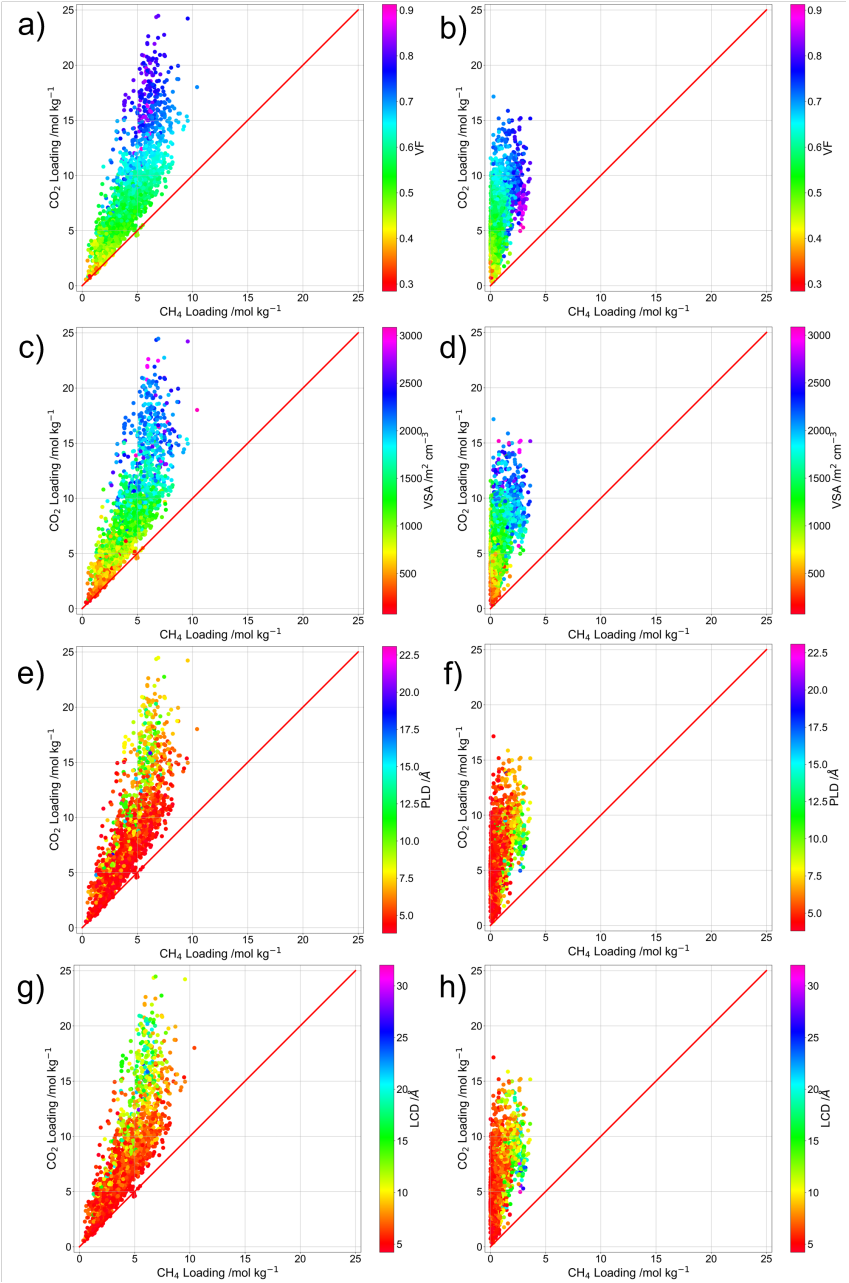

Figure 4.1: *N_{CH_4} plotted against N_{CO_2} for the MOFs in the curated dataset as calculated using single component GCMC at 10 bar and 298 K, on the same scale for comparison and with the line $y = x$ given in red in each case. Left: single component case; right: binary mixture case. Four versions of each plot are shown with points coloured according to different structural descriptors as calculated by Glover and Besley: [18] a and b: coloured by void fraction (VF); c and d: coloured by volumetric surface area (VSA); e and f: coloured by pore limiting diameter (PLD); g and h: coloured by largest cavity diameter (LCD).*

CO_2, which, although not the primary focus of this study, are also active fields and merit mention. As a single component, CO_2 loading reaches nearly 25 mol kg^{-1} in the best performing MOFs, and CH_4 loading reaches just over 10 mol kg^{-1}. For the use of CH_4 as vehicular fuel, storage targets set by the US department of energy (DOE) in 2012 [243, 244] and pursued since [245–247] include a gravimetric uptake which amounts to around 31.2 mol kg^{-1}. Clearly the maximum CH_4 uptakes observed in the curated MOF dataset are lower than this, but for CH_4 storage applications pressure is higher, with 35 bar rather than 10 bar being standard, [33] resulting in higher uptake. Pursuit of the high-uptake MOFs identified here at higher pressures may be fruitful in cases where CH_4 storage is desired. Turning to $N_{CO_2}(sc)$, a range of uptakes under different conditions have previously been considered useful, [55] with values above 20 mol kg^{-1} standing out as particularly high. A selection of the MOFs in the curated dataset have predicted uptakes of CO_2 at 10 bar above 20 mol kg^{-1}, which compare well to literature values. Further pursuit of these MOFs for CO_2 storage applications under relevant conditions may be useful.

For binary mixtures, the maximum total uptake of both CO_2 and CH_4 in the curated dataset decreases compared to the single component cases as each gas experiences competition with the other. This is particularly true of CH_4 uptake; the stronger adsorbent, CO_2 is able to outcompete the weaker CH_4. This results in many MOFs displaying very low calculated CH_4 uptake in the mixture and the highest CH_4 uptake in the mixture being less than half as high as the highest uptake of pure CH_4. This is by no means negative when seeking MOFs for separation applications, in which desired material properties include high uptake of CO_2 combined with low uptake of CH_4. The most promising MOFs are likely to be those which appear in the top left area of the plots in Figure 4.1. There are indeed a few points in that area, though there is also a high density of points with a low uptake of both gases. The relationship between $N_{CH_4}(bm)$ and $N_{CO_2}(bm)$ and its implications on separation applications is treated in greater detail in the next section.

Considering the structural features indicated by the colouring of the plots, it is clear that increasing both VF and VSA is in general accompanied by an increase in loading of both gases, although this changes once a certain point is reached: the MOFs displaying the very highest CO_2 loading possess VF and VSA somewhat below the very highest observed. Some of the MOFs with the highest VF display both a drop in CO_2 uptake compared to the maximum and a relatively high CH_4 uptake, which is not promising for selectivity. In general, both high VF and high VSA correspond to a large amount of space for guest molecules measured, respectively, by total void space and by adsorption site space. However, with adsorption commonly being favoured in geometrically compatible areas in which a molecule can experience multiple different simultaneous interactions with the framework, there is a point at which both an increase in VF and an increase in VSA, while increasing total space, decrease favourable adsorption sites and so reduce total uptake. This is related to the trend seen for the pore size measures, PLD and LCD: the highest uptakes of both gases are not seen for the MOFs containing the largest pores, but instead for MOFs with some small to intermediate pore size. In the binary mixture case, the highest CH_4 loading is seen for larger pores than the highest CO_2 loading. For very small pores, CO_2 can outcompete CH_4 for limited binding sites, whereas in slightly larger pores the two may be able to coexist or in some cases CH_4, being larger and more likely to have a geometrical fit with the framework, may have a higher chance of withstanding the competition.

Trade-off Between Uptake and Selectivity

As in section 3.2.2, the relationship between uptakes of the two gases is here quantified by the metric S, selectivity using Equation 3.1 (section 3.2.2). Uptake of the dominant gas and selectivity of the dominant gas over its competitor(s) both being of importance to gas separations, the relationship between $N_{CO_2}(bm)$ and S_{CO_2/CH_4} as calculated by GCMC simulations is examined. In Figure 4.2, N_{CO_2} is plotted against S_{CO_2/CH_4} using both a linear and a loga-

rithmic scale. Using the linear scale much information is unclear as the scale of S_{CO_2/CH_4} is large with the vast majority of MOFs having values towards the lower end in relative terms. This is alleviated by the logarithmic scale, which helps to elucidate some of the obscured information. Figure 4.2, like Figure 4.1, contains four versions of each plot, one with points coloured according to each of the structural features VF, VSA, PLD and LCD. [18]

The upper bounds of both uptake and selectivity are clear from Figure 4.2: there exist MOFs in the curated dataset with high values of S_{CO_2/CH_4} reaching above 10^4, and with binary mixture CO_2 uptakes at 10 bar reaching almost 17.5 mol kg^{-1}. Both upper bounds, in particular selectivity, may be considered among well-performing literature examples, [55] and so some members of the curated dataset may prove useful for practical separations. The chance that structures described as well-performing in this study may be truly useful MOFs is maximised by the curation steps described in section 4.1 which minimise the chance of poorly described MOFs persisting in the dataset.

However, Figure 4.2 also illustrates the existence of the trade-off relationship between gas uptake and gas selectivity. Using the linear scale gives an idea of this, as none of the very high-selectivity MOFs which stand out from the rest in this plot also stand out as having high loading. The logarithmic plot illustrates this in more detail. An ideal MOF would possess both high uptake and high selectivity, and be found in the upper right portion of the plot. Instead, there is a significant population of MOFs in the lower left, with branches into the lower right (high loading, low selectivity) and upper left (high selectivity, low loading). In the absence of MOFs which unite both very high loading and very high selectivity, it is instructive to identify those which find a useful compromise between the two metrics, and that is where TSN_{CO_2/CH_4} becomes useful.

The individual metrics N_{CO_2} and S_{CO_2/CH_4} are separately plotted against TSN_{CO_2/CH_4} in Figure 4.3, which uses a logarithmic scale for S_{CO_2/CH_4} and in which four versions of each plot again appear, with colouring of points

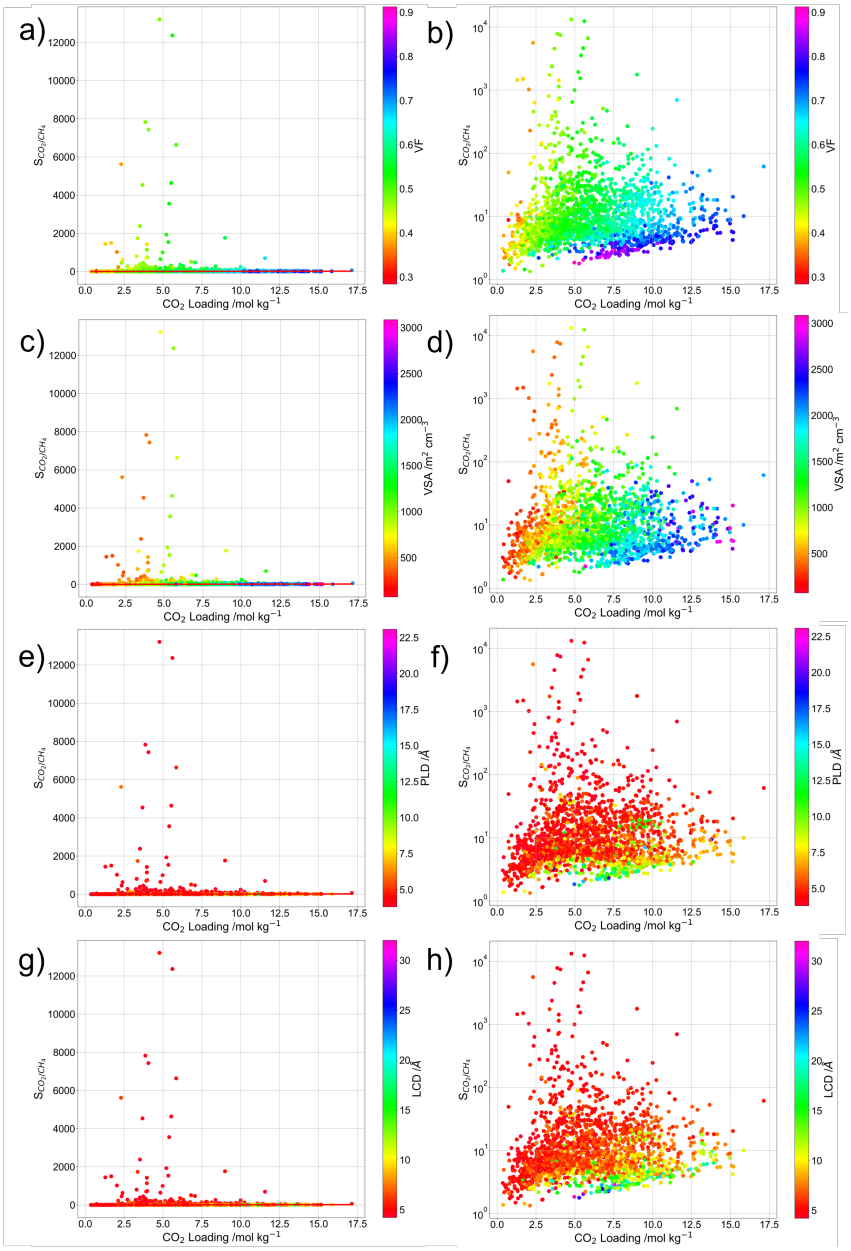

Figure 4.2: *N_{CO_2} plotted against S_{CO_2/CH_4} for the MOFs in the curated dataset as calculated using binary mixture GCMC at 10 bar and 298 K. Left: linear scale, line of $y = x$ given in red; right: logarithmic scale. Four versions of each plot are shown with points coloured according to different structural descriptors as calculated by Glover and Besley: [18] a and b: coloured by void fraction (VF); c and d: coloured by volumetric surface area (VSA); e and f: coloured by pore limiting diameter (PLD); g and h: coloured by largest cavity diameter (LCD).*

determined by VF, VSA, PLD and LCD. [18] High values of TSN_{CO_2/CH_4} capture values of both uptake and of selectivity that are fairly high, which confirms its usefulness as a metric. An upper limit for TSN_{CO_2/CH_4} around 30 mol kg^{-1} is observed, though with only a few MOFs reaching these lofty heights.

As was done for loading using Figure 4.1, Figures 4.2 and 4.3 may be used to identify relationships between the structural properties shown and the different metrics plotted. While high values of VF and VSA (up to a peak) correspond to high CO_2 uptake, the highest values of selectivity occur for ranges of these two descriptors which are lower, though not in general the very lowest. For fairly low VF and VSA, MOFs are less likely to have sufficient adsorption sites to adsorb both competing gases, but remain likely to possess sufficient adsorption sites for the dominant gas, and selectivity may be maximised. In line with the observed trade-off relationship, the highest values of TSN_{CO_2/CH_4} occur for some larger range of VF and VSA values falling between the optimum for uptake and the optimum for selectivity. When it comes to pore size, small pores and channels are seen to facilitate selectivity, with the pores allowing a close fit of the dominant gas without sufficient additional space for the secondary gas. What is also clear from Figures 4.1, 4.2 and 4.3 is that the relationships between the structural features and the biogas separation metrics are complex and interconnected. A single structural feature alone is not sufficient to make a confident prediction about any of the metrics, but the combination of multiple features in a machine learning model may be instructive.

It is useful to identify the very high TSN MOFs, and this is done in Figure 4.4, which contains a table of the names of the seven MOFs whose TSN_{CO_2/CH_4} is greater than 22. The table additionally includes the pore limiting diameter (PLD) and largest cavity diameter (LCD) of each of these 7 MOFs to allow continued consideration of the effect of structural properties on uptake properties. As discussed in section 2.1.1 both PLD and LCD may be relevant to different gas sorption situations, with LCD describing the pore size most di-

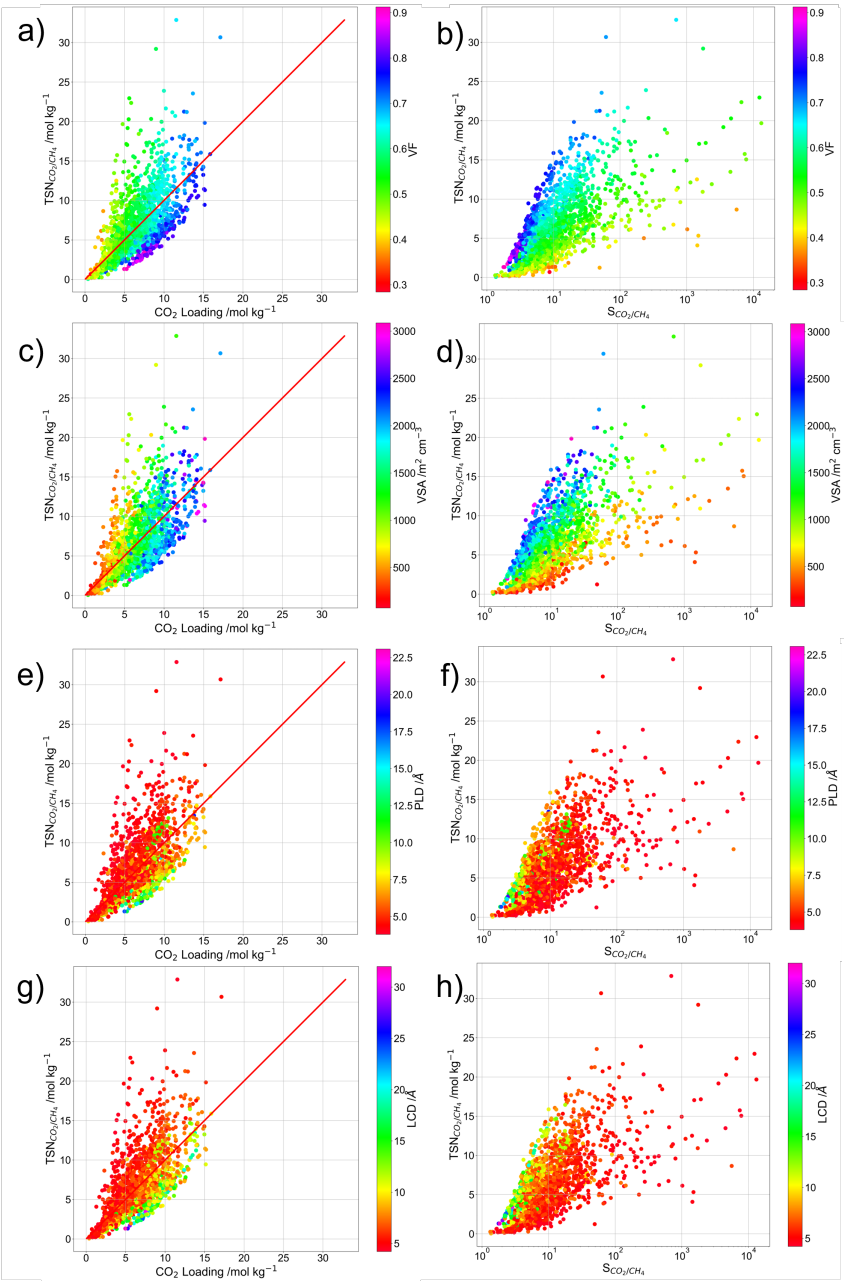

Figure 4.3: *Uptake and selectivity plotted against TSN_{CO_2/CH_4} for the MOFs in the curated dataset as calculated using GCMC at 10 bar and 298 K. Left: N_{CO_2} against TSN_{CO_2/CH_4}; right: S_{CO_2/CH_4} (logarithmic scale) against TSN_{CO_2/CH_4}. Four versions of each plot are shown with points coloured according to different structural descriptors as calculated by Glover and Besley: [18] a and b: coloured by void fraction (VF); c and d: coloured by volumetric surface area (VSA); e and f: coloured by pore limiting diameter (PLD); g and h: coloured by largest cavity diameter (LCD).*

rectly by indicating the largest space available for adsorption within the MOF, while PLD reflects the size of the cavities within the MOF where the strongest adsorption is likely to occur. To further facilitate detailed examination of the MOFs with stand-out TSN_{CO_2/CH_4} values, visualisations of each of the seven structures are also presented in Figure 4.4.

Interestingly, many of the high-TSN_{CO_2/CH_4} MOFs have structural features in common. The top four all have Zn as the metal and four of the top five have 2D structures. The exception is the MOF with refcode UQUVOS, which has a 3D structure made up of connected 2D sheets, and so is likely to share some structural properties with a 2D MOF. Alongside or instead of 2D structures, several of the top 7 MOFs possess channels which are approximately square shaped. Further, the seven MOFs have fairly small PLD, including two among the top four which are barely larger than 3.80 Å, the kinetic diameter of methane and the smallest PLD that was allowed into the dataset. The LCD of these MOFs is only a small amount larger than their PLD, indicating the MOFs are composed of channels of close to uniform nature, and do not have large cavities, with specific adsorption sites defining much of the internal volume. Thus, 2D structures with small interlayer separation and structures with small square shaped channels are highlighted as useful aims when seeking MOFs which have both high CO_2 uptake and high selectivity of CO_2 over CH_4, and Zn is highlighted as a strong choice for a metal centre.

It appears that the confined environment between two close planar sheets or within narrow square channels is conducive to high TSN_{CO_2/CH_4}. As noted in Figure 4.4, the top four MOFs all rank reasonably highly for both selectivity and uptake, suggesting that both selectivity and uptake are enhanced by this environment. The small size is likely to allow a close fit for CO_2, permitting interactions on multiple sides by a similar logic to that applied to Xe in Chapter 3. For CO_2, unlike Xe, the inhomogeneity of the molecule may additionally lead to specific interactions with binding sites of a MOF and may fit well with the 2D structures that appear to be favoured. For example, in a case where the

a)

MOF Refcode	TSN /mol kg^{-1}	S (rank)	BM CO_2 Loading /mol kg^{-1} (rank)	PLD /Å	LCD /Å
DEPJIR02	32.86	695.89 (24)	11.86 (88)	3.81	4.39
QUDJEF	30.67	61.50 (130)	17.15 (1)	4.27	5.23
UQUVOS	29.19	1,765.15 (12)	8.99 (398)	4.42	5.33
AQOWIN	23.89	244.94 (44)	10.00 (209)	3.80	4.92
SIKYIV	23.55	52.86 (144)	13.67 (29)	4.30	6.41
GAJZAU	22.95	12,363.53 (2)	5.61 (1,288)	4.30	4.80
YOCSEQ	22.36	6,633.00 (5)	5.85 (1,214)	4.83	5.30

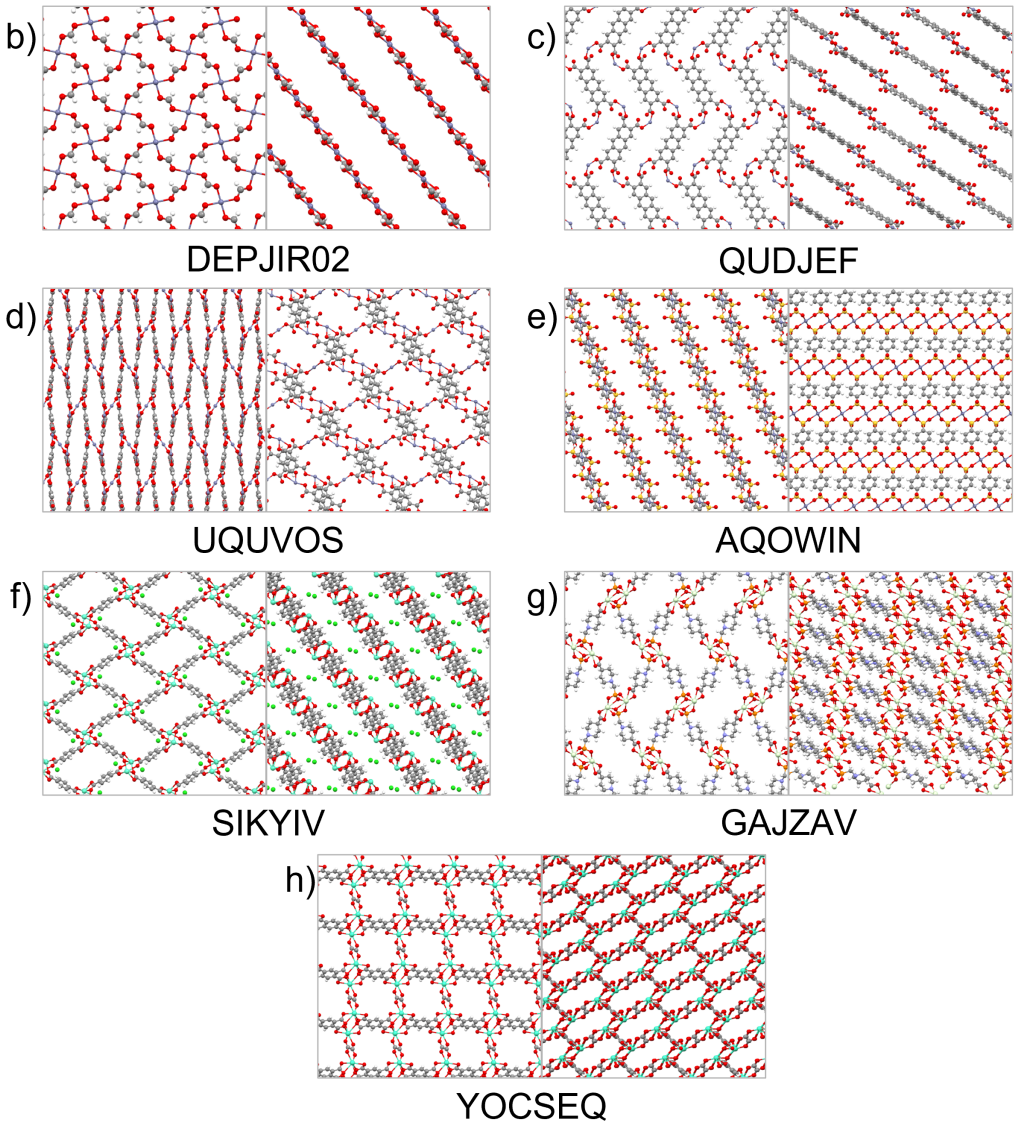

Figure 4.4: *Structures of the seven MOFs for which TSN_{CO_2/CH_4} at 10 bar and 298 K is greater than 22 as predicted by GCMC simulations, visualised along the a-axis (left) and the b-axis (right); a) table giving details of the MOFs, their TSN_{CO_2/CH_4}, selectivity and CO_2 loading at 10 bar and 298 K, along with their PLD and LCD. [18] b) DEPJIR02, metal is Zn c) QUDJEF, metal is Zn d) UQUVOS, metal is Zn e) AQOWIN, metal is Zn f) SIKYV, metal is Eu g) GAJZAV, metal is Pr h) YOCSEQ, metal is Cd.*

O of CO_2 experiences the strongest interaction, O exists on both sides of the molecule and so one molecule may experience its strongest interaction with two planes of a geometrically compatible 2D structure. Alternatively, for narrower interplanar distances, a linear molecule like CO_2 may fit well in line with the planes. Selectivity may additionally be maximised by a difficulty presented in fitting CH_4 into the observed small pores, particularly for the cases where the PLD is very close to 3.80 Å.

4.3 Prediction of Biogas Separation Properties Using Machine Learning Methods

Use of GCMC calculations to determine TSN, while cheaper than ab initio methods, can prove cumbersome, particularly at elevated pressures and particularly when applied to a large volume of structures. With that in mind, calculated TSN_{CO_2/CH_4} data can be used to form the labels of a machine learning model, that is, the data that the model is trained to predict (see section 2.4). A model that has been trained on the curated dataset to reliably predict which unseen MOFs are likely to possess high TSN_{CO_2/CH_4} can facilitate greater efficiency in future high throughput studies by allowing CH_4 and CO_2 separation properties to be predicted without need for lengthy GCMC calculations. Using only the real curated MOFs as training and validation data limits the possible size of the training set to significantly below some previous examples of related machine learning studies, but it maximises the viability of MOFs used as training data, while also minimising the computational cost of building the model. The attempt to train a strong model on this relatively small dataset also has the benefit of demonstrating the capabilities and limitations of smaller, cheaper datasets.

Prior to building any ML model, it was necessary to define training, validation and test data. The source for training and validation data was the

curated real MOF dataset set out in section 4.1. Training and validation sets were defined within the framework of k-fold cross validation with 10 folds, meaning there were not single training and validation sets, but every data point was used for training 9 times and for validation once (see section 2.4.2). This is a useful approach for dealing with small datasets as it maximises the amount that each data point is used. External test data were taken from the commonly used Northwestern Hypotheticals database. [33] This enables assessment not only of the performance of the models on structures on which they were not trained, but also of their performance on the diverse kinds of structures present in different databases. To make up the external test set, 1,000 MOFs were taken at random from the database and subjected to the same curation procedure that was applied to the real training data, in addition to geometrical and charge criteria used in the screening of Glover and Besley [18] for consistency with the training data, leaving a total of 331 external test data points.

4.3.1 Aims of the Models

Models were trained to make predictions relating to biogas upgrading performance as measured by TSN_{CO_2/CH_4} and related metrics. For raw uptakes, regression models were trained to predict $N_{CO_2}(bm)$, $N_{CH_4}(bm)$, $N_{CO_2}(sc)$ and $N_{CH_4}(sc)$ at 10 bar for a given MOF. The aim of a regression model in this context is clear: predict uptake values with the smallest possible error. The quality of the predictions, and thus of the model, is assessed by root mean squared error (RMSE), the coefficient of determination (R^2) and inspection of plots.

Similar regression models assessed in a similar manner were trained to predict TSN_{CO_2/CH_4} and $log_{10}(TSN_{CO_2/CH_4})$. The unevenly spaced distribution of TSN_{CO_2/CH_4} values seen in Figures 4.3 limits the accuracy reachable by a ML algorithm, particularly for data points in the poorly populated re-

gions of the distribution. In this case, the poorly populated region is the high TSN_{CO_2/CH_4} region, which contains the MOFs most likely to be of interest. The logarithmic model was attempted as a result of this, aiming to rescale the target value and improve predictions of high-valued points.

An alternative setup which can be useful in the context of screening for uptake properties is training a classification, rather than a regression, model. The purpose of machine learning in a ML-assisted high-throughput screening is generally to reduce the necessary volume of more expensive computational chemistry calculations, by separating a dataset into promising structures which are candidates for expensive calculations and those which are not and should be discarded. It can therefore be sufficient to train a model which classifies structures as promising and not promising without directly measuring the numerical quality of its predictions. This can be a useful way to deal with poorly distributed data such as the TSN_{CO_2/CH_4} data, as a classification model can be successful without precise numerical prediction of data in sparsely populated regions.

A binary classification model was therefore trained to identify those MOFs which are likely to have high values of TSN_{CO_2/CH_4}. For the classification model to make useful predictions, some benchmark value of TSN_{CO_2/CH_4} must be selected above which the value of TSN_{CO_2/CH_4} is considered to be high and below which it is considered to be low. Training and test data is classified as high-performing or low-performing according to this benchmark. Depending on the nature of the model used, it is trained to either directly classify new MOFs as high- or low-performing, or to predict a value of TSN_{CO_2/CH_4}, which it then uses to make the classification.

As discussed in Section 4.2.1, TSN has been used in relation to related gas mixtures in previous literature, [214, 215] though it has not been widely applied and therefore a suitable benchmark value is not immediately obvious. As an example, $TSN_{H_2S+CO_2/C_1-C_3}$ was used as a metric in a screening of 606 hydrophobic MOFs for separation of H_2S and CO_2 from a natural gas

mixture containing 7 parts CH_4, 1 part CO_2 and 0.2031 parts other gases by Qiao et al. [215] Hydrophobic MOFs were there defined as those for which $K_H(H_2O) < 2.6 \times 10^{-6}$, and were the focus due to their increased likelihood of water stability. Considering only hydrophobic MOFs, however, is likely to have limited the search to lower values of $TSN_{H_2S+CO_2/C_1-C_3}$ than otherwise, while the limits of TSN also depend on precise conditions. In this case the highest $TSN_{H_2S+CO_2/C_1-C_3}$ value calculated was 4.26 and a benchmark value of 2 mol kg^{-1} was used to identify potentially useful MOFs.

Assessment of the available literature relating to the limits of porous materials for the separation at hand shows that values of TSN_{CO_2/CH_4} notably higher than 2 mol kg^{-1} are achievable. A recent review of experimental studies of MOFs for CO_2 capture and separation [55] identified a wide range of MOFs that have been synthesised and applied to CO_2 uptake problems under various conditions in the past two decades. Over this large time period, various kinds of MOFs were considered in the context of uptake and separation under various conditions of temperature and pressure, resulting in structures being deemed useful with CO_2 uptakes ranging from below 1 mol kg^{-1} [248, 249] to above 20 mol kg^{-1}. [250] Selectivity values are less accessible to experimental studies, but a handful were identified for selectivity of CO_2 over CH_4, also spanning a wide range, from below 10 to well over 100. Selectivity values are prone to fluctuation and strong dependence on conditions, but where they are available in addition to raw uptakes, TSN values may be calculated. Using reported experimental S_{CO_2/CH_4} and N_{CO_2} values, TSN_{CO_2/CH_4} upwards of 30 may be observed. [250] In the context of the search undertaken here, which requires MOFs with high selectivity and uptake, the existence of these very high values of TSN_{CO_2/CH_4} is of particular interest, especially if they can be achieved under relevant conditions.

In line with the potential for selectivity and uptake values that yield high TSN_{CO_2/CH_4}, previous high-throughput and ML screenings seeking related metrics have used benchmarks of $S_{CO_2/CH_4} = 5$ (10 bar, 1:9 CO_2:CH_4) [88],

S_{CO_2/CH_4} = 10 (10 bar, 1:9 CO_2:CH_4) [88], N_{CO_2} = 1mol kg^{-1} (0.15 bar, pure CO_2) [228], and N_{CO_2} = 4 mol kg^{-1} (1 bar, pure CO_2) [228] to define high-performing MOFs. The variation of benchmark values is in some cases down to the conditions at which the uptake processes are studied, and is sometimes down to the desired purpose of the model. A model with a higher threshold will classify fewer members of a dataset as high-performing and so reduce the number of structures on which more expensive calculations are carried out as part of a screening. However, models trained with higher benchmarks identify a greater fraction of false positives, [88] and will also be more likely to miss high-performing structures.

A practically relevant zeolite currently used in natural gas scrubbing [88] to separate CO_2 and CH_4 is zeolite 13X, which has a CO_2 uptake of around 6 mol kg^{-1} at 10 bar [30, 251] and whose S_{CO_2/CH_4} in the presence of a CO_2/CH_4 1:9 gas mixture at 10 bar has been given as 14. This would correspond to a TSN_{CO_2/CH_4} of around 6.9 mol kg^{-1}, although for an equal composition gas mixture S_{CO_2/CH_4} is likely to be lower and N_{CO_2} somewhat higher, so TSN_{CO_2/CH_4} would change.

When proposing new materials, it is productive to identify those with better performance than current benchmarks. With this in mind, a benchmark value of TSN_{CO_2/CH_4} = 5mol kg^{-1} is used here. Using this value, it is expected that a high amount of MOFs which a model predicts to have a strong performance will have values of TSN_{CO_2/CH_4} higher than that of materials used in current practice, without overly diluting the high-performing pool with low-performing MOFs. Additionally, all of those MOFs with very high TSN which may be of particular interest are expected to be classified as high performing. This threshold is also likely to identify materials with performance that is similar to or better than benchmarks in recent high-throughput screenings for CO_2 uptake and selectivity. Using the higher of the thresholds taken from previous screenings, [88, 228], a MOF with N_{CO_2} of 4 mol kg^{-1} and S_{CO_2/CH_4} of 10 would possess a TSN_{CO_2/CH_4} of 4 mol kg^{-1}, though precise comparison

is not possible due to the difference in the conditions used here and in previous works.

Examining the GCMC data in the context of the benchmark value of $TSN_{CO_2/CH_4} = 5$ mol kg^{-1} provides further confirmation for the prudence of the choice. In the first place, with a benchmark value of 5 mol kg^{-1}, 1,136 members of the curated MOF dataset are considered to be high-performing according to GCMC calculations, leaving 1,033 low-performing MOFs. This is a useful spread of data: a machine learning model trained to classify MOFs into the two categories will be fed a sufficiently large volume of data from each category to facilitate effective learning. This consideration is lent extra relevance here because of the relative sparsity of total data necessitated by the use of the curated MOF dataset. Meanwhile, if trained effectively and applied to an unseen dataset with a similar distribution of MOFs to the training set, such a model would be expected to identify around half of that dataset as potentially high-performing, allowing half of the structures to be rejected before any GCMC calculations are attempted and significantly improving the efficiency of a screening. The well-performing curated MOFs identified by the $TSN_{CO_2/CH_4} = 5$mol kg^{-1} threshold include a small proportion of hydrophobic MOFs (by the criteria used by Qiao et al and elsewhere in the literature), [215] which may be particularly useful to identify. Additionally, initial regression predictions of TSN_{CO_2/CH_4} displayed statistical deterioration close to a value of 5 mol kg^{-1}, meaning use of a higher benchmark may lead to less confidence in classification.

While the TSN_{CO_2/CH_4} benchmark of 5 mol kg^{-1} results from careful consideration, it may not be the most appropriate benchmark to use in all cases. With the framework for the models set out here, it is readily possible to alter the benchmark for future studies.

4.3.2 Feature Analysis

Any machine learning model requires features. They must be cheaply calculated descriptors which between them display a predictive statistical relationship with the target value (see section 2.4.1). The features used to train the models in this chapter are a combination of geometrical and infinite dilution energy descriptors, which are taken from the data of Glover and Besley. [5] Both geometrical and infinite dilution descriptors are significantly cheaper to calculate than uptake at 10 bar, while several MOF databases are published with pre-calculated geometrical features, reducing their cost to nothing. However, if pre-calculated features are used, care must be taken to ensure they are in the same form as the features a model was trained on: there are several similar definitions for properties such as volume and surface area.

Glover and Besley [5] calculated 21 descriptors for each MOF. A selected smaller subset of these descriptors was used as the set of features. The 21 available descriptors are listed in Table 4.1, with descriptors used as features in this chapter shown in bold. The geometrical features are PLD (pore limiting diamter), LCD (largest cavity diameter), Density, VSA (volumetric surface area), GSA (gravimetric surface area) VF (void fraction), and PV (pore volume). Their nature is described in greater detail in section 2.1.1. The infinite dilution features include Henry constants (K_0), heats of adsorption (Q_0^{st}), diffusion coefficients (D_0^C) and permeabilities (P_0), calculated for the major components of biogas CO_2 and CH_4 as well as for the contaminants H_2O and H_2S which were treated directly in the work of Glover and Besley [18] and so were included in their data. Henry constants and heats of adsorption both relate to the affinity between a guest gas molecule and the framework, and have already been used to describe MOF-guest interactions in chapter 3. Removal of H_2S from a biogas stream is desired due to the risk of its causing corrosion within fuel systems, while the affinity of an adsorbent for water vapour can impact its water stability and therefore the practicality of its application. [18]

Heat of adsorption is a measure of energetic interactions and as such a negative value corresponds to an attractive interaction, but often the negative of Q_{st} is reported as Q_{st} (leading to a positive value being reported for an attractive interaction), as in chapter 3. In this chapter, the raw negative values are used. Diffusion coefficients and permeabilities were relevant to the screening of Glover and Besley, as the focus of that work was on membrane separations. A diffusion coefficient measures the mobility of a guest inside a framework and permeability, which is the product of diffusion coefficient and Henry constant, measures its ability to permeate through the framework.

Here, the 21 descriptors were examined and manipulated, and features were selected from among them based on a combination of statistical analysis and domain knowledge. First, the distributions of the descriptors were assessed among the 2,169 curated MOFs for which biogas separation metrics were available. All 21 distributions are shown by histograms in Figure 4.5. At this stage, presented data covers all descriptors available from Glover and Besley: [18] that is, all of the structural features and energetic features relating to the four gases CO_2, CH_4, H_2O and H_2S listed in Table 4.1. The relevance of these features to the problem at hand is addressed in subsequent stages of feature analysis. From Figure 4.5 it is clear that some descriptors are quite well normally distributed, while others show extreme skew and are unlikely to be useful in their raw form. In particular, Henry constant and permeability features are so skewed as to obscure their distributions on a standard scale. Some descriptors were scaled by application of the log_{10} operator to combat skewness. These are: PLD, LCD, PV ($cm^3\ g^{-1}$), all Henry constants, all diffusion coefficients and all permeabilities.

Correlation between features was then analysed using Pearson's R^2 values for each pair of descriptors (or scaled descriptors where relevant). This resulted in $21^2 = 441$ total R^2 values. Under these circumstances an R^2 of 1 indicates high correlation and an R^2 of zero indicates no correlation; R^2 between a descriptor and itself is always 1. As discussed in section 2.4.1, two features

Table 4.1: List of the 21 descriptors available from the work of Glover and Besley [5], including names of descriptions and the method and software used to calculate them. Descriptors in bold are selected for this work. Where relevant: MC = Monte Carlo, GCMC = Grand canonical Monte Carlo, MD = molecular dynamics, r = probe radius.

Descriptor	Description	Method	Software
PLD /Å	**Diameter of the largest sphere that can percolate through the MOF**	**Voronoi network**	**Zeo++** [81]
LCD /Å	**Diameter of the largest sphere that fits inside the MOF**	**Voronoi network**	**Zeo++** [81]
Density /g cm^{-3}	**Mass of MOF per unit volume**		**Zeo++** [81]
VSA **/m^2 cm^{-3}**	**Surface area accessible to the centre of a probe (r = 1.86 Å) per unit volume**	**Voronoi network MC sampling**	**Zeo++** [81]
GSA /m^2 g^{-1}	Surface area accessible to the centre of a probe (r = 1.86 Å) per unit mass	Voronoi network MC sampling	Zeo++ [81]
VF	**Fraction of the volume not occupied by MOF atoms**	**Voronoi network MC sampling**	**Zeo++** [81]
PV /cm^3 g^{-1}	Volume accessible to a probe (r = 0 Å) per unit mass	Voronoi network MC sampling	Zeo++ [81]
$K_0(CH_4)$ /mol kg^{-1} Pa^{-1}	Henry constant of CH_4 in the MOF at infinite dilution	Force fields GCMC	RASPA [82]
$K_0(CO_2)$ /mol kg^{-1} Pa^{-1}	Henry constant of CO_2 in the MOF at infinite dilution	Force fields GCMC	RASPA [82]
$K_0(H_2S)$ /mol kg^{-1} Pa^{-1}	Henry constant of H_2S in the MOF at infinite dilution	Force fields GCMC	RASPA [82]
$K_0(H_2O)$ /mol kg^{-1} Pa^{-1}	Henry constant of H_2O in the MOF at infinite dilution	Force fields GCMC	RASPA [82]
$Q_0^{st}(CH_4)$ /kJ mol^{-1}	**Heat of adsorption of CH_4 in the MOF at infinite dilution**	**Force fields GCMC**	**RASPA** [82]
$Q_0^{st}(CO_2)$ /kJ mol^{-1}	**Heat of adsorption of CO_2 in the MOF at infinite dilution**	**Force fields GCMC**	**RASPA** [82]
$Q_0^{st}(H_2S)$ /kJ mol^{-1}	**Heat of adsorption of H_2S in the MOF at infinite dilution**	**Force fields GCMC**	**RASPA** [82]
$Q_0^{st}(H_2O)$ /kJ mol^{-1}	**Heat of adsorption of H_2O in the MOF at infinite dilution**	**Force fields GCMC**	**RASPA** [82]
$D_0^C(CH_4)$ /cm^2 s^{-1}	Diffusion coefficient of CH_4 in the MOF at infinite dilution	Force fields MD	RASPA [82]
$D_0^C(CO_2)$ /cm^2 s^{-1}	Diffusion coefficient of CO_2 in the MOF at infinite dilution	Force fields MD	RASPA [82]
$D_0^C(H_2S)$ /cm^2 s^{-1}	Diffusion coefficient of H_2S in the MOF at infinite dilution	Force fields MD	RASPA [82]
$P_0(CH_4)$ /Barrer	Permeability of CH_4 in the MOF at infinite dilution	Force fields GCMC, MD	RASPA [82]
$P_0(CO_2)$ /Barrer	Permeability of CO_2 in the MOF at infinite dilution	Force fields GCMC, MD	RASPA [82]
$P_0(H_2S)$ /Barrer	Permeability of H_2S in the MOF at infinite dilution	Force fields GCMC, MD	RASPA [82]

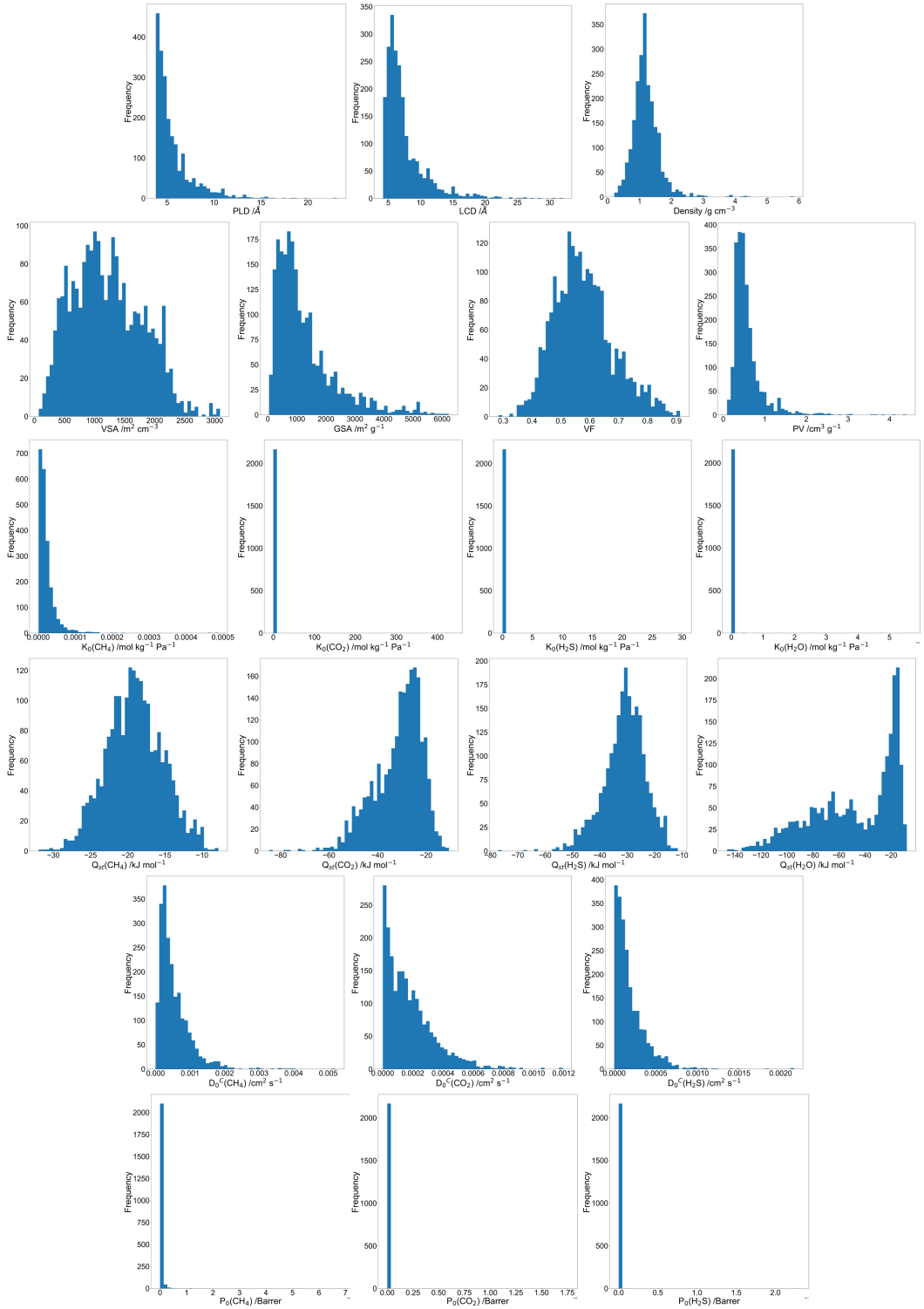

Figure 4.5: *Histograms displaying the distributions of all 21 features available from the study of Glover and Besley. [5]*

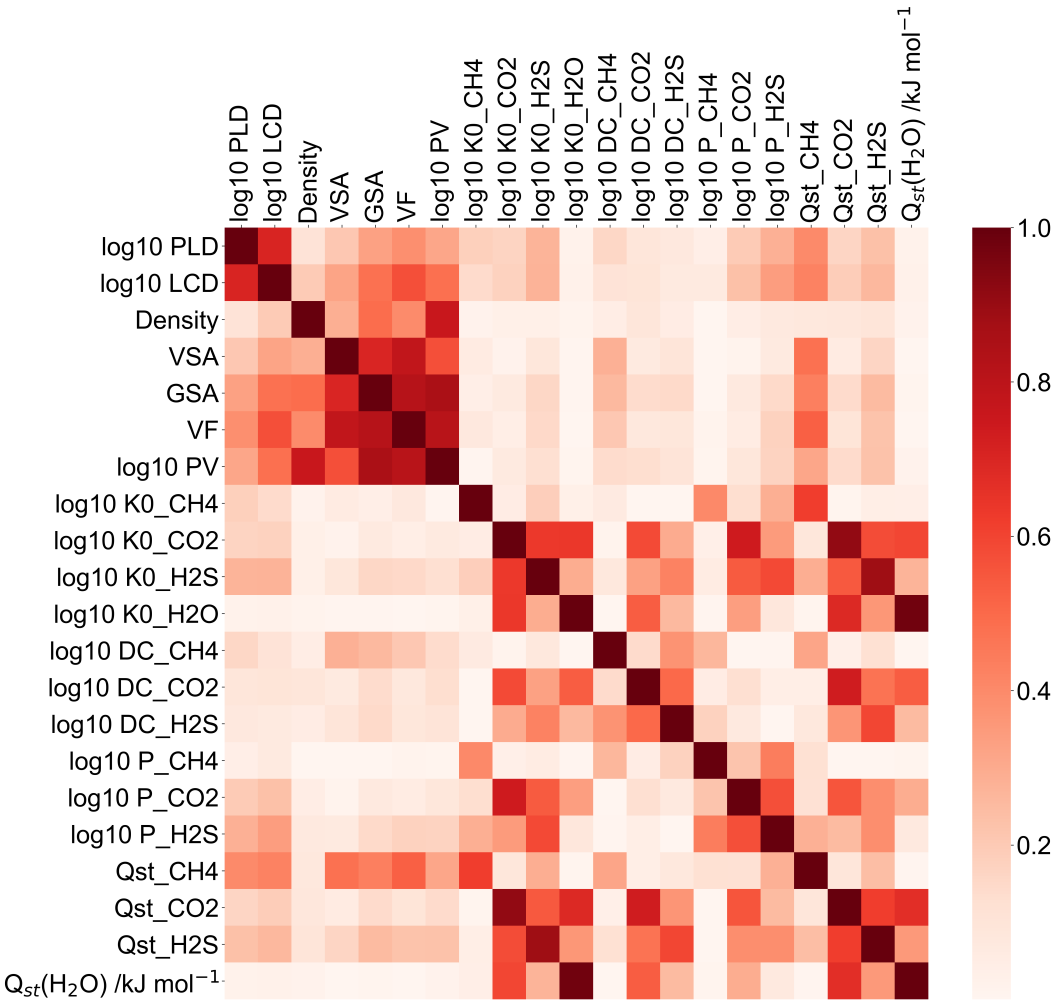

Figure 4.6: *Correlation heatmap of all 21 features available from the work of Glover and Besley, [18] with the log_{10} operator applied where relevant.*

which are strongly correlated bring very similar statistical information to a machine learning model. It is not useful to include both in training, and correlation between features can be used to inform feature selection. Figure 4.6 shows a heatmap of R^2 values between each pair of descriptors.

It is clear that some categories of descriptor display high correlation among their members. Geometrical descriptors tend to correlate highly with other geometrical descriptors. In particular, void fraction and pore volume display high correlation, which is no surprise as both measure the volume available inside a MOF. Void fraction and pore volume also display high correlation with gravimetric surface area. The correlation among these highly correlated structural features is illustrated in Figure 4.7, which contains plots of log_{10}(PV) against VF, GSA against log_{10}(PV) and GSA against VF. Similarly, correlation

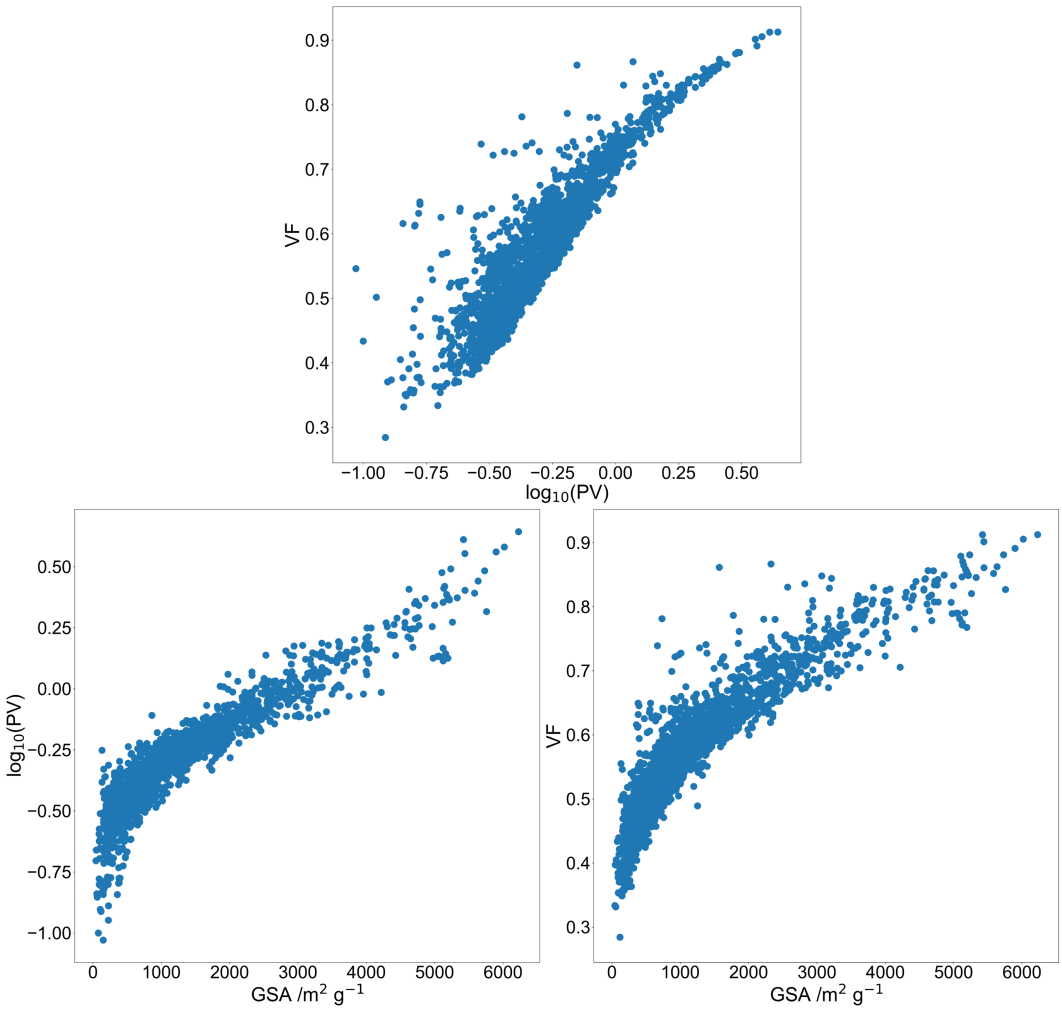

Figure 4.7: *Correlation plots for pore volume (log scaled) against void fraction, gravimetric surface area against pore volume (log scaled) and gravimetic surface area against void fraction in the training set. Data is taken from the work of Glover and Besley. [18]*

is high between Henry constants and heats of adsorption for the four gases for which data was available (that is, the Henry constant of CO_2 correlates with the heat of adsorption of CO_2, and the same is true of CH_4, H_2O, and H_2S). This again is in line with intuition as both are measures of the affinity of a guest gas for a framework at infinite dilution. The correlation among these highly correlated energetic interaction features is illustrated in Figure 4.8, which includes plots of Henry constant against heat of adsorption for CO_2, CH_4, H_2O and H_2S.

It was possible to eliminate selected descriptors from among those identified as highly correlated based on this analysis. From the geometrical de-

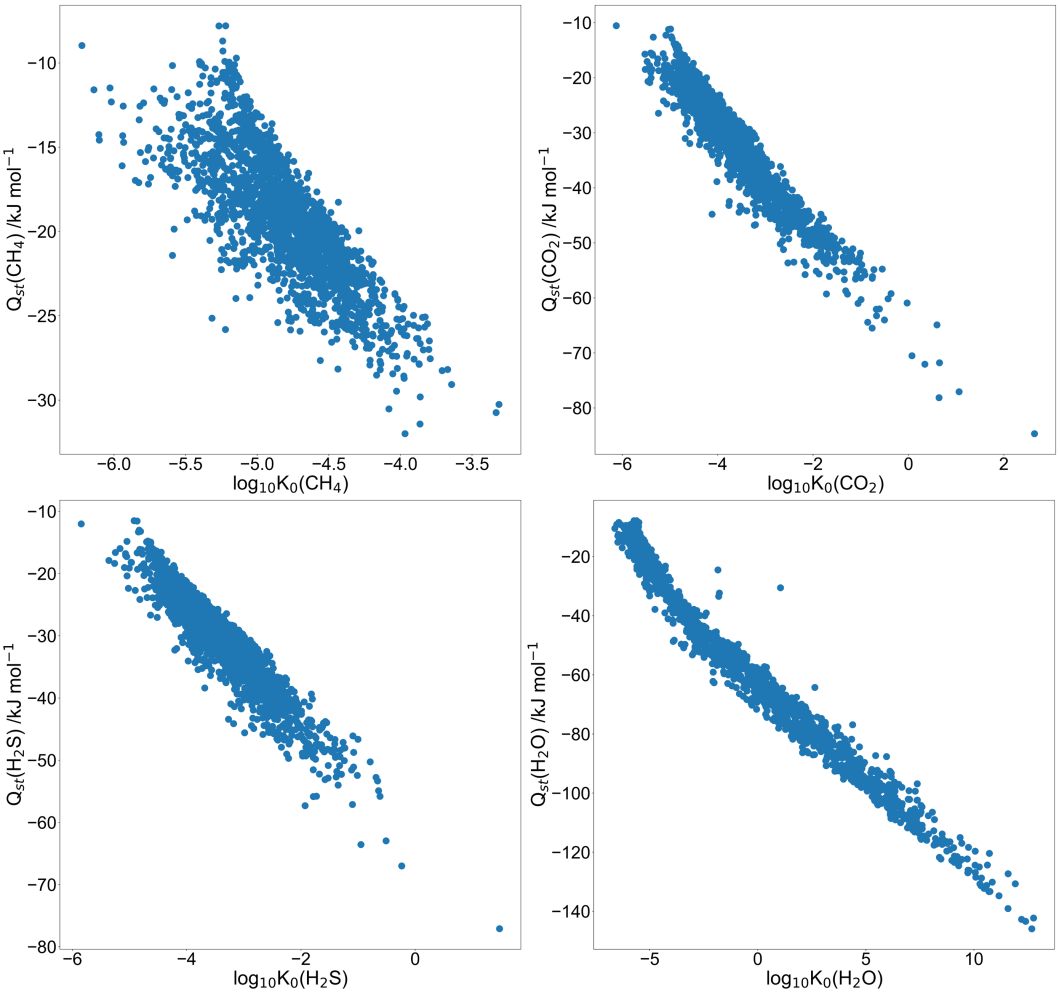

Figure 4.8: *Correlation plots of infinite dilution Henry constant (log scaled) against infinite dilution heat of adsorption for CH_4, CO_2, H_2S and H_2O in the training set. Data is taken from the work of Glover and Besley. [18]*

scriptors, pore volume and gravimetric surface area were removed, with void fraction and volumetric surface area retained as surface area and volume measures. Although PLD and LCD represent similar structural features, both were retained because pore morphology, LCD/PLD, is defined by a combination of the two and has also been seen to be important to gas uptake processes (see chapter 3). From the highly correlated Henry constants and heats of adsorption, Henry constants were discarded and heats of adsorption retained.

Finally, descriptors were selected to discard based on their relevance to the problem at hand. The diffusion coefficients and permeabilities calculated by Glover and Besley [5], although highly relevant to membrane studies, are less likely to have predictive power for purely uptake considerations. All were removed. This leaves a total of 9 descriptors selected as features for training ML models to predict TSN_{CO_2/CH_4} and related metrics. There remained in the feature list heat of adsorption values for all four gases considered by Glover and Besley, [18] including those for H_2O and H_2S which are not accounted for in the biogas upgrading metrics which this work attempts to predict. These features were not removed, as it is possible that the interactions they describe may be relevant to the MOF behaviour sought. For example, the heat of adsorption of water can give a strong indication of the polarity of a MOF, which can relate to its adsorption and selectivity. The 9 features, which are shown in bold in Table 4.1, were retained as a reasonable number of relevant and uncorrelated features.

4.3.3 Trained Machine Learning Models and their Predictions

Validation Set

Machine learning models were trained to address both the regression and the classification problems laid out in this chapter. Three different kinds of models were trained for each type of problem, enabling comparison between the mod-

els and judgement of the most appropriate models for the situation. For each of the six regression problems (prediction of TSN_{CO_2/CH_4}, $log_{10}TSN_{CO_2/CH_4}$, $N_{CO_2}(bm)$, $N_{CH_4}(bm)$, $N_{CO_2}(sc)$ and $N_{CH_4}(sc)$) the models trained were multiple linear regression (MLR), support vector machine (SVM) and random forest (RF). For the TSN_{CO_2/CH_4} classification problem, the models trained were k-nearest neighbour (kNN), SVM and RF.

First, the regression models are considered, and their performance on the validation sets engendered by k-fold cross validation is assessed. Methods to assess the performance of ML models were introduced in section 2.4. The performance of regression models can be assessed using coefficient of determination (R^2) and mean absolute error (MAE) to measure how closely on average predicted values resemble true values. The coefficient of determination is analogous to the Pearson's R^2 used in section 4.3.2 for simple linear regression. For more complex methods it is similar, but is not identical and is no longer strictly limited to falling between 0 and 1. The performance metrics R^2 and MAE are summarised for the validation sets in Table 4.2 with associated errors, along with the standard deviation of each target value.

Varying performance is seen among the models. Higher R^2 values indicate stronger performance. As for the MAE, the lower the value, the better the performance, and values of MAE lower than the standard deviation can as a general rule be said to indicate good performance. All of the models pass this test, even the benchmark MLR models, which display lower R^2 and higher MAE in every case than their more advanced non-linear counterparts. The two higher-quality models, SVM and RF, display similar performance over all measures. Considering the different target values, the raw uptake labels are well-predicted, with R^2 values above 0.8 for all SVM and RF models, and reasonable R^2 values for the MLR models. The best performance is seen for single component CO_2 uptake. The TSN labels, although displaying reasonable R^2 and MAE measures using SVM and RF, are less well predicted by the regression models than the uptake labels, with R^2 below 0.5 using MLR.

Table 4.2: Summary of R^2 and MAE performance measures for the regression models trained for each of the six target values using each of the three regression methods, multiple linear regression (MLR), support vector machine (SVM) and random forest (RF). Also shown is the standard deviation (SD) of the target values. Data used to create the table was produced as part of this work and analysed by Dr Samuel Boobier.

	Target	R^2	Error	MAE	Error	SD
MLR	TSN_{CO_2/CH_4}	0.479	0.120	1.600	0.078	4.022
	$log_{10}TSN_{CO_2/CH_4}$	0.494	0.076	0.146	0.009	0.340
	$N_{CO_2}(sc)$	0.880	0.015	1.060	0.062	4.463
	$N_{CO_2}(bm)$	0.792	0.042	0.888	0.061	2.944
	$N_{CH_4}(sc)$	0.699	0.019	0.649	0.034	1.785
	$N_{CH_4}(bm)$	0.764	0.031	0.238	0.010	0.717
SVM	TSN_{CO_2/CH_4}	0.711	0.071	1.115	0.114	4.022
	$log_{10}TSN_{CO_2/CH_4}$	0.829	0.028	0.096	0.005	0.340
	$N_{CO_2}(sc)$	0.955	0.007	0.621	0.047	4.463
	$N_{CO_2}(bm)$	0.902	0.016	0.601	0.033	2.944
	$N_{CH_4}(sc)$	0.896	0.024	0.372	0.028	1.785
	$N_{CH_4}(bm)$	0.941	0.018	0.116	0.008	0.717
RF	TSN_{CO_2/CH_4}	0.727	0.052	1.147	0.089	4.022
	$log_{10}TSN_{CO_2/CH_4}$	0.804	0.049	0.091	0.006	0.340
	$N_{CO_2}(sc)$	0.953	0.008	0.612	0.037	4.463
	$N_{CO_2}(bm)$	0.901	0.013	0.595	0.040	2.944
	$N_{CH_4}(sc)$	0.872	0.016	0.403	0.019	1.785
	$N_{CH_4}(bm)$	0.926	0.021	0.125	0.010	0.717

Application of the log_{10} operator to the TSN_{CO_2/CH_4} data engenders only a small improvement in the predictions.

Within the RF framework, it is possible to assign importance to features based on which were assigned the highest weights during the training process. The features with the highest weights have the largest impact on predictions for new MOFs. The magnitude of weights can indicate possible strength of relationships between features and labels, although a feature with a large weight is not definitively known to have a stronger relationship with the label than another feature with a smaller weight; it is simply put to greater use by the particular model, and may be put to less use by a model under slightly altered circumstances. The weights assigned to each feature by the RF models presented here are shown in Figure 4.9a. Presented weights are mean weights over all cross-validation folds, and error bars are standard deviations. The most important features, particularly for the TSN targets, are seen to be a mixture of structural and infinite dilution energetic features, supporting evidence from previous studies. [187, 242] For the uptake targets, however, it is structural features which dominate, and energetic features have a relatively small contribution to the models. Predictions of CH_4 loading in both the binary mixture and the single component case display strong dependence on density, while predictions of binary mixture and single component CO_2 loading depend particularly on void fraction. It may be supposed that the dependence on void fraction suggests that for many MOFs at 10 bar sufficient adsorption has occurred that the space available in the MOF is a limiting factor in determining whether new gas molecules can be introduced. To probe this further, void fraction is plotted against calculated (GCMC) $N_{CO_2}(sc)$ in Figure 4.9b, which shows a strong relationship between the two in line with the importance of the VF feature. There is an increase in CO_2 loading with void fraction up to a peak which is not entirely well-defined, and then some decrease once a void fraction of around 0.8 is reached. As previously mentioned, the increase of space within a MOF engendered by a large void fraction facilitates increased uptake up to

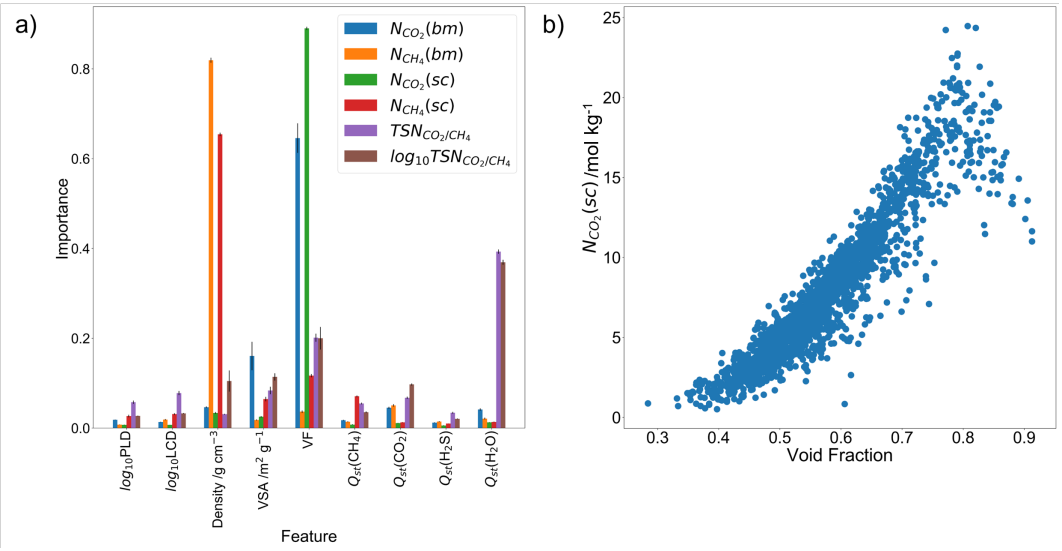

Figure 4.9: *a) Bar chart showing mean importance weights from RF of each feature for the six regression labels, with standard deviation as error bars. b) Plot of void fraction (VF) against $N_{CO_2}(sc)$ as calculated by GCMC. Data used to create the models for (a) was produced as part of this work and analysed by Dr Samuel Boobier.*

a certain point, but it also leads to an increase in pore size and reduction of interactions with pore walls. When pores are large enough, interactions with pore walls are few and uptake decreases.

Between the energetic features, the contribution of $Q_{st}(H_2S)$ is among the smallest. This may be considered to be in line with intuition as H_2S is neither one of the molecules studied nor displays any particular similarity to them. It is interesting, meanwhile, that the energetic feature displaying the most overall importance for predicting the target values is $Q_{st}(H_2O)$, which also corresponds to neither of the studied gas molecules. However, $Q_{st}(H_2O)$ can provide a description of the hydrophilicity and polarity of a structure. Such a description appears able to account for the selectivity contribution to TSN, with highly polar and hydrophilic MOFs likely to unite affinity for CO_2 with lack of affinity for CH_4. In any case, $Q_{st}(H_2O)$ provides a more important description of TSN_{CO_2/CH_4} to the RF models than the heat of adsorption of either of the two individual gases for both the regression and the classification models.

Predictions of the trained RF models are further assessed to examine their ability to predict values of the labels in different regions. In Figure 4.10, calcu-

lated values of each label in the training and validation set are plotted against corresponding predicted values. Fairly strong linear fits with a few outliers are seen for all of the uptake labels; the label $N_{CO_2}(sc)$, which has already been seen to display the strongest performance metrics, has visibly the tightest fit. For TSN_{CO_2/CH_4}, the fit is quite good at lower total values, but deteriorates for the higher values, with deterioration beginning between 5 and 10 and increasing to the point that very high values of TSN_{CO_2/CH_4} around 30 mol kg^{-1} are predicted to be only around 15 mol kg^{-1}. The training and validation sets are poorly populated for values this high, which explains the model's inability to obtain highly accurate predictions in this region. This could be particularly problematic, as the MOFs which researchers seek are those displaying high TSN_{CO_2/CH_4}, so it is desirable for these to be well predicted. The performance plot for $log_{10}(TSN_{CO_2/CH_4})$ displays some improvement in linearity compared to the TSN_{CO_2/CH_4} plot, but it retains outliers and the improvement is exaggerated by scale. As mentioned previously, it only translates to a small improvement in the performance metrics shown in Table 4.2.

The lack of highly accurate predictions for high-TSN MOFs does not mean that a machine learning model cannot be useful in helping to identify promising structures. It is clear from Figure 4.10 that the MOFs with the very highest TSN_{CO_2/CH_4} are in general given predictions which, while not as high as their calculated values, are still reasonably high. This lends support to the use of classification models to determine whether MOFs are likely to be high-performing or low-performing. If a MOF with a calculated TSN_{CO_2/CH_4} of 30 is predicted to have a TSN_{CO_2/CH_4} of 15 mol kg^{-1} and thereby classified as high-performing, a classification model designed to be used in the early stages of a high-throughput screening has performed well for that structure.

The performance of classification models can be measured using precision and recall, as introduced in section 2.4. Precision measures the proportion of points predicted to fall into a class that do fall into that class according to their true values, and recall measures the proportion of points that fall into a class

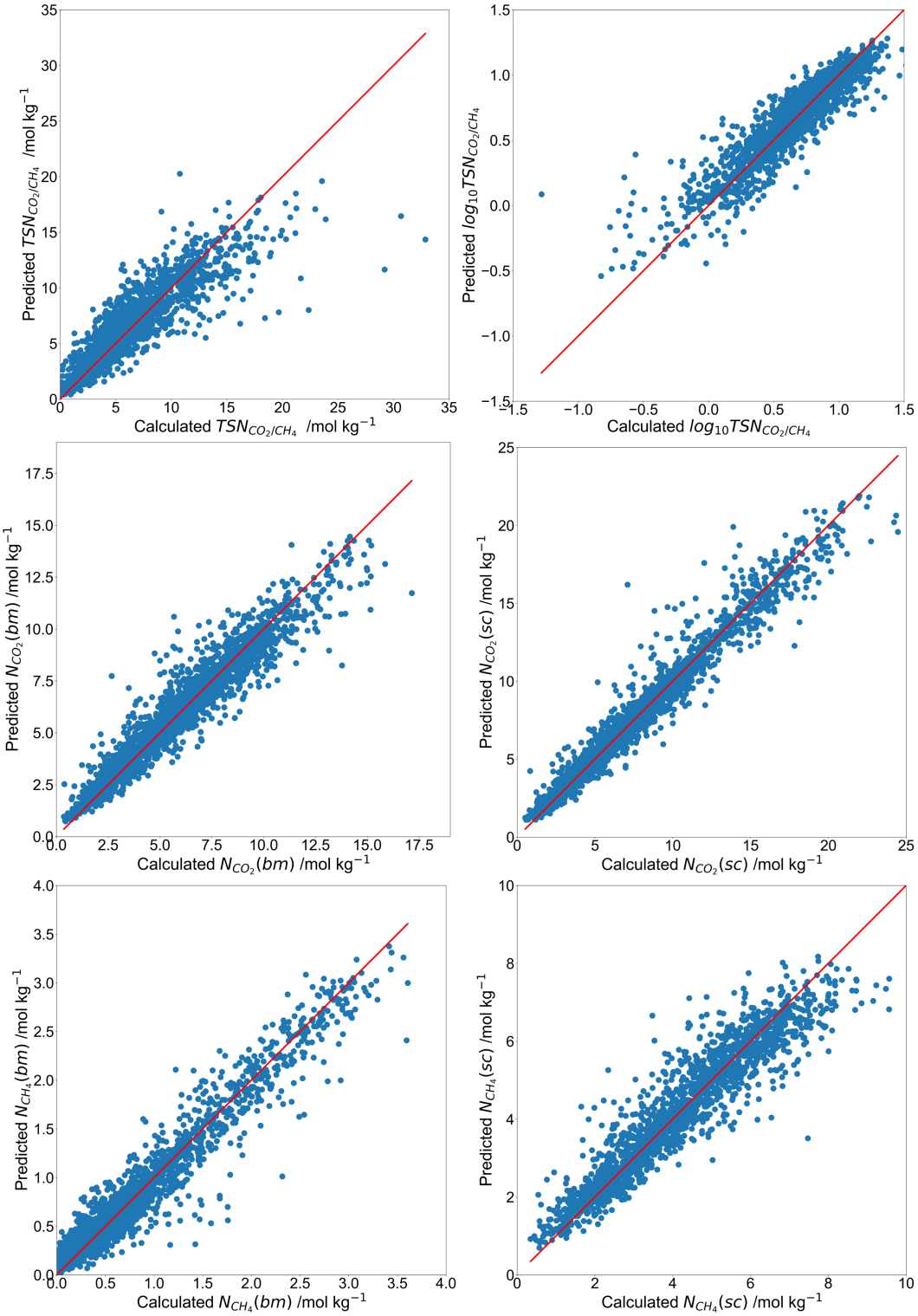

Figure 4.10: *Calculated values of each of the six labels used for regression, TSN_{CO_2/CH_4}, $log_{10}(TSN_{CO_2/CH_4})$, $N_{CO_2}(bm)$, $N_{CH_4}(bm)$, $N_{CO_2}(sc)$ and $N_{CH_4}(sc)$ from the training and validation set plotted against corresponding values predicted using the trained random forest model. All values are for performance at 10 bar and 298 K. Data used to create the models for predicted values was produced as part of this work and analysed by Dr Samuel Boobier.*

Table 4.3: Precision, recall, and their associated errors for the high-performing and low performing classes as classified by k-nearest neighbour (kNN), support vector machine (SVM) and random forest (RF). Data used to create the table was produced as part of this work and analysed by Dr Samuel Boobier.

	Metric	High-performing	Error	Low-performing	Error
kNN	Precision	0.868	0.024	0.886	0.043
	Recall	0.900	0.037	0.851	0.021
SVM	Precision	0.872	0.028	0.874	0.023
	Recall	0.887	0.025	0.859	0.024
RF	Precision	0.888	0.030	0.907	0.032
	Recall	0.919	0.027	0.874	0.022

according to their true values that were predicted to fall into that class. A binary classification involves two classes, both of which have a precision and a recall. Precision and recall (and their associated errors, calculated by standard deviation over the 10 folds) are given in Table 4.3 for kNN, SVM and RF models trained to classify MOFs as high-performing or low-performing based on the TSN_{CO_2/CH_4} threshold of 5 mol kg^{-1}. For example, 86.8% of MOFs predicted to be high-performing by the kNN model truly have calculated TSN_{CO_2/CH_4} greater than 5 mol kg^{-1}, and 90.0% of MOFs whose calculated TSN_{CO_2/CH_4} is greater than 5 mol kg^{-1} were predicted to be high-performing by the same model.

Both precision and recall can range between 0 and 1, and values closer to 1 indicate better performance. All of the trained models perform well for the validation data, with both precision and recall exceeding 0.85 for both classes, meaning the clear majority of MOFs are placed in the correct categories. The random forest model displays the best performance, having the highest precision and recall for both classes. It is therefore carried forwards for further evaluation. The importance weights for each feature, like those in Figure 4.9 for the regression model, are shown in Figure 4.11. The weights are such that importance is spread among the 9 features, with $Q_{st}(H_2O)$ contributing the most to the model, followed by VF, which previously displayed the highest performance for selected regression models.

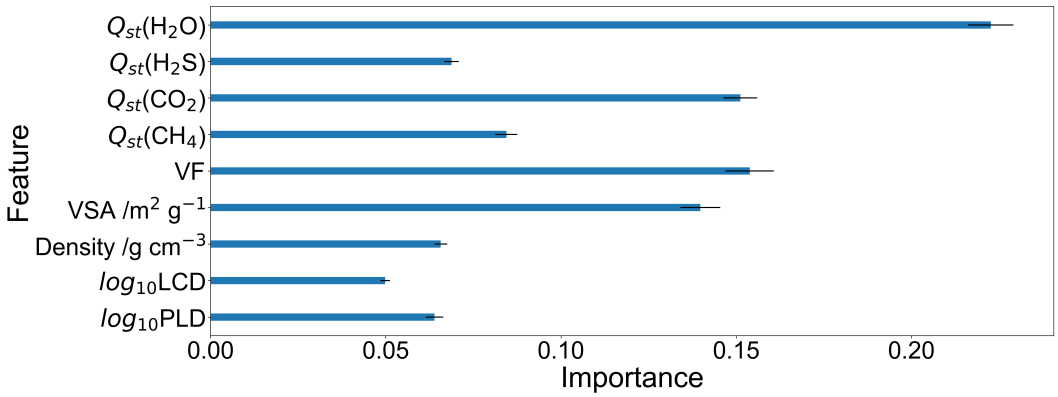

Figure 4.11: *Bar chart showing the importance weights for all features for the random forest classification model. Values are mean weights over the 10 folds, and error bars are standard deviations. Data used to create the figure was produced as part of this work and analysed by Dr Samuel Boobier.*

For a situation like the present one where a continuous value is converted into classes, it is not only the number of structures which are correctly classified that determines the quality of a model, but also which structures are incorrectly predicted. If a model predicts extremely high-performing MOFs to be low performing, and extremely low-performing MOFs to be high-performing, it is less useful than one with the same precision and recall that predicts MOFs whose TSN_{CO_2/CH_4} is slightly above the threshold to be low performing and those whose TSN_{CO_2/CH_4} is slightly below to be high performing. A screening is less likely to be negatively impacted by erroneous exclusion of MOFs with performance metrics close to the threshold than erroneous exclusion of MOFs with performance metrics far from the threshold. Figure 4.12a displays distributions of GCMC-calculated TSN_{CO_2/CH_4} among the MOFs classed as high-performing and those classed as low performing by the RF classification model, along with a vertical line indicating the threshold. It confirms that the majority of erroneously classified MOFs in both classes have calculated TSN_{CO_2/CH_4} close to the threshold and therefore that the model is unlikely to erroneously classify MOFs lying at the extremes of the TSN_{CO_2/CH_4} range.

The quality of the model can also be assessed by the confidence it places in its classification of different structures. The classification RF model functions by assigning a probability of being high-performing to each structure

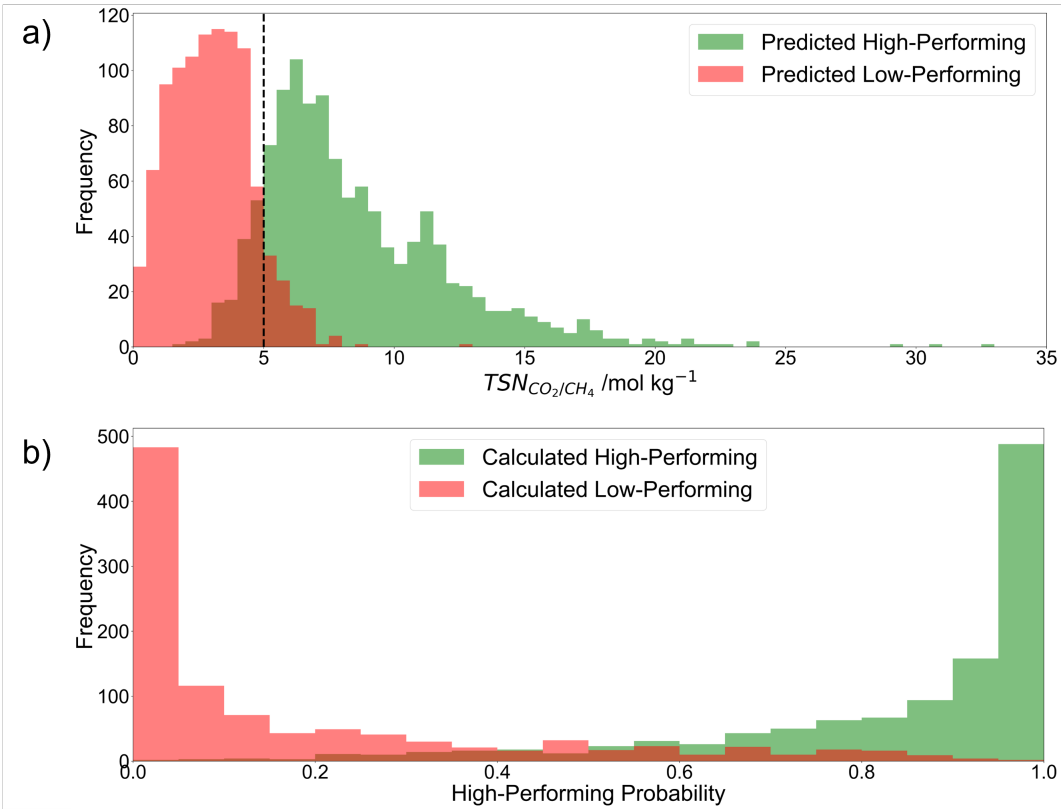

Figure 4.12: *Histograms relating to performance evaluation of the RF classification model for the training and validation set: a) Distributions of* TSN_{CO_2/CH_4} *for the MOFs classed as high-performing and low-performing by the RF model b) Distributions of high-performing probability for the MOFs classed as high-performing and low-performing by the GCMC data. Data used to create the figure was produced as part of this work and analysed by Dr Samuel Boobier.*

and classifying all MOFs with a high-performing probability less than 0.5 as low-performing and those with a high-performing probability greater than or equal to 0.5 as high-performing. A strong model will assign high-performing probabilities close to 1 to correctly-classified high-performing MOFs, and probabilities close to 0 to correctly-classified low-performing MOFs, while MOFs which are erroneously classified will be assigned high-performing probabilities close to 0.5. Figure 4.12b confirms that in the cross-validation set this is generally the case for the RF model fitted here by showing the distribution of the high-performing probabilities of MOFs which were placed in each class based on GCMC data. Most MOFs for which predictions are confident are correctly classified, with only very few MOFs classified erroneously with high probability.

In addition to analysis on a statistical scale, the performance of models of both kinds was assessed on an individual-MOF level by analysis of the structural features of outlying MOFs. The random forest model for $N_{CO_2}(sc)$ is taken as an example of the best-performing regression model. Calculated and predicted loading of the six MOFs with the highest absolute prediction error are shown in Table 4.4, along with relevant structural features, VF, which showed the highest importance in the model, and PLD. To ensure any uniting factors observed among the six were not simply artifacts of the dataset as a whole, the six structures with the lowest prediction errors are also given in Table 4.4. The majority of the largest errors correspond to an overprediction of loading rather than an underprediction, and are associated with large-pored structures displaying high void fraction (at least 0.74), while the majority of the best-predicted structures display low void fraction and smaller PLD than the poorly predicted MOFs. The prevalence of high void fractions among the poorly-predicted structures likely relates to the peak in the relationship between void fraction and calculated $N_{CO_2}(sc)$ seen in Figure 4.9. There are few structures with VF larger than that of the peak, so the model may have received insufficient training data in this region to effectively learn the dip in

Table 4.4: Details of the six MOFs with the highest absolute prediction error and the six MOFs with the lowest absolute prediction error from the RF regression model for $N_{CO_2}(sc)$, along with their calculated and predicted loading values, prediction error, pore limiting diamter (PLD) and void fraction (VF).

	Refcode	Calculated Loading /mol kg^{-1}	Predicted Loading /mol kg^{-1}	Prediction Error /mol kg^{-1}	PLD Å	VF
Large Error	LAWGEW	7.09	16.20	+9.11	6.78	0.74
	OWITIY	13.89	19.91	+6.02	6.78	0.77
	JUTCIM	8.82	14.52	+5.70	14.88	0.74
	TUWVEO	12.02	17.59	+5.57	9.32	0.83
	HAYZAL	17.82	12.27	−5.55	18.55	0.78
	IPUPIU	9.66	14.76	+5.10	8.73	0.75
Small Error	SOTWOP	10.52	10.52	0	5.19	0.66
	VETMIS	15.76	15.76	0	11.95	0.86
	HIFTOG	8.27	8.27	0	4.14	0.63
	WOFSIU	4.01	4.01	0	5.74	0.51
	BECTAH	5.05	5.05	0	5.74	0.51
	DOHDAH	6.72	6.71	+0.01	4.71	0.55

performance when VF is very high, leading to overprediction.

Outliers of the classification model are also identified. They include both structures with high TSN_{CO_2/CH_4} which were incorrectly classified as low, and structures with low TSN_{CO_2/CH_4} which were incorrectly classified as high. Details of the six worst-classified structures of each category are given in Table 4.5, with the worst-classified structures defined as those which are incorrectly classified and whose calculated TSN_{CO_2/CH_4} is furthest from the threshold of 5 mol kg^{-1} (highest value for those incorrectly classified as low and lowest value for those incorrectly classified as high). Again, the best-classified MOFs are also included in the table for comparison, with the best-classified MOFs defined as those which are correctly classified with the highest confidence (highest probability for high-performing MOFs and lowest probability for low-performing MOFs). For the training and validation data several MOFs are classified with a probability of 1 or 0, so to obtain a range of well-classified MOFs, two sets of correctly classified high-performing MOFs are listed. These are the 6 correctly classified as high-performing with a probability of 1 and the

highest GCMC TSN_{CO_2/CH_4} and the six correctly classified as high performing with a probability of 1 and the lowest GCMC TSN_{CO_2/CH_4}. A similar two sets of correctly classified low-performing MOFs are listed.

Several observations can be made relating to Table 4.5. For one thing, of the 12 selected incorrectly classified MOFs with the largest distances from the TSN_{CO_2/CH_4} threshold, most are classified with probabilities fairly close to 0.5. Along with Figure 4.12, this highlights the positive fact that the model is not confidently wrong about values of TSN_{CO_2/CH_4} at either extreme. In a similar vein, the very confidently classified MOFs do not include either low performers or high performers whose values of TSN_{CO_2/CH_4} are very close to 5, showing that it is *not* highly confident about borderline MOFs. Among the high-performing MOFs classified with a probability of 1 is QUDJEF, a structure that was identified as among the highest performers across the whole dataset in Figure 4.4. It was confirmed that the other 6 top performers identified in Figure 4.4 were also correctly classified by the model, all with high-performing probability above 0.9 with the exception of YOCSEQ (probability=0.67).

In terms of structural features among the well-classified and poorly-classified MOFs, PLD and VF vary significantly among incorrectly-classified MOFs of both kinds. High-performing MOFs which were well classified all have VF within a similar range, close to 0.6, and well-classified low-performing MOFs have smaller VFs in the range 0.4-0.55. The poorly classified MOFs do not have VFs which fit as neatly into this pattern. In particular, low-performing MOFs which were incorrectly classified as high-performing tend to have larger VF than well-classified low-performing MOFs. A similar effect appears to be seen here for the relationship between VF and TSN_{CO_2/CH_4} as was seen in Figure 4.9b for the relationship between VF and N_{CO_2}.

External Test Set

The models have been seen to perform well for the training and validation sets and are now moved on to external testing. The performance of the TSN

Table 4.5: Details of the MOFs which are worst-classified and best-classified by the RF TSN_{CO_2/CH_4} model. Details are: Refcode, TSN_{CO_2/CH_4}, High-performing probability, pore limiting diameter (PLD) and void fraction (VF). Within categories, incorrectly classified MOFs are selected in order of distance from the TSN_{CO_2/CH_4} threshold and correctly classified MOFs are selected in order of the high-performing probability. Since several correctly classified MOFs have a probability of 0 or 1, they are selected first in descending order and then in ascending order of TSN_{CO_2/CH_4}.

	Refcode	TSN /mol kg^{-1}	Probability	PLD /Å	VF
	KEFZEC	12.87	0.318	4.06	0.48
High	WITLUI	8.76	0.366	4.54	0.51
TSN_{CO_2/CH_4},	POTREX	7.81	0.4	10.80	0.79
incorrectly	VULKOD	7.80	0.202	3.95	0.45
classified	MIJSII	7.74	0.408	6.51	0.53
	KEQJEX	7.55	0.376	5.17	0.53
High	QUDJEF	30.67	1	4.27	0.70
TSN_{CO_2/CH_4},	COTVIS	18.96	1	4.32	0.62
correctly	PASMUT	18.87	1	3.99	0.57
classified,	SIKYOB	18.06	1	4.44	0.69
maximise	QISNEN	17.83	1	5.40	0.73
TSN_{CO_2/CH_4}	MOHLIF	17.49	1	4.67	0.64
High	UKAQAZ	6.45	1	4.81	0.59
TSN_{CO_2/CH_4},	JURYOM	6.64	1	4.68	0.58
correctly	UZIJUJ	6.80	1	5.85	0.61
classified,	KOCPOI	7.09	1	4.68	0.57
minimise	YEBGOD	7.27	1	4.90	0.62
TSN_{CO_2/CH_4}	ZEZPUR	8.21	1	4.74	0.57
	OWITIY	1.96	0.554	5.80	0.77
Low	PUZLUZ	2.32	0.504	8.40	0.71
TSN_{CO_2/CH_4},	IPUPIU	2.46	0.808	8.73	0.75
incorrectly	PURQOJ	2.57	0.664	6.85	0.69
classified	BIPKUI	2.73	0.502	11.99	0.58
	GITVIP01	2.95	0.724	5.47	0.63
Low	ZIVDIT	0.47	0	4.58	0.44
TSN_{CO_2/CH_4},	OSIWOD01	0.52	0	4.04	0.44
correctly	NENVAE	0.37	0	4.66	0.39
classified,	VULRAW	0.62	0	4.57	0.45
minimise	ZIVFIV	0.68	0	4.59	0.45
TSN_{CO_2/CH_4}	MABKEG	0.70	0	4.29	0.42
Low	TEQTEQ	3.51	0	5.80	0.52
TSN_{CO_2/CH_4},	ADODII	3.30	0	3.52	0.54
correctly	OKIRIK	3.30	0	5.14	0.54
classified,	IVITOY	3.29	0	4.17	0.46
maximise	LAGDAB	3.23	0	6.01	0.51
TSN_{CO_2/CH_4}	PEKTOQ	3.14	0	4.18	0.48

Table 4.6: Summary of R^2 and MAE performance measures for the three regression models trained to predict $N_{CO_2}(sc)$ at 10 bar and 298 K and tested on the external test set. Data used to create the table was produced as part of this work and analysed by Dr Samuel Boobier.

Method	R^2	MAE	SD
MLR	-0.256	3.027	3.798
SVM	0.116	2.841	3.798
RF	-0.077	2.264	3.798

classification models and the performance of the best-performing regression models, those for $N_{CO_2}(sc)$, were further tested on the external test set of 331 hypothetical MOFs taken from the Northwestern Hypotheticals database. [33] For comparison to predicted values, relevant calculated values were obtained for the members of the test set using the methods outlined in section 4.2.2.

The performance metrics for the $N_{CO_2}(sc)$ regression models are summarised in Table 4.6. The MAE remains lower than the standard deviation in all cases, which suggests a reasonable performance of the model, particularly for the RF model. However, the R^2 values for the external data are far less promising than the equivalent values for the training and validation set, all being much lower and some being negative, suggesting overfitting. It appears counter-intuitive that an R^2 value could be negative, but for cases more complex than simple linear regression, R^2 is not described simply by the square of a value, and negative values are possible, but not desired.

The mixed messages given by the performance metrics for the external test set are examined further. In Figure 4.13a, $N_{CO_2}(sc)$ calculated by GCMC for the external test is plotted against predictions of $N_{CO_2}(sc)$ made by the RF model, which displays the best performance as measured by MAE. Additionally, Figure 4.13b includes a plot of VF against $N_{CO_2}(sc)$ as calculated by GCMC among the external test set. VF is the descriptor which is weighted highest by the RF model, and variations in VF have a strong impact on predictions, so the relationship between VF and calculated loading in the unseen external data is highly relevant. Figure 4.13a shows that there is a strong rela-

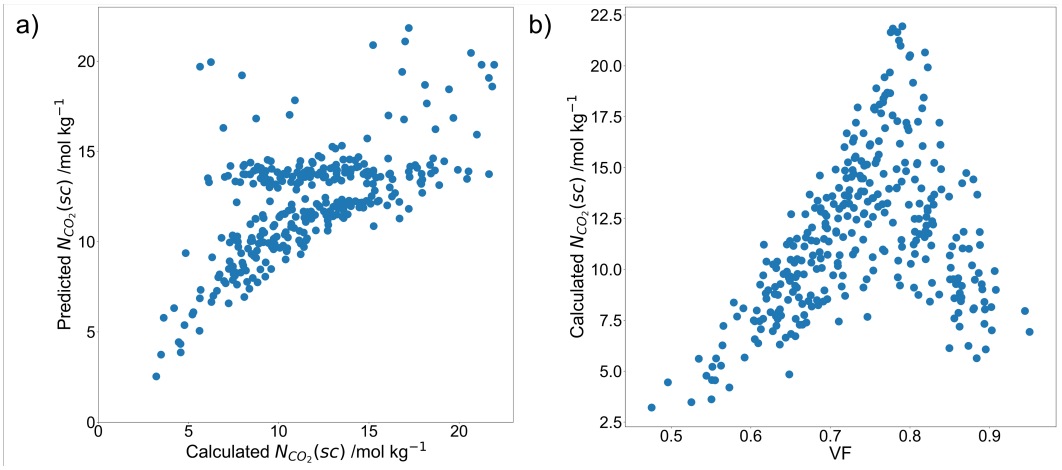

Figure 4.13: *a) Plot of calculated $N_{CO_2}(sc)$ against $N_{CO_2}(sc)$ predicted by the RF regression model for the external test set. b) Plot of void fraction against calculated $N_{CO_2}(sc)$ for the external test set. Data used to create the model used in (a) was produced as part of this work and analysed by Dr Samuel Boobier.*

tionship between calculated and predicted values for a selection of the MOFs up to some peak, but that significant numbers of outliers have overpredicted $N_{CO_2}(sc)$ and that for a selection of MOFs with varying calculated $N_{CO_2}(sc)$, the predicted value is around 15 mol kg^{-1}. It therefore appears there may be a particular type of MOF present in the test set which is not well-predicted by the model, while MOFs of other kinds are.

Interestingly, the shape of the plot in Figure 4.13a resembles the (reflected) shape of the VF against calculated $N_{CO_2}(sc)$ plot in Figure 4.13b. Having been assigned the highest weight, void fraction is likely to play a significant part in determining the quality of the predictions. The plot of VF against calculated $N_{CO_2}(sc)$ in figure 4.13b itself somewhat resembles the similar plot for the training and validation data given in Figure 4.9b, although there are (proportionally) many more points with VF higher than the $N_{CO_2}(sc)$ peak around 0.8 for the test data, and they reach a higher maximum VF. The test set covers this particular region of feature space to a greater extent than the training set does. This offers an explanation for the strong performance of the model for some MOFs compared to its poor performance for others: the poorly predicted MOFs are likely those with particularly high void fractions, which lead to a comparatively low loading that has not been seen by the model in

training. The model trained here, while effective for MOFs of certain kinds, is insufficient to model with high accuracy the GCMC-calculated properties of MOFs occupying the entire feature space of hypothetical datasets. It may be possible to improve the model by judicious addition of training data.

Meanwhile, the difference between the real curated training set and the hypothetical curated test set that is highlighted here is an interesting point for future consideration; detailed study of the similarities between commonly used hypothetical and real sets of MOFs is merited. The benefits of hypothetical MOF databases are defined by the fact that they contain MOFs which have not been produced experimentally, so some level of dissimilarity between the two is advantageous. However, a database which significantly departs from structures similar to those known to be experimentally synthesisable risks containing sets of coordinates which are not useful to pursue.

The external test set was also used to assess the performance and level of overfitting for the classification models. The performance metrics precision and recall for each model are given in Table 4.7. All models retain a reasonable performance by most metrics, in particular displaying high recall of low-performing MOFs (meaning that a high proportion of the low-TSN_{CO_2/CH_4} MOFs are correctly classed as low-performing). Precision is reasonable for both classes, but recall of high-performing MOFs has suffered the most significant drop in performance compared to the same metric in the training and validation set, being below 0.5 for the non-linear models: a lower proportion of high-TSN_{CO_2/CH_4} MOFs are correctly classified as high-performing. The high low-performing recall comes at the expense of the high-performing recall: a threshold which classes more total structures as low-performing will both *correctly* class more structures as low-performing and *incorrectly* class more as low-performing. There would be scope to adjust all precision and recall values in relation to each other by adjustment of the TSN_{CO_2/CH_4} threshold.

The low recall for high-performing MOFs does not necessarily prevent the models from being useful in a high-throughput screening situation. If

Table 4.7: Precision, recall, and their associated errors for the high-performing and low performing TSN classes as classified by k-nearest neighbour (kNN), support vector machine (SVM) and random forest (RF) among the external test set. Data used to create the table was produced as part of this work and analysed by Dr Samuel Boobier.

	Metric	High-performing	Low-performing
kNN	Precision	0.776	0.790
	Recall	0.638	0.881
SVM	Precision	0.871	0.736
	Recall	0.469	0.955
RF	Precision	0.789	0.725
	Recall	0.462	0.920

the contribution of a machine learning model to a screening is to filter out low-performing structures to reduce the amount of expensive calculations that must be done on structures which are not of interest, then high recall of low-performing structures is particularly important. Of course, it is still desirable to ensure that those structures which are most promising are retained. Therefore, the question of *which* high-performing MOFs the model is able to correctly classify becomes important. If it incorrectly classifies structures whose TSN_{CO_2/CH_4} is close to the threshold, this is far more promising than if it incorrectly classifies those with extreme values of TSN_{CO_2/CH_4}.

With that in mind, the performance of the RF classification model on the external training set is assessed in more detail using histograms presented in Figure 4.14 as it was for the training and validation set (see Figure 4.12). Figure 4.14a contains histograms of the calculated TSN_{CO_2/CH_4} values for those MOFs which are classed as high-performing and those MOFs which are classed as low-performing by the RF model. There are proportionally fewer MOFs classed as high-performing for the external test set than there were for the training and validation sets (76 of the 331 MOFs), and this reflects calculated values of TSN_{CO_2/CH_4} in the external set, which have a lower maximum and with fewer MOFs whose GCMC TSN_{CO_2/CH_4} is above the threshold of 5 mol kg^{-1} (130 of the 331 MOFs). MOFs of both classes which are incor-

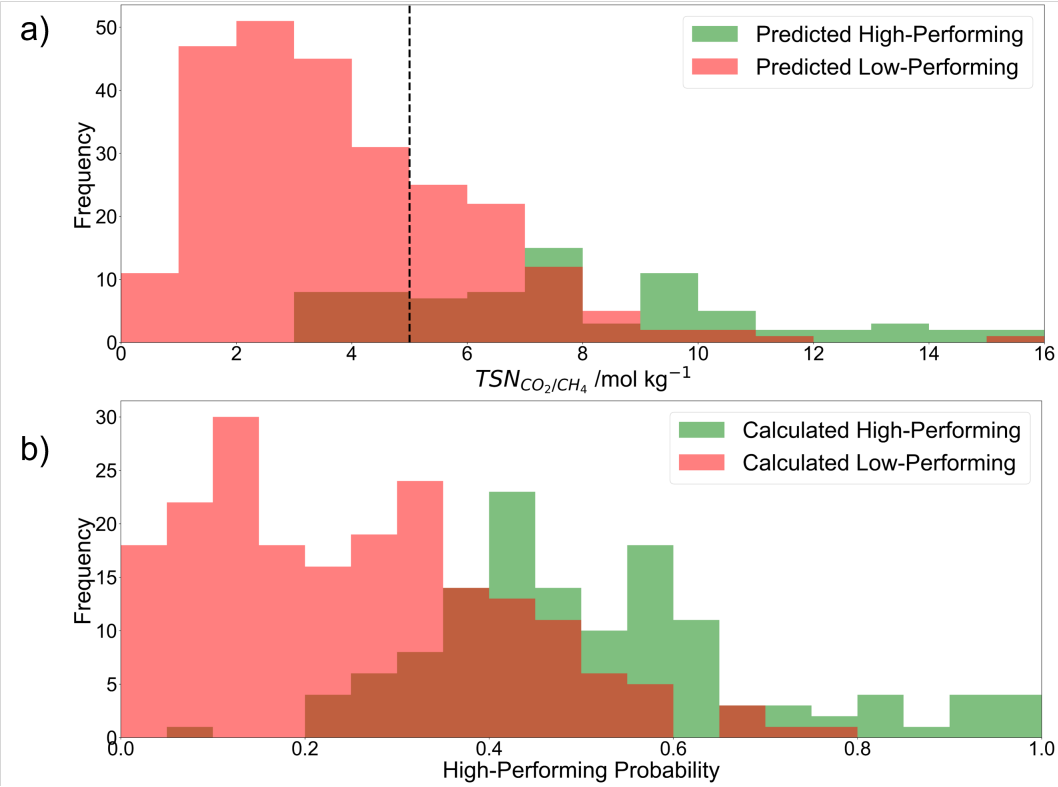

Figure 4.14: *Histograms relating to performance evaluation of the RF classification model for the external test set: a) Distributions of TSN_{CO_2/CH_4} for the MOFs classed as high-performing and low-performing by the RF model b) Distributions of high-performing probability for the MOFs classed as high-performing and low-performing by the GCMC data. Data used to create the figure was produced as part of this work and analysed by Dr Samuel Boobier.*

rectly classified again have calculated TSN_{CO_2/CH_4} close to the threshold in general, a promising sign for the potential of the model in high-throughput screening situations. There is, however, a tail of high-performing MOFs which are predicted to be low-performing, with one of the structures with the highest calculated TSN_{CO_2/CH_4} in the dataset being classed as low-performing by the model. This TSN_{CO_2/CH_4} is still significantly lower than the highest values observed in the training and validation set.

Figure 4.14b examines the confidence with which the RF model makes correct and incorrect classifications. Most of the MOFs whose calculated TSN_{CO_2/CH_4} indicates low performance are assigned low probability of high performance by the model, including several with probability close to 0. Few of these MOFs are assigned very high probabilities of high performance, with none assigned a probability above 0.8. The MOFs whose calculated TSN_{CO_2/CH_4}

indicates high performance are far more likely to be assigned probabilities close to 0.5, though some are very confidently correctly classed with probabilities close to 1, and very few are assigned probabilities very close to zero (confidently incorrectly classified).

The specifics of the outlying MOFs which are poorly predicted by the regression and classification models can be analysed for the test set as they were for the training set (see Tables 4.4 and 4.5). This can give more information about the kinds of MOFs in diverse datasets which are well predicted and poorly predicted by the trained models. Outliers from the RF regression model trained to predict $N_{CO_2}(sc)$ and from the RF classification model trained to classify MOFs according to TSN_{CO_2/CH_4} are taken as examples.

Table 4.8 includes details of the six MOFs from the test set whose $N_{CO_2}(sc)$ was predicted with the largest error by the RF model. It also includes details of the six MOFs whose predictions have the smallest error for comparison. In the Northwestern Hypotheticals dataset, each MOF is associated with a numerical identifier rather than the 6-letter refcode of the CSD. These identifiers are used in Table 4.8. The majority of the most poorly-predicted structures (5 of the 6) have their TSN_{CO_2/CH_4} overpredicted by a significant margin, larger than the margins seen for the training and validation set. These structures have very large pores. Several of them display pores of similar kinds, with large hexagonal structures common. They reach PLD as high as 30 and VF as high as 0.95. All 5 have VF as high as or higher than the value of around 0.8 at which calculated $N_{CO_2}(sc)$ is seen to stop increasing for both the test set and the training and validation set (see Figures 4.9b and 4.13b). They are in the region that is poorly represented in the training set and is therefore not well predicted, and they contrast markedly with the six well-predicted structures, all of which have much smaller internal space as measured by both PLD and VF, and which commonly have small square or rectangular channels. This is a similar effect to that seen in the training and validation sets, but is more marked for the test set which possesses a different distribution of structures. Since high

Table 4.8: The hMOF number codes of the six MOFs from the external test set with the highest absolute prediction error and the six MOFs with the lowest absolute prediction error from the RF regression model for $N_{CO_2}(sc)$, along with their calculated and predicted loading values, prediction error pore limiting diamter (PLD) and void fraction (VF).

	Numerical Identifier	Calculated Loading /mol kg^{-1}	Predicted Loading /mol kg^{-1}	Prediction Error /mol kg^{-1}	PLD Å	VF
	3000087	5.64	19.70	+14.06	10.63	0.88
	3000121	6.24	19.95	+13.71	10.46	0.87
Large	3001228	7.96	19.22	+11.26	23.70	0.94
Error	3001337	6.93	16.31	+9.38	30.55	0.95
	7002417	8.75	16.83	+8.08	7.63	0.78
	5035859	21.65	13.75	−7.9	9.58	0.78
	5035323	12.10	12.10	0	4.75	0.69
	5017848	14.02	14.01	−0.01	8.19	0.78
Small	5070732	4.45	4.44	−0.01	3.84	0.50
Error	5042760	9.66	9.70	+0.04	4.15	0.65
	5018534	11.59	11.64	+0.05	5.87	0.72
	14366	7.69	7.63	−0.06	4.99	0.58

loading is more likely to be sought than low loading, it is encouraging that the 6 best-predicted structures generally (but not exclusively) display higher calculated loading than the 6 worst-predicted.

To afford similar analysis for an example of a classification model, Table 4.9 gives details of the most incorrectly classified structures by the RF classification model among the test set. It includes details of the six MOFs with the highest calculated TSN_{CO_2/CH_4} which were incorrectly classified as low-performing and of the six MOFs with the lowest TSN_{CO_2/CH_4} which were incorrectly classified as high-performing. For comparison, it also includes the six MOFs of each class that were correctly classified with the highest confidence (high-performing probability closest to one for the high-TSN_{CO_2/CH_4} MOFs and high-performing closest to zero for the low-TSN_{CO_2/CH_4} MOFs). In this case there were not several MOFs classified with probabilities of one or zero, so only one set of six well-classified MOFs is used for each class. Most of the incorrectly-classified MOFs are classified with probabilities close to 0.5, some very close (one with a probability of 0.5 and another with a probability of 0.49). Meanwhile, the

MOFs correctly classified with high probability have TSN_{CO_2/CH_4} which is far from the threshold of 5 mol kg^{-1}. Indeed, the well-classified low-performers all have TSN_{CO_2/CH_4} less than 2. Unlike for the RF regression model, several of the well-classified low performing MOFs have very large pores as measured by both PLD and VF, with VF values towards the top of the possible range. It has been seen that very high VF can lead to lower uptake of CO_2 which is less well-represented in the training set than the test set. TSN_{CO_2/CH_4} is influenced both by CO_2 uptake and by CH_4 uptake. Large VF and PLD are likely to lead to situations in which both gases do not primarily interact with pore walls, so the uptake of both is less defined by competition between the two for binding sites. This will lead to higher CH_4 uptake relative to CO_2 uptake, a reduction in selectivity and a reduction in TSN_{CO_2/CH_4} which appears to have been identified by the RF classification model among the test set. This was not seen in outlier analysis for the training set, but that may be because the training set contained MOFs of other kinds which are absent in the test set and which the model could classify with even more confidence.

Table 4.9: Details of the MOFs from the external test set which are worst-classified and best-classified by the RF TSN_{CO_2/CH_4} model. Details are: Refcode, TSN_{CO_2/CH_4}, High-performing probability, pore limiting diameter (PLD) and void fraction (VF). Within categories, incorrectly classified MOFs are selected in order of distance from the TSN_{CO_2/CH_4} threshold and correctly classified MOFs are selected in order of the high-performing probability.

	Numerical Identifier	TSN /mol kg^{-1}	Probability	PLD Å	VF
	5013254	15.80	0.464	5.99	0.76
High	5012751	11.25	0.446	5.12	0.68
TSN_{CO_2/CH_4},	23544	10.98	0.32	6.42	0.72
incorrectly	5020995	10.00	0.49	4.33	0.73
classified	5012751	9.99	0.3	10.33	0.76
	16103	9.47	0.35	5.68	0.74
	13060	11.01	1	4.71	0.62
High	5030899	9.60	0.996	5.11	0.67
TSN_{CO_2/CH_4},	3886	12.23	0.972	6.69	0.79
correctly	5030750	10.91	0.962	4.40	0.65
classified	5022978	6.33	0.94	5.26	0.71
	5022844	7.54	0.916	5.98	0.76
	5025294	3.14	0.746	6.64	0.77
Low	5063900	3.29	0.5	4.85	0.60
TSN_{CO_2/CH_4},	5023812	3.50	0.564	6.66	0.80
incorrectly	2512	3.62	0.68	7.04	0.80
classified	5027410	3.86	0.662	6.96	0.80
	2994	3.70	0.786	6.51	0.75
	5082265	1.90	0	18.57	0.88
Low	5070957	0.83	0.006	3.91	0.48
TSN_{CO_2/CH_4},	5058049	1.28	0.01	10.50	0.87
correctly	3001337	0.39	0.014	24.49	0.95
classified	5074040	1.36	0.036	11.27	0.85
	5058049	1.31	0.036	10.65	0.85

4.3.4 Conclusions

In this chapter, GCMC simulations have been used to obtain computational measures of CO_2 and CH_4 uptake and separation properties in a range of curated MOF structures with a view to identify MOFs promising for biogas upgrading. In addition to uptake and selectivity, the metric TSN_{CO_2/CH_4} is used to measure biogas upgrading performance as it takes into account the trade-off between uptake and selectivity which is an essential consideration when seeking a promising MOF. The results of the GCMC simulations were used to identify selected MOFs which may be promising for biogas upgrading, and also to identify structural features which are common to high-TSN_{CO_2/CH_4} MOFs. It was seen that 2D MOFs with narrow separation between sheets are among those which may be useful for the application.

The results of the GCMC simulations were also used to train a range of machine learning models to predict uptake of the two gases CO_2 and CH_4 in MOFs, as well as to predict TSN_{CO_2/CH_4} and to classify MOFs as high-performing or low-performing based on a TSN_{CO_2/CH_4} threshold. Models were trained on a carefully curated set of real structures, which necessitated a small training set.

Models displayed strong performance in various areas, with random forest models showing the best overall performance throughout. Models were tested on unseen data from a selection of hypothetical MOF structures. Certain aspects of unseen test data were not well-predicted by the models. This appears to be related to the populations of the training and test sets, which do not cover the same areas of feature space. There is therefore scope for future improvement of the trained models by supplementing the training data. The difference between the performance of the models on data from different sources also highlights the diversity of data between different kinds of MOF databases, and the differences between hypothetical and real MOF datasets.

Chapter 5

The Influence of Solvent Molecules on Gas Sorption in Metal Organic Frameworks

Gas storage and separation by adsorption in porous materials have been central to previous chapters as processes of industrial and environmental relevance. Among the many important separations, it has been discussed that porous materials represent a promising avenue for safe and efficient uptake and delivery of next-generation fuels such as CH_4 [245, 252] and H_2, [44, 45] and have for some time been under active consideration for capture and sequestration of greenhouse gases such as CO_2 [46, 47] It has also been discussed that adsorptive separation by membrane methods or by methods involving swing of a selected condition can be a cheaper and less energy-intensive alternative to inefficient distillation-based gas separations. [24] A pertinent example is the separation of CO_2 and CH_4, which is necessary to several processes including biogas upgrading (see Chapter 4). [18, 48] Metal organic frameworks (MOFs) stand out throughout previous chapters and wider literature as candidates which may in principle be readily tuned to gas sorption situations thanks to the modular nature of their assembly permitting a vast quantity of possible MOF structures and compositions.

To fully benefit from the attractive promise of adsorptive separation in porous materials requires a robust understanding of the relationship between the adsorption process and the real structures in which it occurs. Several studies, as has been discussed, exist which examine structure-property relationships in MOFs and related families of porous materials. [11, 88, 104, 180, 187] These include high-throughput computational screenings which obtain predictions of behaviour on a statistical scale and can be invaluable in detecting geometrical and other qualities that can promote selected gas uptake properties as well as in identifying promising individual structures. Computational studies typically make use of MOFs that have been synthesised, had their structures crystallographically resolved, and have been placed in databases. They can also employ hypothetical MOFs that have been generated by combining structural moieties taken from these. [33] However, computational studies almost exclusively make use of pristine, evacuated structures. By contrast, experimental samples of porous materials may include various kinds of defect, including partial collapse, missing portions and residual synthesis solvent trapped in pores. It has been demonstrated that the presence of defects and disorder can have significant impact on gas sorption behaviour. [253] Although the desire to minimise defects and disorder appears natural and routine, their impact on behaviour can be positive or negative depending on the desired use of the sorption process. [253]

Of defects and disorder which may impact sorption properties, the presence of solvent molecules is a pertinent example which has already been referred to in chapter 1. Immediately after synthesis, it is very common for residual synthesis solvent to remain within MOF pores (and consequently to be included in published structural information files). Experimental gas uptake studies are habitually preceded by evacuation processes with the purpose of solvent removal, meaning the effect of solvent is not commonly assessed in depth. However, selected experimental studies have demonstrated significant solvent effects, rendering it necessary to consider the potential of incomplete

or unsuccessful desolvation processes when recommending materials. Erhart et al, [106] for example showed effects of the nature of the desolvation process on geometrical properties. Turning to sorption behaviour, Konik et al [104] observed effects of the nature of the solvent: three different coordinated amide solvents reduced CO_2 and CH_4 uptake in two MOFs, and the magnitude of the reduction among the three solvents depended on properties of the MOFs. Ethiraj et al [107] examined the effect of the degree of desolvation in MOF-76-Ce on CO_2/N_2 separation. They found that a fully desolvated form of the MOF displayed higher total uptake, but a partially desolvated form displayed higher selectivity, bearing out both the common trade-off relationship between selectivity and uptake (see in particular chapter 4), and the observation that the influence of defects can be positive or negative depending on the desired outcome.

When it comes to computational studies, evacuation is commonly reflected by algorithmic stripping of free and coordinating solvent molecules, [11, 20, 21] again meaning that solvent effects are likely to be overlooked. This additionally means that the specific procedure used to eliminate the solvent molecules can impact reported properties. Different solvent removal algorithms use different procedures to characterise parts of structures as solvent for removal. For example, the solvent removal process used in creation of the CSD MOF subset is implemented in a published Python script [21]. A structure is split into components based on binding and only the heaviest part of the structure retained, removing any solvent not bound to the MOF. Solvents coordinated to selected types of metal centre are then additionally removed. This is achieved by identifying metal atoms in the heaviest polymeric part of the structure, computationally removing all bonds to those metal atoms and assessing the resulting separate components. Any component which either corresponds to the common solvents listed by the Cambridge Crystallographic Data Centre, or is a lone oxygen atom, is removed. [21] Meanwhile, the process applied in the creation of the CoRE MOF database [11, 20] also involves a non-bound

solvent removal step in which all components other than the MOF framework (or charge balancing ions) are removed, with components of a structure defined by criteria within the Zeo++ software package. [81] The step is again followed by removal of coordinated solvents. The coordinated solvent removal step is achieved by temporarily removing bonds between metal atoms and oxygen atoms. If the removal resulted in creation of a new separate component, that separate component is removed. Otherwise, the bond is restored. [20]

Where solvent effects have been studied computationally, their potential to impact gas sorption processes has been evident. A study by Haldoupis et al [85] involved developing optimal computational models for CO_2 adsorption in a small group of MOFs through tuning force field parameters and adjusting the degree of solvation in the structures with water and methanol solvents. It concluded that residual solvent molecules were likely to influence gas sorption in MOFs whilst commonly being overlooked. Haldoupis et al [85] found that solvation of up to around 30% of metal centres in the MOFs they studied could not be detected based on surface area. They were also able to analyse the effects of the solvents on uptake events, and found that, while both solvents reduced total CO_2 loading due to free volume effects, the presence of water molecules led to an increase in the heat of adsorption of CO_2 while the presence of methanol led to a decrease.

In later work, Altintas et al [19] compared CH_4 and H_2 uptake at 1 bar in pairs of structures ostensibly representing the same MOF but taken from two different common MOF databases, the CSD MOF subset [21] and the CoRE MOF database. [20] Different algorithms for solvent removal used in these databases account for a substantial proportion of the observed structural differences: of the 387 MOFs whose gas uptake properties differed between the two databases, the differences in 116 were defined by the presence of bound solvent. That is, in one database the MOF structure was given in a form in which some solvent molecule existed which was absent in the other database. Although comparing uptake in solvated and desolvated forms of structures

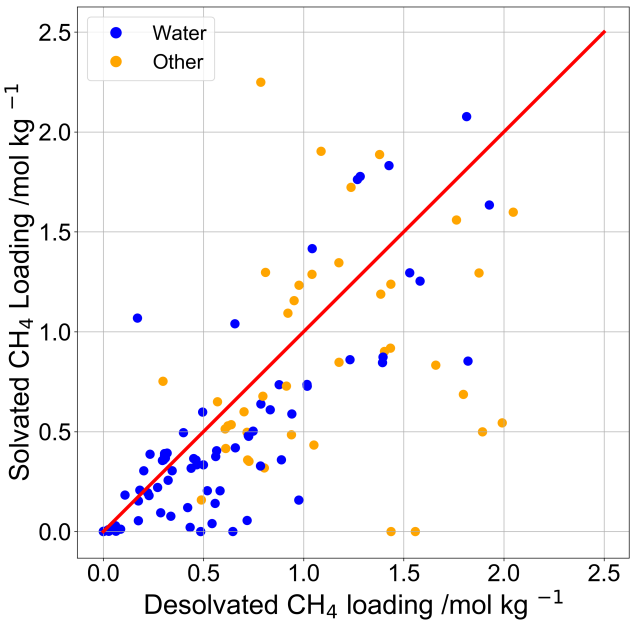

Figure 5.1: *Plot of data taken from Altintas et al [19] of desolvated CH_4 loading against solvated CH_4 loading at 1 bar and 298 K in MOFs taken from the CoRE MOF database [20] and CSD MOF subset. [21] Blue points: solvent is water, orange points: solvent is not water.*

was not the primary purpose of the study of Altintas et al, the comparison is possible using their data [19]. For CH_4 in the data of Altintas et al, situations in which loading is higher in the desolvated form (solvent has a negative effect on loading) are common, and situations in which loading is higher for the solvated form (solvent has a positive effect on loading) are also observed. This is shown in Figure 5.1, in which desolvated CH_4 loading at 1 bar is plotted against solvated CH_4 loading at 1 bar for the MOFs of Altintas et al. [19]

Similarly, while not directly considering solvent effect on uptake, the work of Nazarian et al comparing MOF structures optimised by DFT to those which appear in databases [93] is instructive. The structures were taken from the 2014 CoRE MOF database, [20] where they were deposited following crystallographic structure determination and algorithmic cleaning procedures, but no optimisation. The results of Nazarian et al showed that structures from which solvent had been stripped were more likely to experience a change in sorption behaviour following structural optimisation than those which had not contained solvent to begin with. This finding particularly highlights the

geometrical effects of residual solvent and the need ultimately to verify the geometries of computationally evacuated structures, though gas uptake in solvated structures and therefore a full analysis of solvent effects is outside the scope of the work.

A high-throughput computational work which makes direct consideration of solvent effects is the publication of the 2019 iteration of the CoRE MOF database. [11] Distinction is made in the database between 'bound' solvent, which is directly associated with an otherwise undercoordinated metal site, and 'free' solvent, which interacts only loosely with the framework walls. Two forms of each MOF are published, an 'all solvent removed' form (referred to henceforth here as the desolvated form), which underwent the full desolvation procedure, and a 'free solvent removed form' (referred to here as solvated), in which solvent identified as bound remains and only free solvent was stripped. The utility of a published database containing both forms was demonstrated by high-throughput calculation of Xe and Kr uptake properties in a search for Xe-selective MOFs. It was observed that in some cases solvent blocked favourable Xe binding sites and so reduced uptake and selectivity. In other cases solvent helped to provide conditions conducive to Xe uptake. [11]

Later, Velioglu and Keskin [105] additionally measured solvent effect on processes relevant to CO_2 adsorption in a small selection of MOFs taken from the CoRE MOF database [11] and Cambridge Structural Database (CSD) [254] MOF subset. [21] Solvent presence was found to generally hinder CO_2 uptake processes. The negative effect of solvent on CO_2 uptake is of high relevance to several applications of MOFs, though it has not until now been studied on a high-throughput scale. For the CO_2 uptake process that was the focus of the study of Velioglu and Keskin, [105] positive solvent effects were not observed among the MOFs used. Removal of solvent in computational CO_2 sorption studies as a matter of course was advised. Note, however, the previous studies of both computational and experimental nature [11, 107, 253] which have indicated that negative effects are not exclusively observed for

every process and that positive effects of solvent and other defects can also be seen.

The culmination of these valuable observations of the fact that solvent can have a large impact on gas uptake and of the presence of both positive and negative solvent effects prompts further study. It is desirable to achieve greater understanding of the effect of residual solvent on common adsorption processes in MOFs and related porous materials to enable identification of situations in which elimination of solvent is of particular importance, and of situations in which solvent presence should be promoted. To maximise the usefulness of this understanding, it must ultimately encompass the different behaviour of MOFs and solvents of different kinds, as well as the response of solvent effects to changes in conditions such as pressure.

Here, solvent effects are assessed on a number of levels, each relating to interaction of CO_2 and CH_4 with host frameworks. These are common molecules of vital importance. Aside from the usefulness of the individual gases, they carry the advantage of being widely studied, with computational methods well-validated and not prohibitively slow. They may also to an extent be considered representative of their molecule types: CH_4 is a simple hydrocarbon, commonly modelled as entirely non-polar with no Coulomb interactions, while polarity and partial charge are often relevant to interactions of CO_2, which possesses a quadrupole moment. It may be supposed that the two gases would exhibit distinct behaviour in relation to solvent effects on interaction with host structures, particularly with respect to the open metal sites which desolvation may reveal. First, solvent effects in a small but geometrically diverse selection of MOFs are assessed on a classical level, then this assessment is taken further in high-throughput examination of solvent effects in a dataset of hundreds of structures. Taking a step beyond classical simulations, the effect of solvent on binding energies of small gas molecules to metal centres is then further studied using ab initio methods.

For the first two cases, host structures in both the solvated and desolvated

forms are taken from the 2019 CoRE MOF database. [11] No further costly optimisation of structures in the style of Nazarian et al [93] is attempted, as this work aims to consider only the effect of the presence of solvent on uptake. The same choice was made by Chung et al in the original publication of the database. [11] GCMC calculations are employed to calculate uptake of both gases at 0.1 bar and 1 bar and solvent effect is assessed using the calculated uptakes. Different metrics of solvent effect can be relevant. Perhaps the most obvious are the difference between desolvated and solvated uptake and the ratio of desolvated to solvated uptake, which can communicate different information about the systems in question. The relationships between solvent effect and identity of the guest gas are considered, as well as a number of geometrical properties available as part of the CoRE database. The impact of external pressure is additionally addressed.

In a high-throughput screening of MOFs from a database, consideration of structural viability is necessary. As discussed in chapters 1 and 4, the persistence of unviable structures in published databases is gaining attention as a limitation of studies of this nature. [19, 25, 102, 105, 109, 255] Structural problems in crystallographic information files (cifs) can come about as a result of issues with experimental crystallographic structure determination or with computational cleaning and processing of structures. Symptoms include missing hydrogen atoms, overlapping atoms, and omission of integral parts of structures stripped by overzealous solvent removal algorithms. Oxidation state counting has previously been used as an efficient and effective way to detect problematic structures as those with unviable or highly unlikely oxidation states. Here, a curation workflow detailed in section 5.2.1 and similar to that described in section 4.1 is applied to the database prior to GCMC calculations to minimise the presence of problematic structures.

5.1 Introductory Exploration

For an initial exploration of solvent effect in MOFs of different geometries, gas uptake in a small but geometrically diverse selection of both solvated and desolvated structures was predicted using GCMC simulations. Individual structures from this group were selected for closer inspection and analysis. Focus, as for subsequent sections, is on the selected gases CO_2 and CH_4. Structures for both forms of each MOFs were taken from the 2019 iteration of the CoRE MOF database. [11] In selecting the small subset of the CoRE MOF database to be used for this initial exploration, a representative range of geometries was desired. Therefore, the database was split into bins and a MOF structure was selected from each populated bin as described in section 5.1.1. The bins were defined using structural descriptors that were deemed likely to have an influence on solvent effect.

5.1.1 Structure Selection

In order to select initial structures from geometrical categories expected to span a range of solvent effects, some guidance was needed regarding the geometrical properties most likely to have an impact on these effects. Solvent effects are not routinely examined in studies of gas uptake, and a close analysis of geometry and uptake in solvated and desovlated forms does not yet exist in the literature. Nevertheless, a literature search [11, 19, 93, 106] identified features which may be notable. These are the nature of the solvent, the presence or absence of an open metal site (OMS), and the MOF geometry, most notably PLD, LCD, void fraction, and pore morphology, defined as the ratio LCD/PLD.

The CoRE MOF 2019 database contains readily accessible data on a number of structural features, and those features which are both available and likely to be relevant were prioritised for exploration. The database was partitioned into bins according to relevant features as follows. First, it was refined to include only those materials whose desolvated structures differ from their sol-

vated structures. From the database, 5,928 MOFs were initially taken in each form (11,856 total structures). These were algorithmically checked for similarity and 3,912 were found to contain exactly the same atoms in the desolvated and solvated forms, and were excluded. This left a total of 2,016 members of the database with different solvated and desolvated form. This dataset was then split into those which contain an OMS in the desolvated form and those which do not. This was an unequal split of 1,911 MOF possessing an OMS and 95 MOFs with no OMS. Within the bins, the database was further split based on PLD. Five different PLD size ranges were used, 0-4 Å, 4-6.5 Å, 6.5-8 Å, 8-13 Å and 13+ Å. This led to bins of varying populations, with some very sparsely populated bins and one not populated at all. Bins were further split according to pore morphology. Each bin was split into those structures for which LCD/PLD is less than 2, and those for which LCD/PLD is greater than 2. The ultimate result of this partitioning is a dataset split into 20 bins, each of which defines a certain category of structure. Of the 20 bins, 15 are populated. One MOF was selected at random from each populated bin. Covering both desolvated and solvated forms, there are 30 structures to consider. The bins were not split further to incorporate void fraction because splitting each bin again would have led to an unmanageable number of structures. Table 5.1 displays the population of each bin and basic details about the MOF selected from it (identifying code, PLD in the desolvated form, and the metal(s) it contains).

It is possible a further difficulty is presented by the range of different open metal sites in the selected MOFs. Open metal sites are notoriously difficult to accurately model with standard force field parameters, [84] and can require development of parameters specific to interactions of the metal with relevant guest groups. Such an undertaking is cumbersome, particularly with eight different open metal sites to consider. A possible option which was considered to address this was the restriction of selection of MOFs to only one type of metal site. To explore this possibility, an initial reduced dataset was created

Table 5.1: The bins into which the CoRE MOF database was partitioned and details of each bin's selected MOF. Details are: Identifying code, PLD in Å (desolvated form) and metal centre. The structure files of the MOF HAKSAQ suggest that it contains the metal Zn in the solvated but not the desolvated form

	LCD/PLD< 2		LCD/PLD> 2	
OMS?	Yes	No	Yes	No
0-4 Å	706 MOFs GACPUX PLD: 3.77 Metal: Zn	33 MOFs NIKDEQ PLD: 2.64 Metal: Gd	90 MOFs ILUFIF PLD: 3.09 Metal: Eu, Co	5 MOFs HAKSAQ PLD: 3.94 Metal: In (Zn solv.)
4-6.5 Å	698 MOFs PEKCUF PLD: 5.43 Metal: Mn	33 MOFs ODODOC PLD: 4.84 Metal: Cd, Si	61 MOFs FORPOS PLD: 4.25 Metal: Ce	5 MOFs RIFSOQ PLD: 4.84 Metal: Co
6.5-8 Å	149 MOFs ja302340b_si_002 PLD: 7.66 Metal: Cd, Ni	10 MOFs EBEXAM PLD: 6.79 Metal: Gd	29 MOFs IWEVAJ PLD: 6.68 Metal: Cu	0 MOFs
8-13 Å	132 MOFs OKABAE PLD: 9.53 Metal: Cu	9 MOFs ja406844r_si_002 PLD: 8.44 Metal: Zr	5 MOFs DITJIB PLD: 9.01 Metal: Cd	0 MOFs
13+ Å	41 MOFs LOXKAM PLD: 13.78 Metal: Cd	0 MOFs	0 MOFs	0 MOFs

in which the only metal present was Cu. Cu was considered at this stage as it is a commonly-occurring metal centre in useful MOFs and one of the most frequently occurring in the CoRE dataset (see section 5.2.1). It has been widely studied and readily available adjusted parameters for relevant interactions exist. [85] It was proposed that if only Cu sites were considered, adjusted parameters may be applied to selected MOFs efficiently. However, ultimately this was not deemed possible: the reduced dataset contained only 200 MOFs in total prior to any curation, of which only one possessed no open metal site in the desolvated form. The size of the dataset was considered excessively reduced, and the full dataset was used instead. Due to the diversity of possible kinds of binding present, standard force field parameters are used to represent all interactions.

5.1.2 Grand Canonical Monte Carlo Setup

For introductory exploration of uptake of both gases, single-component GCMC calculations were used for each of the 15 desolvated and 15 solvated MOFs at 0.1 bar and at 1 bar and at 298 K. The setup resembles that used in previous chapters for similar simulations (see Chapters 3 and 4). For framework atoms, Lennard-Jones force field parameters were taken from the UFF [4] and charges were determined using the extended charge equilibration (eQEQ)[162] method available in the RASPA software package. [82] Guest atoms were represented using the TraPPE forcefield, [152, 153] whose three-site CO_2 model includes partial charges while its single-site CH_4 model includes none. Interactions between unlike atoms were calculated using Lorentz-Berthelot mixing rules. As usual, a rigid framework was assumed with host-guest and guest-guest interactions considered within a van der Waals cutoff of 12.8 Å. Bound solvent molecules were treated as part of the rigid framework. Although solvent molecules may be more labile than non-solvent MOF atoms, their movement is severely restricted, justifying their representation as rigid. [85] Similar setups

have been used in the previous studies which have made explicit consideration of solvent effects. [11, 105] Long-range charge interactions were computed using Ewald summation with a cutoff of 10^{-6}.

All Monte Carlo simulations employed at this stage of the study were single component, and as such the available moves were translation, rotation, insertion, deletion, and reinsertion. The process was run for 12,500 equilibration cycles followed by 12,500 production cycles. As before, a Monte Carlo cycle is here defined as a number of trial moves, that number being the minimum of 20 and the number of guest molecules in the system.

5.1.3 Gas Uptake in the Selected Desolvaed and Solvated Structures

Gas uptake in the desolvated form is plotted against uptake in the solvated form for CO_2 and CH_4 in each of the 15 selected structures in Figure 5.2, along with a line $y = x$ in each case for guidance and points coloured, in separate plots, according to the PLD and LCD of the desolvated form of each MOF. Points which lie on or close to the $y = x$ line represent cases where solvated loading is close to desolvated loading: the solvent effect is small. Points lying below the line have a larger desolvated than solvated loading: bound solvent inhibits adsorption, and solvent can be described as having a negative effect. For points above the line, the solvated form adsorbs more readily than the desolvated form, and bound solvent is beneficial to adsorption: it has a positive effect.

A number of points lie on or close to the line $y = x$ for both gases, although a clear effect of solvent is also observed in some cases. For the CH_4 plot at 0.1 bar, several points fall above the line, but at 1 bar these all shift to fall either close to it or or below it. This perhaps indicates a favourable energetic interaction with solvent which becomes less relevant as pressure increases. At increased pressure, available volume, which is always larger in the desolvated

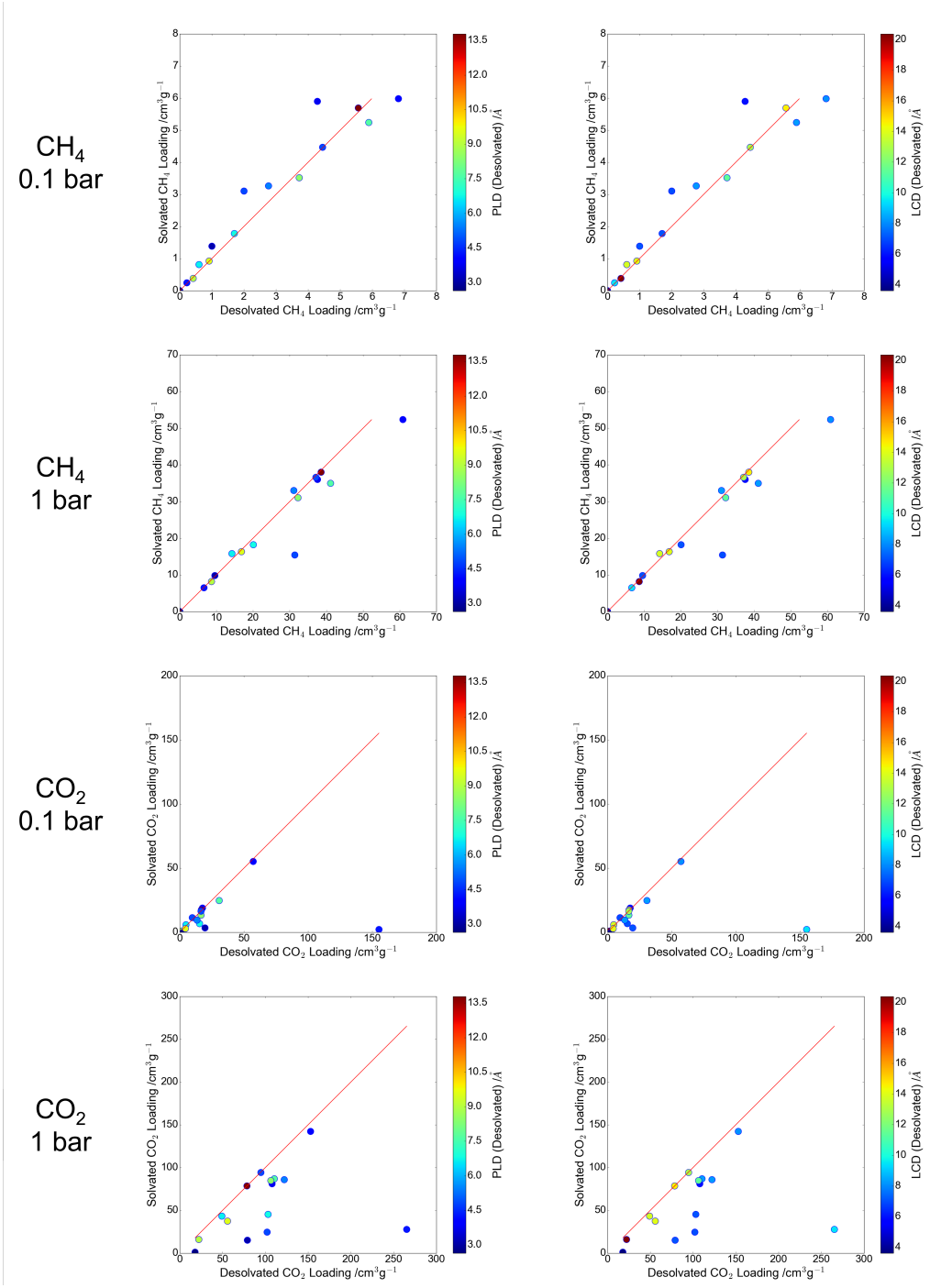

Figure 5.2: *Plots of solvated loading against desolvated loading for the 15 selected MOFs all at 298 K, with points coloured according to PLD (left) and LCD (right). Top row: CH_4, 0.1 bar; second row: CH_4, 1 bar; third row: CO_2, 0.1 bar; bottom row: CO_2, 1 bar. Red lines: $y = x$.*

form, has a more dominant effect on uptake.

There is not evidence of the same observation of positive solvent effect on adsorption at 0.1 bar and negative effect at 1 bar for CO_2: all of the selected MOFs are either not affected or affected negatively by solvent even at 0.1 bar. Negative effects are, however, larger at 1 bar than at 0.1 bar. The lack of any positive solvent effect on CO_2 adsorption at low pressure may be directly because of the nature of the energetic interactions of CO_2 with the host frameworks, although it may also be down to the higher overall loading of CO_2 than CH_4. At 0.1 bar, total CO_2 uptake may already be high enough that volume effects dominate.

The most notable outliers in each plot (that is, the points that fall farthest from the line $y = x$ in either direction), are those for which solvent effect, positive or negative, is greatest. Those in the CH_4 plots are the MOFs with refcodes HAKSAQ, GACPUX, and ODODOC. GACPUX and ODODOC lie above the line $y = x$ at 0.1 bar, while HAKSAQ lies below the line. HAKSAQ remains in a similar position at 1 bar, while GACPUX moves close to the line and ODODOC crosses the line at the higher pressure. The most notable outlier in the CO_2 plots is the point in the lower right, which denotes the MOF with refcode FORPOS. For each outlier, the PLD and LCD in the desolvated form and relevant gas uptakes are given in Table 5.2. As discussed previously, LCD is likely to relate more closely to the space within the pores, while PLD relates to the strongest interactions experienced by guests.

The difference in solvent-influenced behaviour at different pressure points is next assessed in focus. The ratio *solvated loading/desolvated loading* is calculated for each gas at 0.1 bar and at 1 bar. Ratios at 0.1 bar are plotted against those at 1 bar in Figure 5.3, along with the lines $y = x$, $x = 1$, and $y = 1$ in each case, and with points coloured according to PLD and LCD. If the ratio *solvated loading/desolvated loading* equals 1, there is no solvent effect on loading. If a point on the plot lies close to the line $y = x$, the solvent effect on the corresponding structure is similar at 0.1 bar to at 1 bar. Points

Table 5.2: Details regarding the MOFs from the small selection with the most notable solvent effects: PLD and LCD in the desolvated form and uptake of the gas which experiences the strong solvent effect. Desolv. = desolvated, solv. = sovlated.

MOF Refcode	Gas	Desolv. PLD /Å	Desolv. LCD /Å	Uptake 0.1 bar /cm^3g^{-1}		Uptake 1 bar /cm^3g^{-1}	
				Desolv.	Solv.	Desolv.	Solv.
HAKSAQ	CH_4	3.94	8.07	6.82	5.98	60.9	52.3
GACPUX	CH_4	3.77	5.95	4.29	5.90	37.6	36.0
ODOCOC	CH_4	4.84	7.00	2.00	3.10	31.4	15.4
FORPOS	CO_2	4.25	9.45	155	2.18	265	27.2

far from the line display a marked difference in solvent behaviour at different pressures. When considering the effect of pressure, the most significant outlier for both gases is ODODOC (desolvated PLD: 4.84 Å, desolvated LCD: 7.00 Å, CH_4 0.1 bar ratio: 1.55, CH_4 1 bar ratio: 0.491, CO_2 0.1 bar ratio: 1.07, CO_2 ratio 0.751). CO_2 trends appear to depend less on pressure than CH_4 trends.

Based on the plots in Figure 5.2 and 5.3, the MOFs from this selection with the most marked influence of solvent on behaviour appear to be HAKSAQ, ODODOC, GACPUX and FORPOS. Interestingly, the PLD of all four of these outliers in the desolvated form is lower than 5 Å. This highlights small PLD as a geometrical feature which appears to be linked to heightened solvent effect. LCD, meanwhile, is larger than PLD (LCD must be larger than or equal to PLD), but is still fairly small, indicating also a link between LCD and solvent effects.

5.1.4 Detailed Consideration of Individual Structures

Selected structures from among the 15 MOFs are examined in detail, along with favourable locations for gas binding within these structures. Since this would be impractical for all 15 MOFs (30 structures), two MOFs (four structures) are selected, one small-pored example which experiences large solvent effect (GACPUX) and one large-pored example which experiences small solvent effect (LOXKAM). Visualisations of the two MOFs are given in Figure

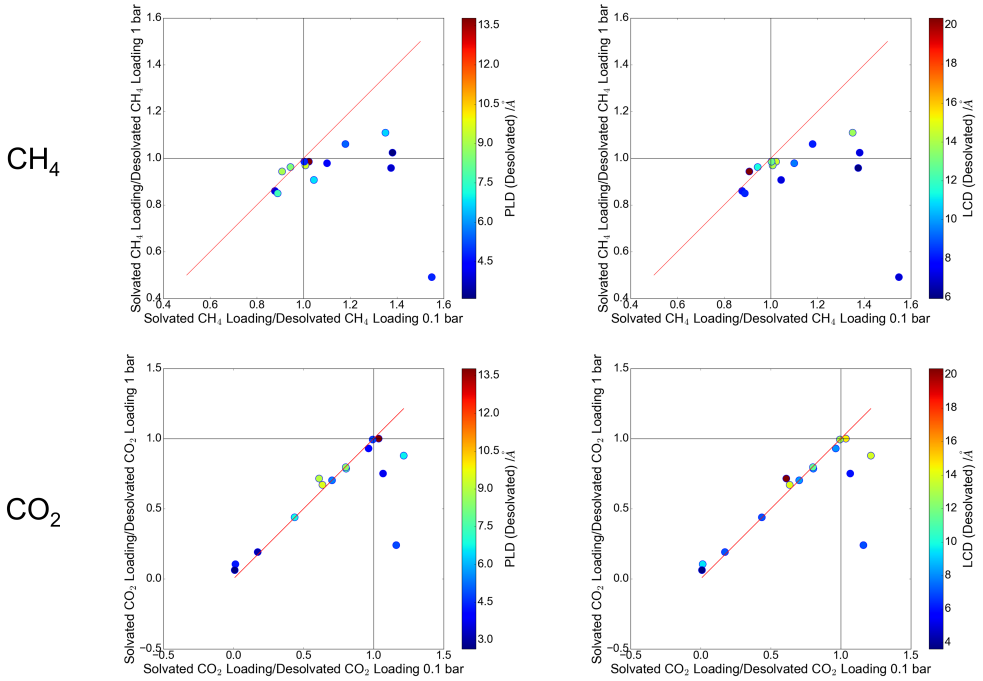

Figure 5.3: *Plots of solvated loading/desolvated loading at 0.1 bar against solvated loading/desolvated loading at 1 bar for CH_4 (top row) and CO_2 (bottom row) for the 15 selected MOFs, all at 298 K. Left: points are coloured according to PLD; right: points are coloured according to LCD.*

5.4 and 5.5. GACPUX is a Zn-based MOF with narrow channels whose solvent is dimethylformamide (DMF). LOXKAM is a Cd-based MOF whose solvent is water. It possesses two types of channel, one large and one medium-sized. The water solvent is located in the corners of the large pores, but the size of the pores means that the solvent is only visible in the image on close inspection.

Locations of adsorbed guests during a simulation are analysed for the solvated and desolvated forms of GACPUX and LOXKAM. Probability density plots of guest location over the course of the GCMC production run were obtained using kernel density estimation and are given for CH_4 and CO_2 at 1 bar in Figure 5.6 for GACPUX and Figure 5.7 for LOXKAM. Similar to the Xe and Kr uptake visualisation in Figure 3.14 of Chapter 3, these figures do not represent a single configuration of molecules that may be observed in a particular snapshot of the simulation, but instead represent the probability density of finding guests in given positions over the whole course of a simulation. The visualisations may be used to aid speculation about the origins of the various

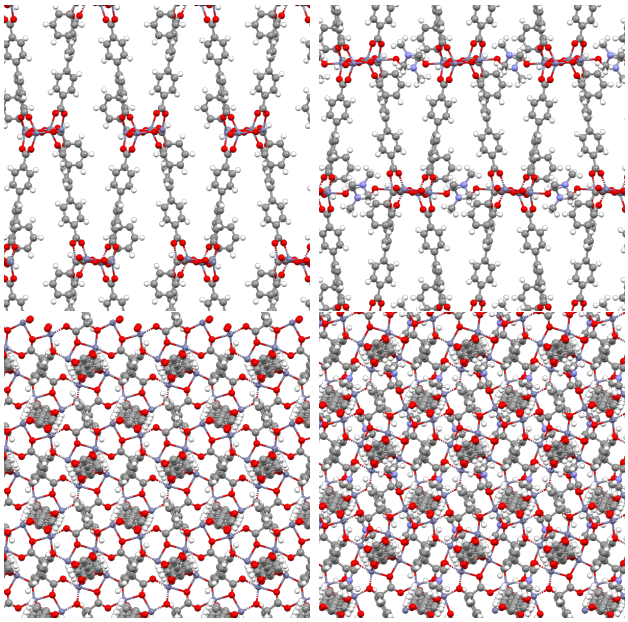

Figure 5.4: *Visualisations of the MOF GACPUX in the solvated and desolvated forms. Top left: the desolvated form, along the a-axis; top right: the solvated form, along the a-axis; bottom left: the desolvated form, along the c-axis; bottom right: the solvated form, along the c-axis.*

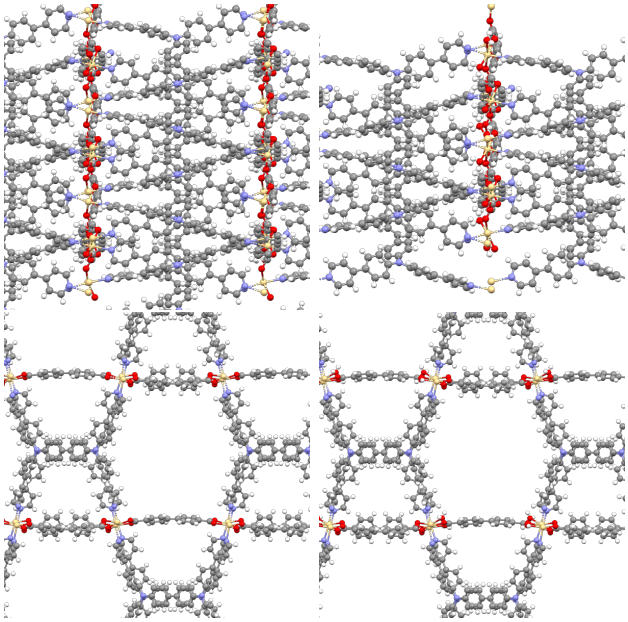

Figure 5.5: *Visualisations of the MOF LOXKAM in the solvated and desolvated forms. Top left: the desolvated form, along the a-axis; top right: the solvated form, along the a-axis; bottom left: the desolvated form, along the c-axis; bottom right: the solvated form, along the c-axis.*

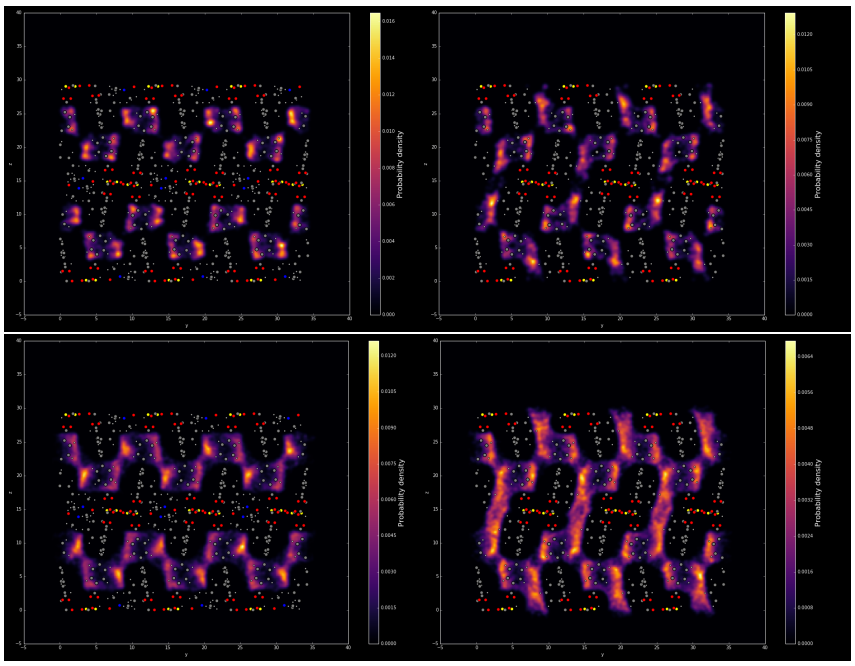

Figure 5.6: *Probability density plots of guest locations over the course of the simulations at 1 bar for GACPUX. Top left: solvated/CH_4; top right: desolvated/CH4; bottom left: solvated/CO_2, bottom right: desolvated/CO_2.*

solvent effects observed, though they give information only about location of adsorption at certain sites, and not directly about the binding energy experienced there. Additionally, since they are only in two dimensions the precise location of adsorption in the third dimension cannot be obtained from them. Suggestions made about the origins of solvent effect using these figures would benefit from confirmation by other methods.

For CH_4 loading in GACPUX, there does not appear to be significantly higher loading in the vicinity of the metals than elsewhere in the pores. Guests appear to cluster around aromatic rings of ligands even in the desolvated form. It is possible that CH_4 has a higher affinity for the aromatic ligands than for the Zn centre, or that the restricted geometry means that desolvation is not sufficient to make the metal accessible to CH_4. If indeed aromatic rings are more important than metal centres to CH_4 adsorption in GACPUX, the lack of improvement in uptake with desolvation at 0.1 bar is explained: the bound solvent does not restrict the most important binding sites. It does, however, affect available volume, which becomes important at higher pressure.

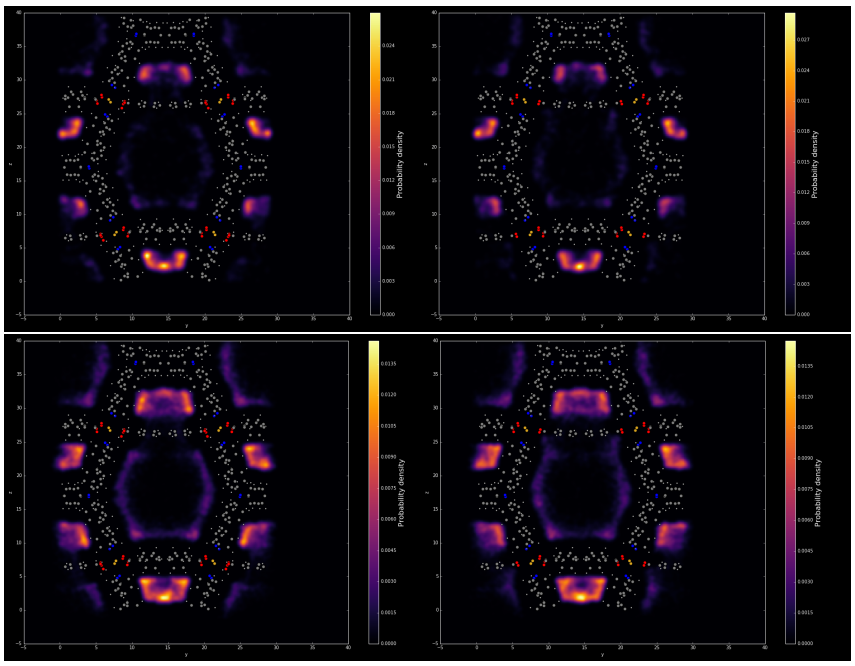

Figure 5.7: *Probability density plots of guest locations over the course of the simulations at 1 bar for LOXKAM. Top left: solvated/CH_4; top right: desolvated/CH4; bottom left: solvated/CO_2, bottom right: desolvated/CO_2.*

The solvated form does not simply exhibit equal loading to the desolvated form at 0.1 bar: it performs better. The reduction in pore size engendered by solvent molecules may influence this, by allowing a closer fit of the guests within the pore, leading to a favourable situation in which interactions are experienced on multiple sides at once.

The difference between uptake locations in the solvated and desolvated forms of GACPUX is more marked for CO_2 than for CH_4. Uptake of CO_2 appears to occur closer to the metal centres. There are two potential reasons that this might be the case: increased affinity with the metal centre due to polarity and size effects. For the first, the polar bonds of CO_2 allow it to exhibit stronger van der Waals interactions and partial charges which may interact with the partially charged environment of a metal centre. For the second, CO_2 possesses a smaller kinetic diameter than CH_4, 3.30 Å compared to 3.80 Å. This difference may be sufficiently large for CO_2 to adsorb significantly closer to the Zn centre and therefore in a more favourable part of the Zn-guest potential well than CH_4.

Turning to adsorption of both gases in LOXKAM, there is very little change in loading locations with dseolvation. Instead, in all considered cases the bulk of loading occurs in the smaller pores where geometrical interactions are likely to play a larger role. In these pores, access to the metal is more restricted. It is no surprise that uptake dominated by adsorption in this location is not strongly affected by solvent. The direct interaction with the metal as described by the UFF and the partial charges assigned by eQEQ is clearly insufficient to have a significant effect on loading. Further consideration of the particular interactions between the two guests and the Zn centre may be instructive to determine whether the UFF provides a sufficient representation and the metal-guest interactions are not strong enough in the large pores to engender a significant solvent effect.

5.1.5 Relationship Between Pore Size and Solvent Effect

The apparent much higher prevalence of solvent effects in small-pored MOFs merits further consideration. A relationship is therefore sought among the 15 MOFs between solvent effect and two geometrical features describing pore size, PLD and LCD. It has been indicated by the PLD and LCD values of the outlying MOFs that these two properties may be related to solvent effect. Previous discussion in this work has addressed how, although correlated with one another, the two relate to different aspects of gas storage and separation, PLD to transport and the close fitting of guest molecules in small pores, and LCD to available space and the fit of guests in large pores.

Solvent effect for this purpose is described by loading difference, defined as the difference between loading in the desolvated form and loading in the solvated form. Desolvated PLD and desolvated LCD are plotted against loading difference in Figure 5.8 along with a line $y = 0$ at which there is zero difference between solvated and desovlated loading for the 15 selected MOFs. In each

case it appears there is a peak or trough at some low PLD or LCD where points are distributed above, both sides of, or below the zero difference line, followed by a tapering off as PLD or LCD increases. The PLD plots are similar to the LCD plots, although for the 15 MOFs studied the LCD maxima and minima appear slightly sharper than those for PLD. This behaviour among only 15 MOFs suggests an interesting relationship between solvent effect and pore size measures, which may be confirmed or refuted by expansion of the study to a larger number of structures.

Indeed, throughout this initial exploration, several interesting solvent effects have been observed which may be better appreciated by expansion of the number of MOFs considered. Study of solvent effects on a high-throughput level is therefore conducted in Section 5.2. This includes measuring of solvent effect metrics at two pressures for hundreds of MOFs, and comparison of solvent effect not only to PLD and LCD but also to other potentially relevant structural properties on this large scale.

5.2 High Throughput Calculation of Solvent Effects

The results presented in Section 5.1 indicate the possibility of interesting statistical relationships relating to the nature and extent of solvent effect on gas uptake processes in MOFs. With this in mind, a high throughput screening was undertaken seeking further understanding of solvent effect on gas uptake in MOFs, and more specifically of solvent effect on CO_2 and CH_4 uptake and separation behaviour.

Structures for the screening, as in Section 5.1, were taken from the CoRE MOF database [11] in both the desolvated and the solvated forms. A curation procedure similar to that employed in Chapter 4 was necessary, and is outlined in section 5.2.1. Uptake of CO_2 and CH_4 was calculated using GCMC

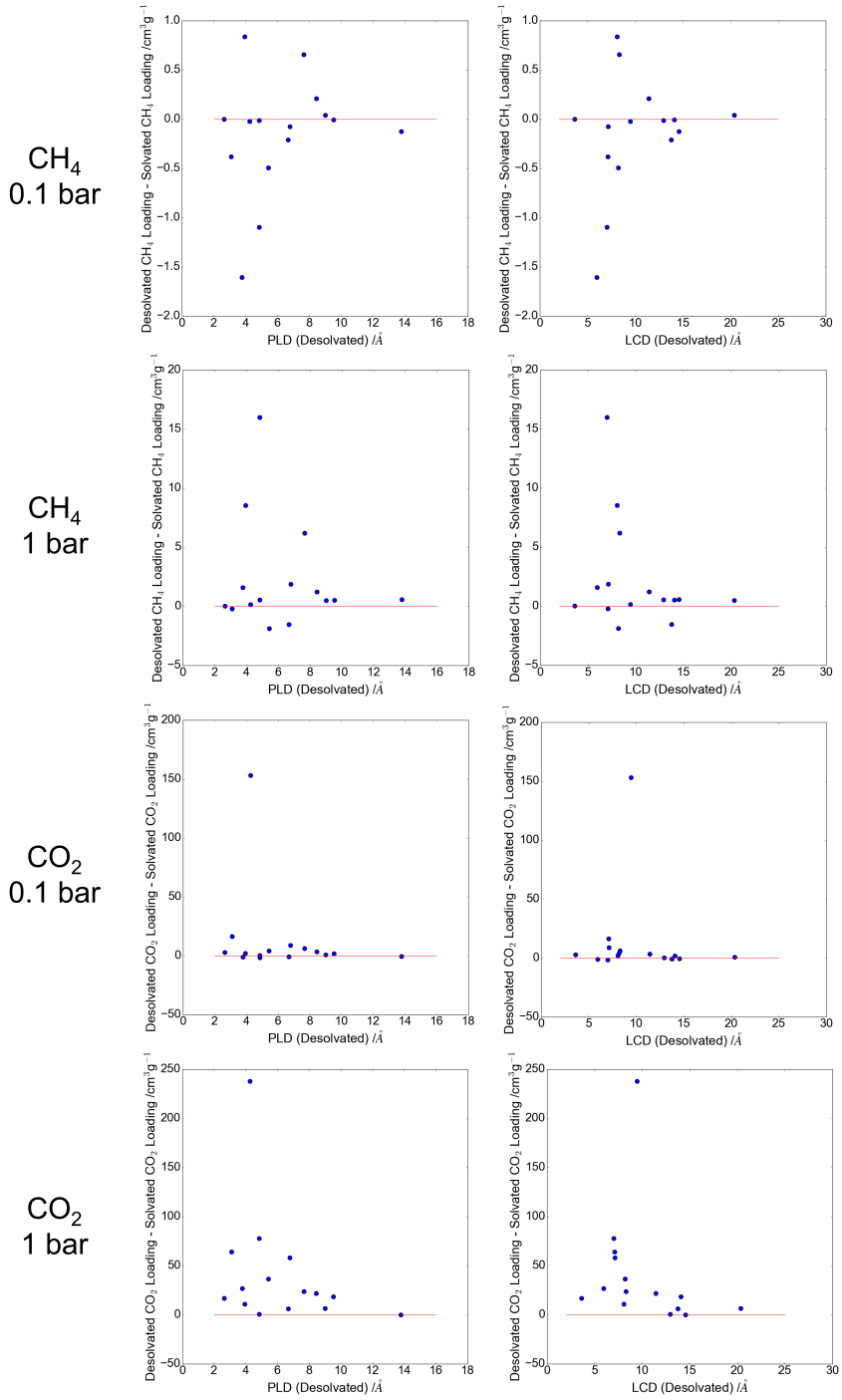

Figure 5.8: *Plots of Desolvated PLD (left) and LCD (right) against loading difference for each gas/MOF system at 298 K. Top row: CH_4, 0.1 bar; second row: CH_4, 1 bar; third row: CO_2, 0.1 bar; bottom row: CO_2, 1 bar. Horizontal lines are Desolvated loading − solvated loading = 0, where desolvated loading equals solvated loading.*

simulations for each structure at 298 K and at two pressure points, 0.1 bar and 1 bar. This was done for single component CO_2 and CH_4 adsorption and for adsorption of the two gases in a 50/50 binary mixture. The GCMC setup is detailed in Section 5.2.1.

5.2.1 Computational Setup: Dataset Curation and Monte Carlo Calculations

The data curation procedure is defined by the following steps. As before, structures for which the desolvated and the solvated form are identical were not required, so the 2,016 MOFs whose structures differ were taken as a starting point. The curation stages that followed are designed to minimise the presence of unviable structures in the dataset, and to remove structures with geometries that are not relevant to the task at hand. Any structure for which either the desolvated or the solvated form contained no metal, no carbon, or only one or two atoms was first removed, leaving a total of 1,930 MOFs in each form. MOFs with overlapping atoms were next treated, removing any structure whose solvated or desolvated form contained any atom-atom distances less than 0.5 Å, a lenient threshold which removed only one MOF, leaving 1,929 structures in each form. To address MOFs with missing H atoms, it was first confirmed that no structure file in the database referred to D instead of H, and then any H-free MOFs were removed, leaving 1,769 structures in each form.

It is not particularly instructive to consider MOFs for a gas uptake application whose pores are either nonexistent or too small to admit the gas molecules in question, nor is it instructive to study the nature of guest interactions inside such pores. Any MOFs whose total area or volume is zero were excluded. MOFs were additionally excluded whose LCD is lower than 3.8 Å, the kinetic diameter of CH_4. Considering pore accessibility, a PLD threshold was used. A rigid structure with PLD smaller than 3.8 Å would be unable to admit CH_4 molecules, but it was discussed in chapters 1 and 3 that several MOFs have

been shown to exhibit flexibility allowing uptake of molecules slightly larger than their PLD. [31, 76] Although subsequent simulations apply a rigid approximation, the Monte Carlo method is able to take account of adsorption within ostensibly inaccessible pores. A 'soft' PLD threshold is used, allowing structures with PLD larger than 3.4 Å in both the desolvated and the solvated form, of which there are 1,092. Values of all structural descriptors used in this curation process (total area, total volume, LCD and PLD) are those provided by the CoRE MOF database. [11]

The penultimate step in the curation processes sought MOFs whose atomic structure necessitated assignment of unviable oxidation states using standard counting rules. For this purpose, the MOSAEC code [109] was employed as in Chapter 4. It was applied to both forms of all 1,092 remaining structures and any flagged MOFs were removed. Following this, 228 structures remained which had not been flagged as problematic in either form.

A contribution to interactions between framework atoms and CO_2 is modelled using Coulomb's law and partial atomic charges. The extended charge equilibration (eQEQ) [162] method available in the RASPA software package [82] was applied to the two forms of the 228 structures. The calculation was unable to run to completion for one of the structures, necessitating its removal from the dataset. The eQeq method is not infallible, and may produce physically unreasonable charges. A charge tolerance was applied to the 227 remaining structures to identify MOFs containing atoms with charges lower than -2 or higher than 3.5. These upper and lower bounds are similar to those used previously in the literature. [18, 161] Two MOFs were identified with charges outside the bounds in either form, and these were removed from the dataset. This left 225 structures which had passed all curation stages and were taken forward for high-throughput calculations.

Therefore, following the curation procedure 225 structures remained which (i) possess desolvated structures which are different from their solvated form, (ii) were not identified as unviable structures, (iii) possess geometries which

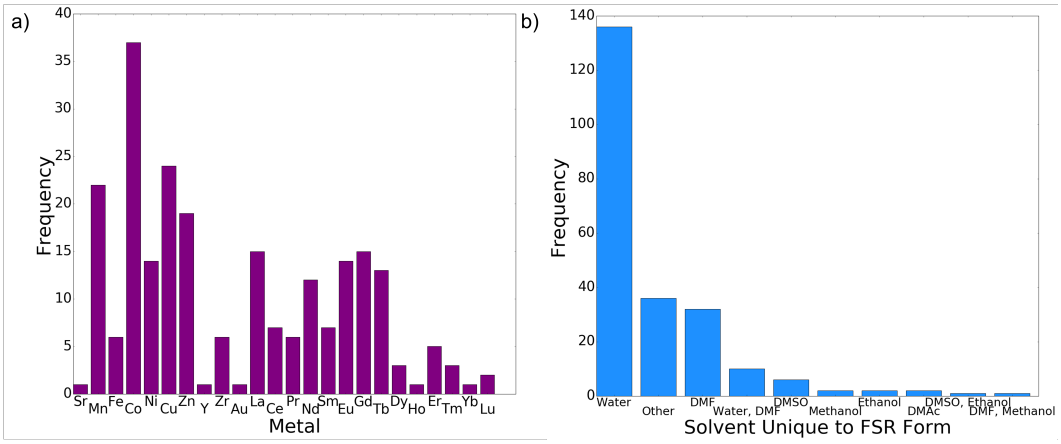

Figure 5.9: *Histograms illustrating the abundance of different metals (a) and solvents (b) in the curated dataset of 225 MOFs.*

indicate the possibility of admitting CO_2 and CH_4 molecules (with some leeway for flexibility) and (iv) were successfully simulated. Among these 225 MOFs, 24 different metal atoms make up the metal centres. In most cases there is only one type of metal in a given MOF, but in a handful of cases there are two. The abundance of each of the 24 metals in the dataset is shown in Figure 5.9a. The most abundant metal is Co with 37 appearances in the curated dataset, followed by Cu with 24 appearances and Mn with 22.

A MOF in the solvated subset differs from its equivalent in the desolvated subset by the presence of one or more solvent molecules. The solvent molecules present in the solvated subset but not the desolvated subset were identified using a modified version of a solvent stripping python script which is available in the literature [21] and which identifies solvent molecules based on comparison of MOF fragments to a list of known solvents. The abundance of each of the solvents in the dataset is shown in Figure 5.9b. By far the most abundant solvent is water. Several solvents were identified in this way, but for some MOFs the solvent which separated the solvated form and the desolvated form was not present in the list utilised by the script. These cases are labelled 'Other' in Figure 5.9b.

For each of the 225 structures, computational predictions of CH_4 and CO_2 uptake under various relevant conditions were calculated using GCMC sim-

ulations with the RASPA software package. [82] Uptake was calculated at 0.1 bar and at 1 bar for pure single component gases and for a binary 50/50 CH_4/CO_2 gas mixture, both at 298 K. The latter case allowed calculation of CO_2 selectivity, S_{CO_2/CH_4} using Equation 3.1, which was defined in Chapter 3.

The GCMC setup was very similar to that used in previous chapters and sections: the rigid approximation was used for host MOFs, with host-guest and guest-guest van der Waals and Coulomb interactions computed. Van der Waals interactions were modelled with a cutoff of 12.8 Å using Lennard-Jones parameters. Parameters were taken from the UFF [4] for the host structures and TraPPE [152, 153] parameters were used for guests. Framework partial charges were calculated using the extended charge equilibration (eQEQ) method available in the RASPA software package [82] using Ewald summation with a cutoff of 10^{-6}. For the Monte Carlo simulations, 15,000 equilibration cycles and 15,000 production cycles were used. The trial moves translation, rotation, insertion and deletion, and, for the binary mixture case, identity change, were available for selection.

5.2.2 Gas Uptake in Solvated and Desolvated Structures

Single component and binary mixture uptake of CH_4 and CO_2, and selectivity for CO_2 over CH_4 in solvated MOFs may be compared to the same data for equivalent desolvated MOFs. The comparison is aided by plotting solvated values against desolvated values. This is done in Figures 5.10 (single component) and 5.11 (binary mixture), in which points are coloured according to LCD in the desolvated form as presented in the CoRE MOF database. [11] This is to facilitate consideration of the relationship between geometrical properties and solvent effect, which is examined in detail for LCD and other geometrical properties in section 5.2.4. A range of solvated and desolvated gas uptakes is observed at both pressures, with generally higher uptake of CO_2 than CH_4.

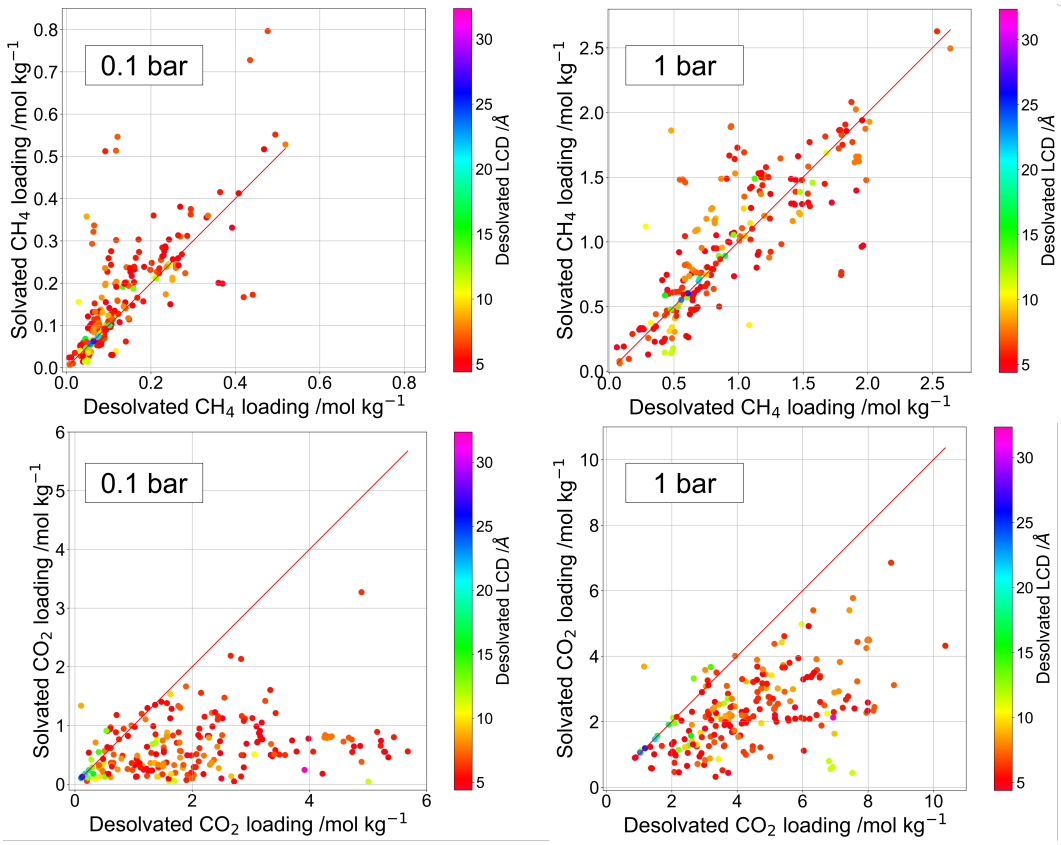

Figure 5.10: *For the 225 MOFs of the curated dataset, single component CH_4 and CO_2 loading in the solvated form against equivalent values in the desolvated form at 0.1 bar and at 1 bar pressure and at 298 K. Points are coloured according to the LCD of the MOF in the desolvated form.*

The line $y = x$ is included in each plot in Figures 5.10 and 5.11 for guidance. Points that lie close to the line $y = x$ correspond to MOFs with similar uptake or selectivity in the solvated and desolvated forms. For points that lie above the line, solvated uptake or selectivity is higher than desolvated uptake or selectivity, meaning solvent has a positive effect on the measured property. For points below the line the reverse is true: solvated uptake or selectivity is lower than the desolvated value, so solvent has a negative effect.

It is useful to define a quantity by which to measure the effect of solvent on the gas sorption properties considered, CH_4 loading, CO_2 loading and, in the case of a binary mixture, CO_2 selectivity. The simple metric of the difference between a property in the solvated form of a MOF and the property in the desolvated form of the MOF is used here to measure solvent effect:

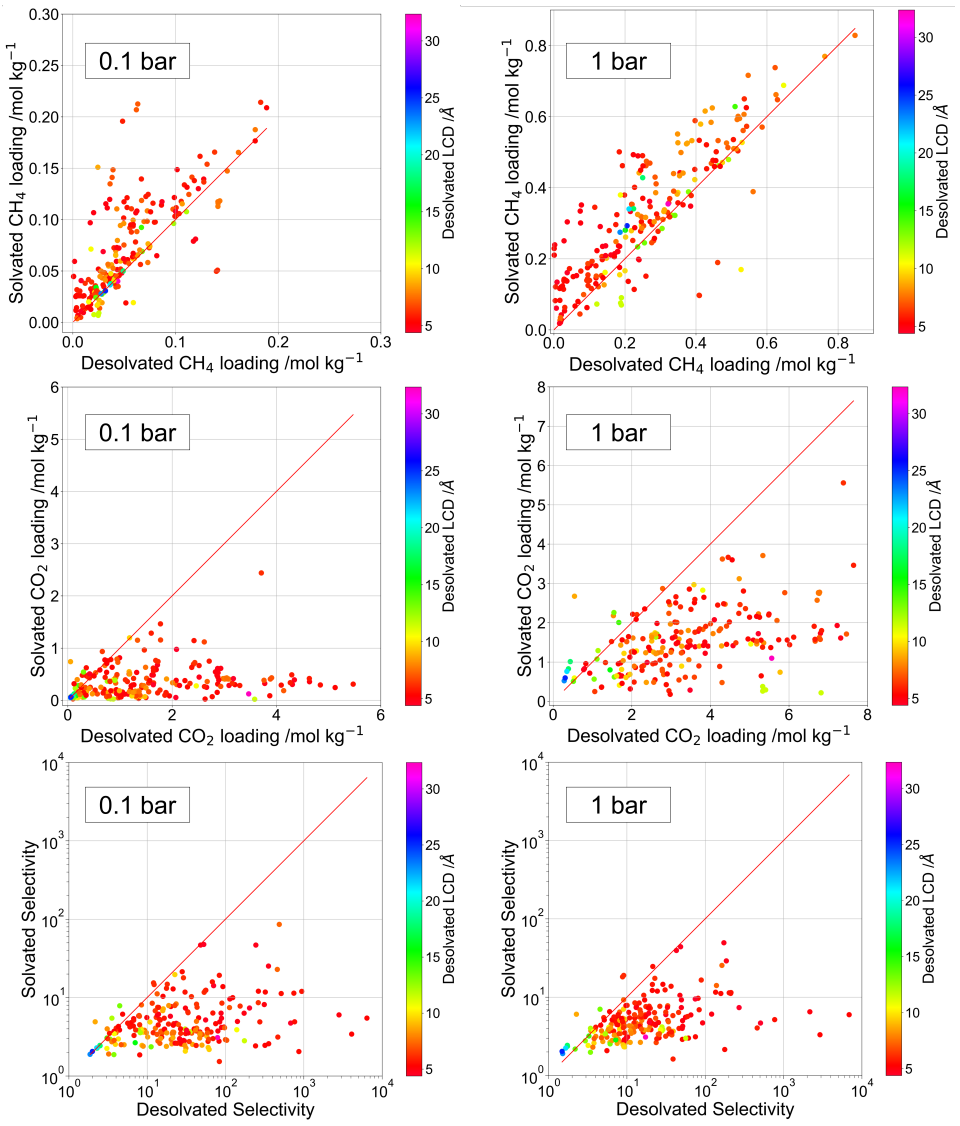

Figure 5.11: *For 225 MOFs of the curated dataset, CH_4 and CO_2 loading in the solvated form against equivalent values in the desolvated form for the 50/50 binary mixture at 0.1 bar and at 1 bar pressure and at 298 K. Points are coloured according to the LCD of the MOF in the desolvated form.*

$$Loading\ difference = Solvated\ loading - Desolvated\ loading \quad (5.1)$$

A positive loading or selectivity difference corresponds to a positive effect of solvent on loading or selectivity (points falling above the line in Figures 5.10 and 5.11), and a negative difference means the reverse. It is useful to identify not only the direction of solvent effect, but also whether the effect is significant: points for which difference is close to zero within a small tolerance can be said to experience little effect of solvent on the measured quantity. Thresholds may be defined to categorise MOFs as experiencing positive, negative or little solvent effect, and the number of MOFs experiencing each for a given metric under a given set of conditions may be quantified. There is a large difference in total magnitude of CH_4 and CO_2 loading. Therefore, different thresholds are used here to define the nature of solvent effect for the two gases: $\pm$ 0.01 mol kg^{-1} for CH_4 and a threshold of $\pm$ 0.1 mol kg^{-1} for CO_2. A threshold of $\pm$ 5 is used for selectivity. The thresholds and the number of structures which fall into each category are summarised in Table 5.3.

For CH_4 in the single component case at both pressures, a substantial proportion of points lie either side of the line $y = x$, meaning solvent may have a positive or a negative effect on adsorption. More points lie above the line than below, and the points farthest from the line are on the positive side: although both positive and negative effects are seen, positive effect is more prevalent and can reach a stronger absolute maximum than negative effect. This is contrary to any perception of solvent presence as always negative and to be minimised, [104, 105] and solvent effect displays a somewhat different distribution to that seen in the results of Altintas et al in figure 5.1, The dominance of positive solvent effect is slightly reduced as pressure increases, at which point the total magnitude of loading increases also. The effect of volume, which is always reduced by solvent presence, thus gains significance. In the binary mixture

Table 5.3: The thresholds used to define positive, negative and little solvent effect for each quantity measured, and the number of MOFs which fall into each category. SC = single component, BM = binary mixture, diff = difference.

	Negative Effect	Little Effect	Positive Effect
SC CH_4	diff $<$ -0.01	-0.01 $\leq$ diff $<$ 0.01	diff $\geq$ 0.01
0.1 bar	49 structures	59 structures	119 structures
SC CO_2	diff $<$ -1	-1 $\leq$ diff $<$ 1	diff $\geq$ 1
0.1 bar	125 structures	101 structures	1 structure
SC CH_4	diff $<$ -0.01	-0.01 $\leq$ diff $<$ 0.01	diff $\geq$ 0.01
1 bar	95 structures	17 structures	115 structures
SC CO_2	diff $<$ -1	-1 $\leq$ diff $<$ 1	diff $\geq$ 1
1 bar	169 structures	57 structures	1 structure
BM CH_4	diff $<$ -0.01	-0.01 $\leq$ diff $<$ 0.01	diff $\geq$ 0.01
0.1 bar	19 structures	96 structures	112 structures
BM CO_2	diff $<$ -1	-1 $\leq$ diff $<$ 1	diff $\geq$ 1
0.1 bar	104 structures	123 structures	0 structures
S	diff $<$ -5	-5 $\leq$ diff $<$ 5	diff $\geq$ 5
0.1 bar	167 structures	59 structures	1 structure
BM CH_4	diff $<$ -0.01	-0.01 $\leq$ diff $<$ 0.01	diff $\geq$ 0.01
1 bar	34 structures	23 structures	170 structures
BM CO_2	diff $<$ -1	-1 $\leq$ diff $<$ 1	diff $\geq$ 1
1 bar	157 structures	69 structures	1 structure
S	diff $<$ -5	-5 $\leq$ diff $<$ 5	diff $\geq$ 5
1 bar	141 structures	83 structures	3 structures

case, although negative solvent effects on CH_4 loading remain present, the increased prevalence of positive effects is exaggerated compared to the single component case at both pressures. Increased CH_4 uptake with solvent presence is now accompanied by reduced CO_2 uptake and reduction in the competition experienced by CH_4.

The effect of solvent on CO_2 uptake under the conditions studied is clear. The majority of points fall below the line $y = x$ indicating a negative effect of solvent on uptake, and those which fall above the line do not fall far above it. This is in line with a desire to minimise solvent presence for CO_2 uptake applications. [105] The significant increase in negative effects of CO_2 loading compared to CH_4 loading may be down to strong interactions of CO_2 with metal sites which are blocked by bound solvent. This reasoning implies that in only very few cases does CO_2 experience interactions with solvent molecules themselves which are stronger than those with the metal sites that they block. The negative effects can additionally or alternatively be attributed to volume effects: the total magnitude of CO_2 loading is in general significantly higher than that of CH_4 loading. At higher total loading, available volume becomes more important and negative solvent effects become more likely.

Selectivity for CO_2 over CH_4 experiences a similar qualitative solvent effect to CO_2 loading, although the scale of the total values of selectivity is much larger (note that the selectivity plots in Figure 5.11 are on a logarithmic scale). The large scale of selectivity is a result of some MOFs displaying very low CH_4 loading. The presence of solvent molecules is likely to reduce CO_2 uptake and increase CH_4 uptake, concurrently reducing CO_2 selectivity. In no case in either the desolvated or the solvated form of any MOF does selectivity fall below 1, which would indicate a reversal of selectivity, a MOF being CH_4-selective rather than CO_2-selective.

In seeking MOFs for CO_2 uptake or for selectivity of CO_2 over CH_4, the results indicate that it is very rare that solvent presence should be encouraged. However, with the CO_2 uptake and selectivity plots including some points close

to the line $y = x$ and others very far from it comes the additional indication that in some cases minimising solvent presence should be a priority, while in others it can be overlooked without significant effect on performance. When it comes to CH_4 capture, MOF structures exist on a significant scale in which the presence of bound solvent ought to be encouraged in order to enhance uptake, particularly for low pressure applications.

5.2.3 Relationship Between Pressure and Solvent Effect

The observed effect engendered by residual solvent differs depending on external pressure, and this relationship is now examined in greater detail. For each gas, loading difference at 0.1 bar is plotted against loading difference at 1 bar in Figures 5.12 (single component case) and 5.13 (binary mixture case). Figure 5.13 additionally contains a plot of selectivity difference at 0.1 bar against selectivity difference at 1 bar, as calculated from binary mixture uptake data and on a symmetric log scale. These are the difference metrics previously defined and discussed.

For a pressure of 0.1 bar, MOFs which lie close to the line $x = 0$ experience little solvent effect on the quantity in question. Those which lie to the left of it experience a negative effect (presence of solvent reduces the quantity) and those to the right of it in the orange hashed region experience a positive effect (solvent presence increases the quantity). For 1 bar pressure, MOFs close to the line $y = 0$ experience little effect; those below it experience a negative effect and those above it in the purple hashed region experience a positive effect. Meanwhile, for points close to the line $y = x$, the magnitude of solvent effect as measured by difference is similar at the two pressure points; for points above the line $y = x$ in the green hashed region the magnitude of solvent effect is more positive at 1 bar than at 0.1 bar and for those below it the reverse is true. Therefore, the octant or quadrant into which a point falls describes the nature of the solvent effect on that MOF as described by the difference metric,

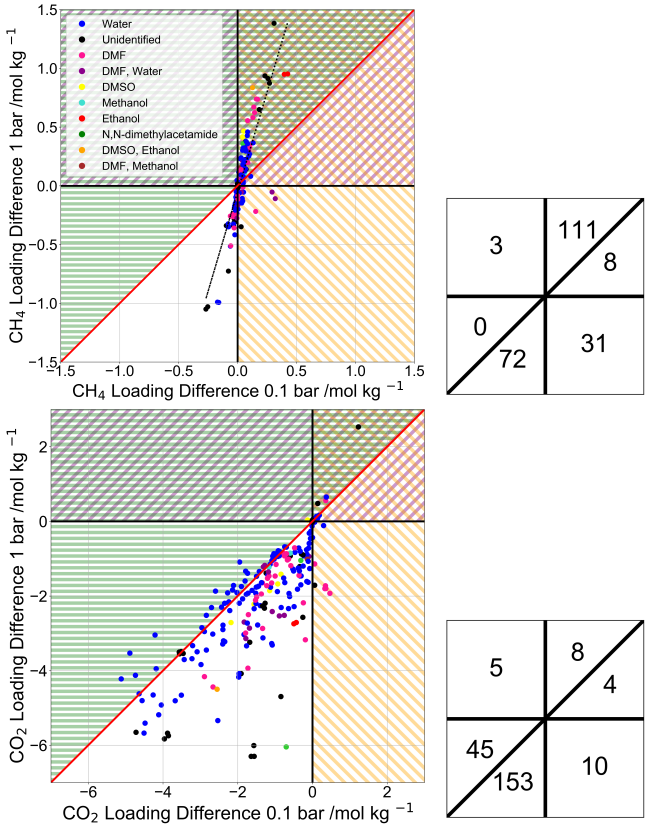

Figure 5.12: *Plots of CH_4 (top) and CO_2 (bottom) loading differnce at 0.1 bar and 298 K against loading difference at 1 bar and 298 K as calculated from single component GCMC simulations. Lines $x = 0$ and $y = 0$ are given in black and line $y = x$ is given in red for guidance. Alongside each plot is a grid showing the population of the octants and quadrants formed by the guiding lines. On the CH_4 plot, a dashed regression line is shown, which has equation $y = 3.39x - 0.051$.*

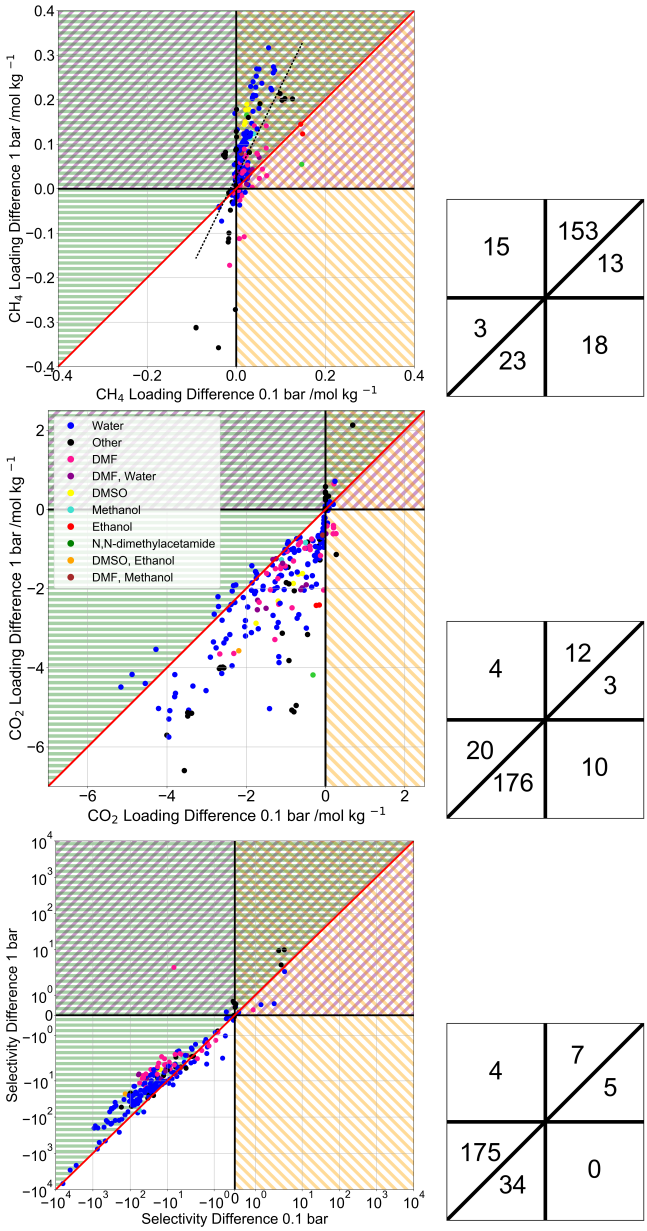

Figure 5.13: *Plots of CH_4 (top) and CO_2 (centre) loading differnce at 0.1 bar and 298 K against loading difference at 1 bar and 298 K, as well as selectivity difference at 0.1 bar and 298 K against selectivity difference at 1 bar and 298 K (bottom), as calculated from binary mixture GCMC simulations. Lines $x = 0$ and $y = 0$ are given in black and line $y = x$ is given in red for guidance. Alongside each plot is a grid showing the population of the octants and quadrants formed by the guiding lines. On the CH_4 plot, a dashed regression line is shown, which has equation $y = 2.03x + 0.027$*

as follows.

MOFs in the upper half of the top right quadrant, (the north-north-east octant as defined by compass points, with orange, purple and green hashes) have a positive solvent effect at both pressures, and this effect is more pronounced at 1 bar than at 0.1 bar. MOFs in the lower half of the top right quadrant (east-north-east, orange and purple hashes) have a positive solvent effect at both pressures, which is more pronounced at 0.1 bar than at 1 bar. MOFs in the bottom right quadrant (south-east, orange hashes only) have a positive solvent effect at 0.1 bar but a negative effect at 1 bar. MOFs in the lower half of the bottom left quadrant (south-south-west, no hashes) experience a negative solvent effect at both pressures, and the effect is more negative at 1 bar than at 0.1 bar. MOFs in the upper half of the bottom left quadrant (west-south-west, green hashes only), also experience negative solvent effects at both pressues, but the effects are more negative at 0.1 bar than at 1 bar. MOFs in the top left quadrant (north-west, green and purple hashes) experience a positive solvent effect at 1 bar and a negative effect at 0.1 bar. Alongside each plot in Figures 5.12 and 5.13 are grids giving the populations of the relevant octants and quadrants in each plot.

The relationship between solvent effect on CH_4 adsorption and pressure is somewhat opposite to that for CO_2 adsorption in both the binary mixture and the single component case. For CO_2 adsorption, the majority of the points fall into the unhashed south-south-west octant, which denotes negative solvent effect at 0.1 bar and more negative solvent effect at 1 bar. Negative solvent effects at 0.1 bar are more likely to be down to strength of framework interactions, while the enhancement of negative effects at 1 bar is more likely to relate to volume availability. For CH_4, on the other hand, the most populated octant in both cases lies opposite: the north-north-east octant dashed in purple, orange and green, in which positive solvent effects as measured by difference are observed at both pressures, and the effects are more positive at the higher pressure. It should be noted that the behaviour of CH_4 is more varied

than that of CO_2, with a more appreciable number of MOFs falling into the opposite south-south-west octant, particularly in the single component plot. However, compared to CO_2, the general behaviour of CH_4 molecules involves a greater likelihood of more positive interactions with solvent molecules than with metal centres. At higher pressure it might be expected that volume effects would countermand any increased loading, but for CH_4 the higher pressure instead facilitates increase in loading difference. In the single component case it is easily seen that the total magnitude of CH_4 loading is not sufficient for volume effects to be particularly relevant. In the binary mixture case, it must be remembered that presence of CO_2 dominates compared to CH_4, so volume effects may be expected. Indeed, the magnitude of loading difference is smaller for the binary mixture case than for the single component, but points remain in the north-north-east octant.

Interestingly, for CH_4 the MOFs fall into a reasonable approximation of a straight line, particularly in the single component case: an approximately linear relationship exists between solvent effect at 0.1 bar and solvent effect at 1 bar. If CH_4 loading difference at one pressure is known, it could be used to predict solvent effect at the other pressure. To this end, regression lines were fitted for both CH_4 loading difference cases. For the binary mixture case, the fitted line is $y = 2.03x + 0.027$ with an R^2 value of 0.44, while for the single component case the fitted line is $y = 3.39x - 0.051$ with a higher R^2 value of 0.71.

A general observation that stems from analysis of all four loading plots in Figures 5.12 and 5.13 is the dominance of the north-east and south-west quadrants and within those the north-north-east and south-south-west octants. Few points fall into either the north-west or south-east quadrants: if solvent has a particular effect on loading at one of the two pressures studied, that effect may be reduced or enhanced at the other pressure, but is unlikely to be reversed. Once positive or negative effect at both pressures is established, an increase in pressure from 0.1 bar to 1 bar is likely to enhance rather than

reduce the effect. That is, if the effect is negative at 0.1 bar it is likely to be more negative at 1 bar, and if the effect is positive at 0.1 bar it is likely to be more positive at 1 bar as measured by the difference metric.

When it comes to selectivity, the majority of MOFs fall into the west-south-west octant of the plot in Figure 5.13. Of those which do not fall into the west-south-west-octant, the majority lie in its neighbouring south-south-west octant and are not far from the line $y = x$ which divides them. The positioning means that in most cases solvent has a negative impact on selectivity at both pressures and the effect is more negative at the lower pressure. Selectivity of CO_2 over CH_4 depends on uptake of the two gases. It relies on the magnitude of uptake, not just on uptake difference. Total selectivity tends to be higher at lower pressures, so any absolute change in selectivity is likely also to be larger, meaning a negative change in selectivity will be more negative. A potentially useful implication is that at high enough pressure there would be very little effect of solvent on selectivity. However, this must be balanced by the fact that at this high pressure selectivity itself would be very low, and solvent effect on CO_2 uptake may be very negative.

Points in Figures 5.12 and 5.13 are coloured according to the nature of the solvent in the MOFs. No clear pattern is observed which separates the effects of one solvent from the effects of another.

It is possible to measure solvent effect using metrics other than loading difference. For example, one could measure the ratio of uptake in one form to uptake in another, as in section 5.1. The effect of pressure on loading ratio is not always the same as on loading difference. The focus here is on difference because of its relevance to useful MOF applications. At very low total uptake a ratio may be very high even where total uptake is not usefully large, while at higher total uptake an increase or decrease in loading that does not amount to a very significant ratio may be relevant to a researcher seeking MOFs for gas sorption applications. Ratio, however, can also be useful in discussions of the nature and origin of solvent effect.

5.2.4 Relationship Between Geometrical Properties and Solvent Effect

It has thus far been established that measures of CO_2 uptake and selectivity in MOFs are likely to be affected by the presence of solvent negatively if at all, while CH_4 uptake may be affected in either direction with a preference for positive effects, especially at low pressures. The effect of solvent, however depends on more than the guest molecule. Understanding of the relationship between solvent effect and MOF properties is also desirable. Such a relationship is indicated in the plots shown in Figures 5.10 and 5.11, in which points are coloured according to LCD in the desolavted form. [11] From this colouring it is clear that the MOFs furthest from the line $y = x$, those experiencing the largest solvent effect on both uptake and selectivity, tend to possess low values of LCD close to 5 Å and in almost all cases certainly below 10 Å. This is particularly true of CO_2 loading, but also seen for CH_4 loading. Meanwhile, MOFs with larger LCD tend to display smaller solvent effects, and also to display smaller total CO_2 loading and selectivity. To further probe relationships between geometric properties and solvent effect, solvent effect difference metrics (defined in equation 5.1) are compared against a selection of geometrical properties available in the CoRE MOF database. [11]

The full list of geometrical properties published in the CoRE MOF database includes largest cavity diameter (LCD), pore limiting diameter (PLD), largest free pore diameter (LFPD), density, accessible surface area (ASA, volumetric and gravimetric), non-accessible surface area (NASA, volumetric and gravimetric), accessible volume (AV, gravimetric and void fraction (VF)), and non-accessible volume (NAV, gravimetric). From these, further geometrical properties can also be obtained. Total area and volume are the sums, respectively of accessible and non-accessible area and volume. Pore morphology, which describes the uniformity of pore diameters, is given by the ratio LCD/PLD. The database also contains information on whether an open metal site (OMS) is

present. All information is contained for both the desolvated and the solvated form of each MOF in the CoRE database. Focus in this section is on information about the desolvated form, as this is the form most likely to be used in a computational study, and its structure is accessible from the solvated form by use of solvent stripping scripts, which may be found in the literature. [11, 21] In most cases, the structure of the solvated form is also available and relationships between solvent effect and the geometry of the solvated form could equally be established.

For the sake of brevity, focus is on only a selection of relevant geometrical properties. Without addressing solvent effect, complex predictive relationships have previously been established between uptake and selectivity of the two gases in question and a combination of pore sizes, surface area, void fraction [88, 180, 228] and more rarely density. [180] Given the importance of these properties to uptake and selectivity themselves, establishment of an understanding of their impact on solvent effects is clearly valuable.

A close analysis of geometry and uptake in solvated and desolvated forms of MOFs does not yet exist in the literature, but early analysis in section 5.1 indicated a particular relationship between degree of solvation, gas uptake, and pore size and shape as defined by PLD, LCD and pore morphology. This has been borne out on a statistical scale for LCD as seen in Figures 5.10 and 5.11, and is also supported by initial indications from the literature. [11, 19, 93, 106] Both LCD and PLD can be of critical importance for gas uptake and selectivity, [89, 187, 256, 257] and it may be readily imagined that the finely tuned interactions engendered by well-selected pore diameters may be altered by solvent presence. This relationship is observed here on a large scale for PLD and pore morphology in addition to LCD. All three metrics are plotted against uptake and selectivity difference in Figures 5.14 (0.1 bar) and 5.15 (1 bar). Only binary mixture plots are given; single component plots display similar behaviour.

A relationship between pore size and solvent effect is apparent, particularly

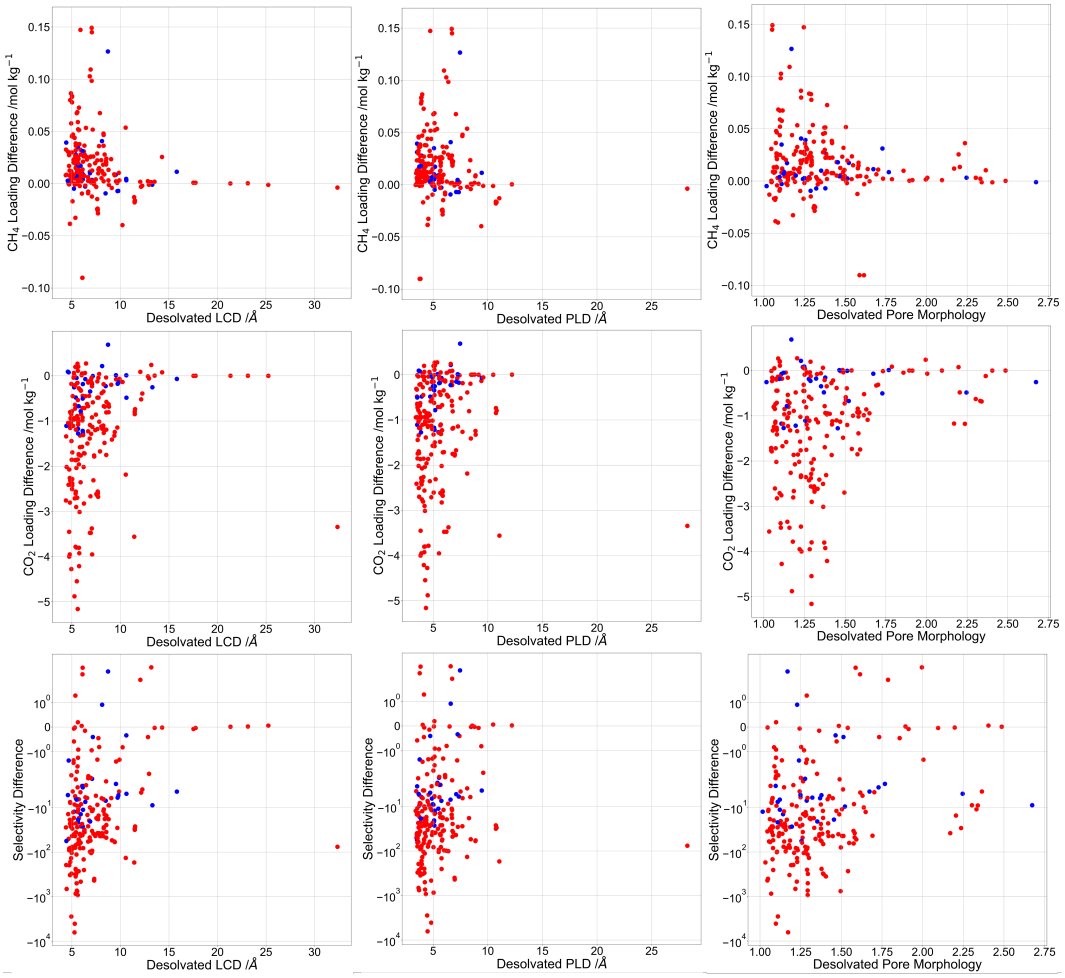

Figure 5.14: *PLD, LCD and pore morphology obtained from the CoRE MOF database [11] plotted against CH_4 uptake, CO_2 uptake and CO_2 selectivity (symmetric log scale) at 0.1 bar and 298 K. Red points: the MOF has no open metal site in the solvated form. Blue points: the MOF has an open metal site in the solvated form.*

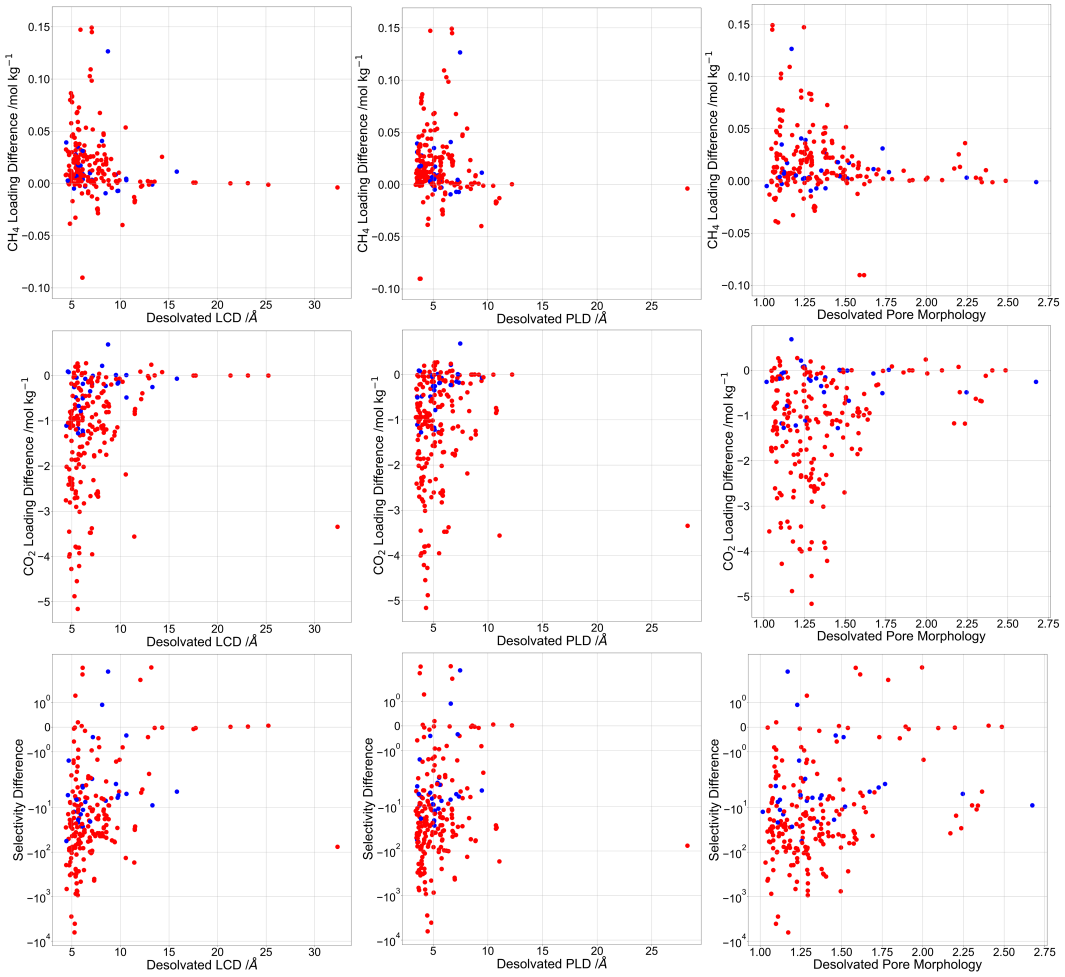

Figure 5.15: *As Figure 5.14, but at 1 bar pressure.*

at 0.1 bar (Figure 5.14). The relationship seen in Figures 5.10 and 5.11 for LCD is also seen in Figures 5.14 and 5.15, and extends also to PLD. While the nature of solvent effect reverses from CH_4 to CO_2 and its magnitude reduces, the qualitative dependence of magnitude on the pore size features is similar for the two gases: small loading difference is possible over the full LCD and PLD ranges, but particularly large effects are for the most part only observed in MOFs with small pores. The very largest solvent effects occur in MOFs with both (desolvated) PLD and LCD around 5 Å, with LCD necessarily slightly larger than PLD; these are the MOFs which should be treated with the most care regarding solvent.

The relationship between pore morphology and solvent effect at 0.1 bar (Figure 5.14) follows a similar pattern to that of LCD and PLD, with the maximum solvent effect on loading occurring for pores with LCD/PLD between 1 and 1.25 for CH_4 and between 1 and 1.5 for CO_2. There is a significantly populated tail of points with LCD/PLD greater than around 1.5 for which solvent effect on loading is not large and it can be surmised that for MOFs of this description significant solvent effect is unlikely.

When it comes to selectivity at 0.1 bar (Figure 5.14), behaviour is in some respects similar to uptake behaviour, although the range of selectivity values, as seen in previous sections, is much larger than the range of uptake values (the selectivity plots are on a symmetric log scale). The relationships of selectivity difference with LCD, PLD and in particular pore morphology are additionally less well-defined than for the uptake differences. For example, there is a cluster of MOFs with LCD/PLD larger than 2 and selectivity difference between 10^1 and 10^2. The largest magnitude of selectivity difference observed approaches 10^4, but a selectivity difference of 10^2 is certainly significant. This cluster of structures makes any general suggestion that solvent effect can be dismissed for selectivity discussions relating to MOFs with LCD/PLD larger than a certain value difficult. The general statement that MOFs with smaller LCD/PLD are more likely to have a larger absolute selectivity difference may nonetheless be

made.

At a higher pressure of 1 bar (Figure 5.15), the general behaviour of solvent effect with changing pore size and shape follows a similar pattern to that at 0.1 bar, although the behaviour is less well-defined and more outliers persist. In the same way as for selectivity at 0.1 bar, the poor definition hinders clear assignment of cutoffs beyond which solvent may as a general rule be ignored, but does not prevent general observations about the likelihood of solvent effect in MOFs of different kinds.

Pore size and shape are not the only geometrical properties which are important to gas uptake in MOFs or relevant to a discussion of the relationship between geometry and solvent effects. In Figures 5.16 and 5.17, ASA (in m^2 g^{-1}), VF and density are plotted against the solvent effect difference metrics. At 0.1 bar, there is some definition in the relationships between these properties and solvent effect, though the relationships are in general less well-defined than the relationships between pore size and shape and solvent effects. At 1 bar, some loose relationships may be seen but they are more poorly defined still.

Across the range of surface areas, very small effect of solvent on all metrics may be observed at 0.1 bar (Figure 5.16). For CH_4 and CO_2 uptake, higher loading difference is more likely to be observed for MOFs with lower ASA, with distributions centred around 1,000 m^2g^{-1}. However, very high loading difference can be observed outside of the general distribution. At 1 bar (Figure 5.17), a loose pattern of a similar nature exists between surface area and solvent effect. At both pressures, selectivity difference is generally large and negative at low surface area and becomes closer to zero with increasing surface area, though with many outliers.

The relationships between VF and the solvent effect difference metrics are the most poorly defined among the geometrical properties considered at both pressures. An intermediate VF range may be seen at which particularly positive or negative CH_4 loading difference and particularly negative CO_2 loading difference are more likely to occur. However, several outliers persist for both

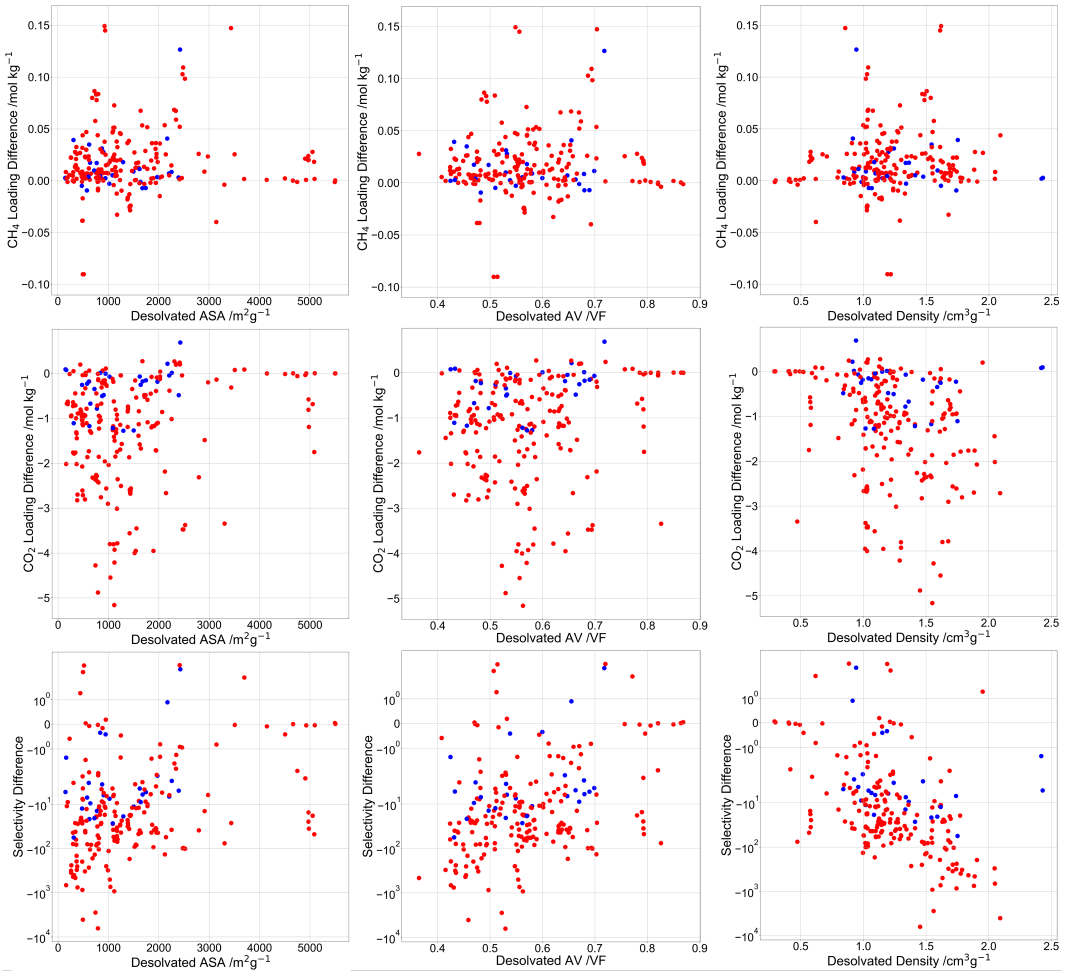

Figure 5.16: *ASA, VF and density plotted against CH_4 uptake, CO_2 uptake and CO_2 selectivity (symmetric log scale) at 0.1 bar and 298 K. Red points: the MOF has no open metal site in the solvated form. Blue points: the MOF has an open metal site in the solvated form.*

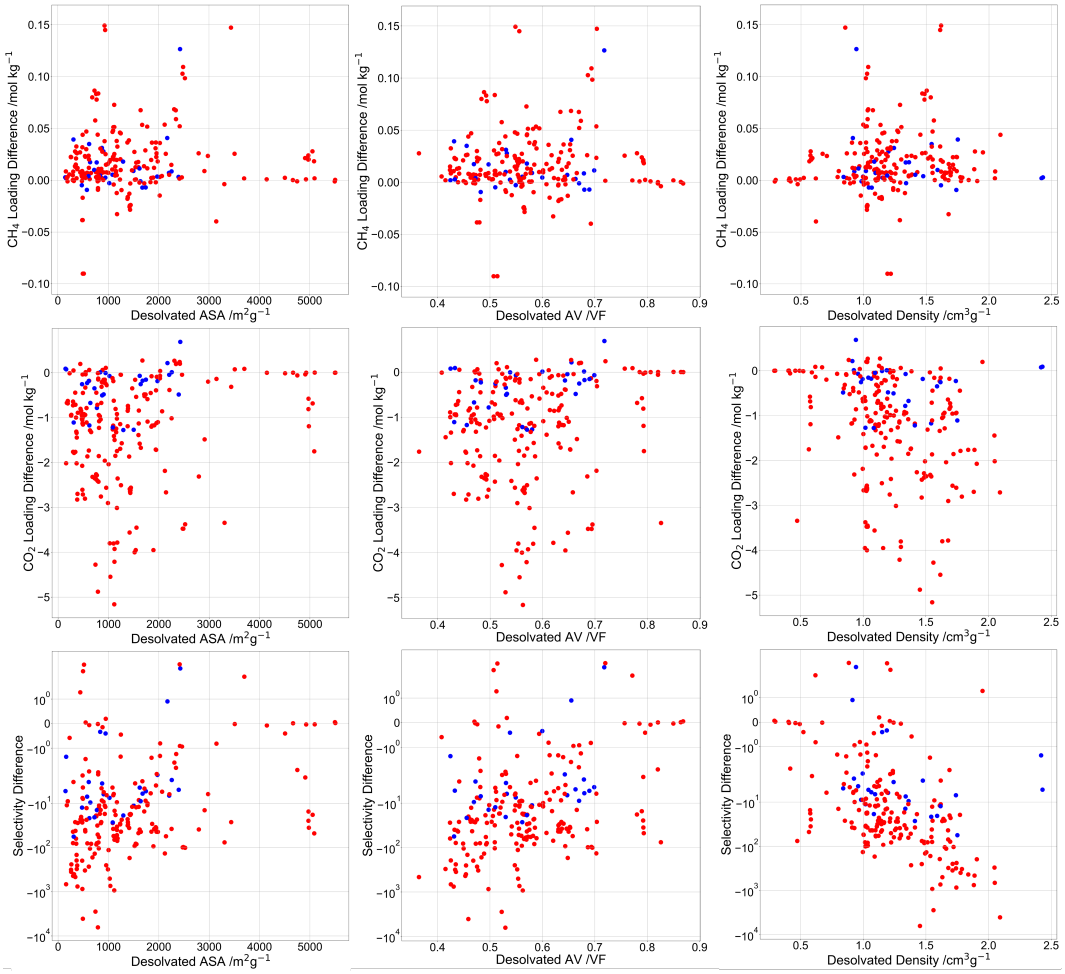

Figure 5.17: *As Figure 5.16 but at 1 bar pressure.*

gases.

An intermediate range of densities characterised by particularly positive CH_4 loading difference and particularly negative CO_2 loading difference is evident at 0.1 bar (Figure 5.16), but less well-defined at 1 bar (Figure 5.17). Furthermore, density is unusual among the geometric properties in that it correlates negatively with selectivity difference, seen at both pressures. Though the relationship is somewhat loose, in general the selectivity of denser MOFs is affected more by solvent than that of less dense MOFs.

For MOFs which possess open metal sites (OMS), interaction of guests directly with these sites has commonly been seen to be an important mechanism of uptake. [84, 258] Since bound solvent is defined as solvent interacting directly with an OMS, it may be supposed that the presence of an OMS relates strongly to solvent effect, and indeed the blocking of an OMS by solvent has previously been discussed in this context. [11] By the definitions used in the CoRE MOF database from which the structures were taken, it may be supposed that desolvated MOFs possess an open metal site. Assessment is made of the relationship between solvent effect and presence of an OMS in the solvated form of a MOF. In Figures 5.14, 5.15, 5.16 and 5.17, all points are coloured according to whether the solvated form of the MOF possesses an open metal site (blue) or not (red).

At both pressures, MOFs which do not possess an OMS in the solvated form cover both small and large solvent effects, but are able to reach more negative CH_4 loading differences, selectivity differences, and in particular CO_2 loading differences than MOFs which possess an OMS in the solvated form. Very strong negative effects are seen in cases where desolvation uncovers an OMS in a MOF which previously contained none: where there is no OMS in the solvated form, the presence of solvent can be instrumental in preventing the strongest interactions possible in a given MOF and so negatively affecting uptake. Positive solvent effects in MOFs with an OMS in the solvated form are more likely to be comparable to positive solvent effects in MOFs with no

OMS in the solvated form. Indeed, those few MOFs which display notable positive solvent effects on CO_2 uptake, including those which display the very highest positive effects, possess an OMS in the solvated form. If the strongest interactions in these MOFs are with an OMS, they need not be blocked in the solvated form. Other effects of solvent, for example in modifying the geometry of the MOF, can therefore dominate, and can readily be positive.

5.3 Effect of Solvent on Guest Binding Energy: Ab Initio Calculations

Speculation has been made in this chapter about the origins of solvent effects on CO_2 and CH_4 uptake properties, but it is instructive to make further consideration of these origins, particularly as UFF parameters may not be sufficient to model all situations encountered. Ab initio binding energy calculations are made to give further insight. Binding of CO_2 and CH_4 to example model complexes displaying different degrees of solvation is considred. These calculations may be considered a starting point for ab initio consideration of solvent effects.

5.3.1 The Model Solvated and Desolvated Complexes

The range of different types of structure in the dataset of 225 MOFs is very wide, with 24 different metals, at least seven solvents in different combinations, surface areas spanning a range of thousands of $m^2\ g^{-1}$ and pore diameters spanning a range of tens of Å. Any one of these, along with additional properties such as nature of the ligand, may have an impact on solvent effect. A large-scale and detailed ab initio analysis of solvent effect on such a range of structure types is not readily accessible. To narrow the range down, it was determined to only make ab initio consideration of one metal and one solvent, copper, the second most abundant metal which appears in a number of well-known MOFs, and water, by far the most abundant solvent.

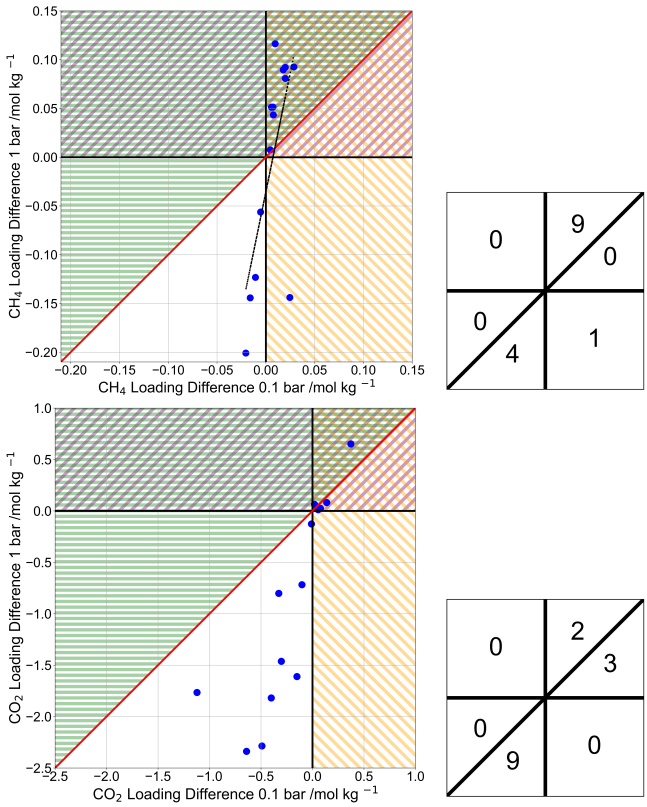

Figure 5.18: *Plots of CH_4 (top) and CO_2 (bottom) loading difference at 0.1 bar against loading difference at 1 bar as calculated from single component GCMC simulations at 298 K for the 14 MOFs with Cu as the only metal and water as the only solvent. Lines $x = 0$ and $y = 0$ are given in black and line $y = x$ is given in red for guidance. Alongside each plot is a grid showing the population of the octants and quadrants formed by the guiding lines.*

There are 14 MOFs in the curated dataset for which Cu is the only metal and water is the only solvent. Plots of single component solvent effect at 0.1 bar against single component solvent effect at 1 bar for these 14 MOFs are given in Figure 5.18. These plots are similar to those in section 5.2.3. Considering only these 14 structures reduces the range of variables which may impact solvent effect and additionally reduces the range of loading difference. However, it is clear from Figure 5.18 that not all Cu/water MOFs display the same kind of behaviour for a given gas.

In line with this, the Cu/water MOFs neither all possess similar structures nor all possess similar Cu environments. This is illustrated in Figure 5.19, which contains visualisations of each MOF. Various different Cu environments exist within the set of structures, including unusual environments, such as the

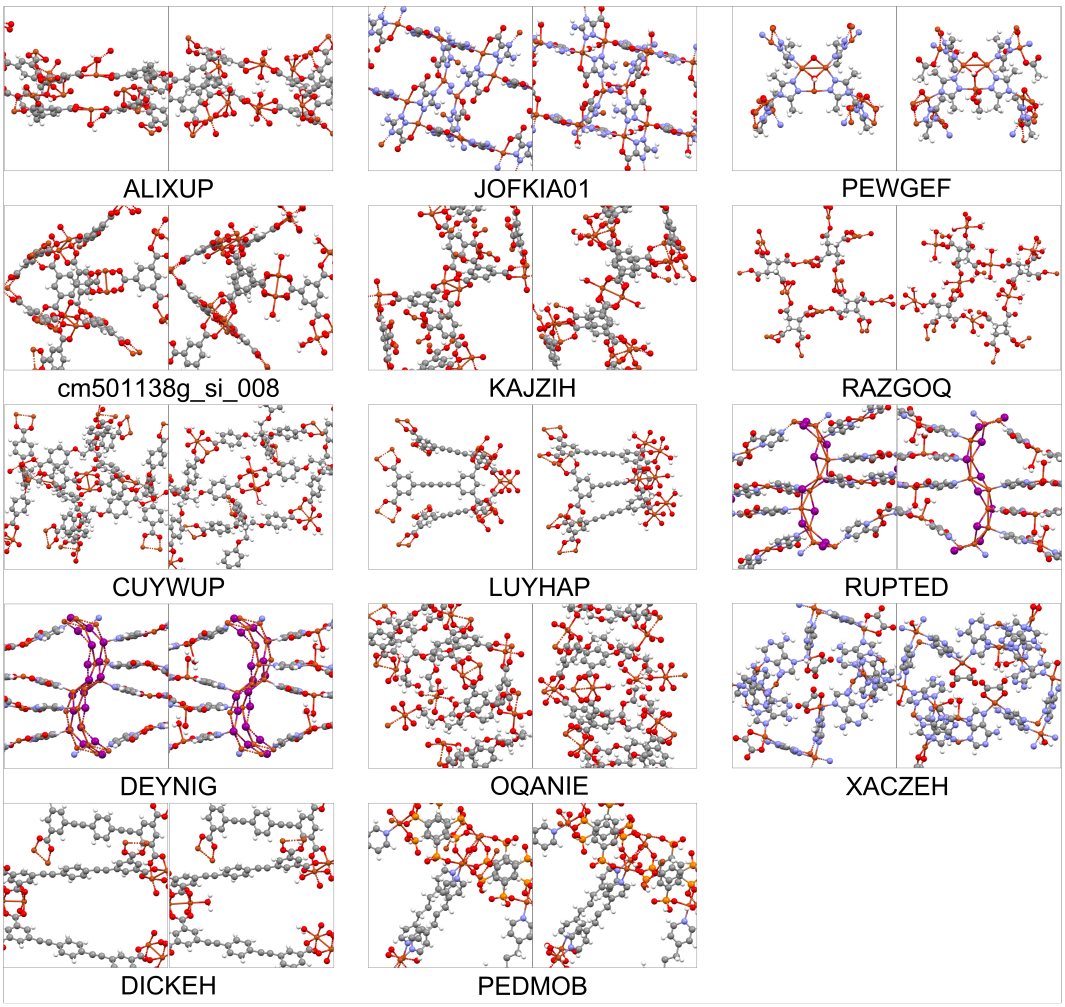

Figure 5.19: *Visualisations of the 14 MOFs in the dataset of 225 structures which contain only Cu metal and only water solvent.*

Cu chain in RUPTED and DEYNIG, which also contain more than one kind of copper site. Even among MOFs with similar copper environment, different solvent behaviour can be displayed: there is more than one Cu paddlewheel MOF in the set, but paddlewheel MOFs display both positive and negative loading difference for both CH_4 and CO_2.

The high complexity of the relationship between MOF properties and solvent effects prevents selection of a single type of metal centre representative of general Cu-water environments whose binding properties may be examined by ab initio methods. Instead, a very simple model copper complex is selected to investigate the general nature of binding to metal complexes displaying different degrees of saturation with solvent. This may initiate ab initio discussion

of solvent effects, but is not an attempt to be fully representative of any single system or system type. To represent a solvated system, the model complex used is the octahedral $[Cu(H_2O)_6]^{2+}$ complex, a copper atom surrounded octahedrally by six water molecules. To represent a desolvated system, one water molecule is removed from the complex, leaving the $[Cu(H_2O)_5]^{2+}$ complex. Both complexes were represented with a spin multiplicity of 2. This model is limited in its representation of real MOF systems. It contains coordinated water molecules in place of real organic linkers, which may affect observed binding, and it is restricted to only one geometry and electronic environment. Future consideration of further MOF models beyond the one used here will be necessary to give a full picture.

5.3.2 Calculation of Binding Energies

Using DFT, sites were sought for CO_2 and CH_4 binding to both the desolvated and the solvated model complex, and binding energies at each site calculated. The general magnitude of the differences between solvated and desolvated binding energy for the two gases can be instructive of the kind of solvent effects that may be observed. All DFT calculations were run using the QChem software package [259] and employed the SCAN functional [171] and the basis set 6-31+G* [260] with D3 dispersion corrections. [261]

Guest-free desolvated and solvated structures were first built in the IQMol visualiser [259] and underwent geometry optimisation. Optimised structures are visualised in Figure 5.20. Atoms are labelled numerically to keep track of geometry and binding site locations. In the desolvated case there is one long Cu-O bond (Cu1-O4) and four shorter Cu-O bonds, one of which is angled away from the plane of the rest (Cu1-O5). In the solvated case there are two long Cu-O bonds opposite to each other (Cu1-O2 and Cu1-O4) and the remaining shorter Cu-O bonds are approximately in plane with each other. All H atoms are aligned in the same orientation as those bonded to opposite O

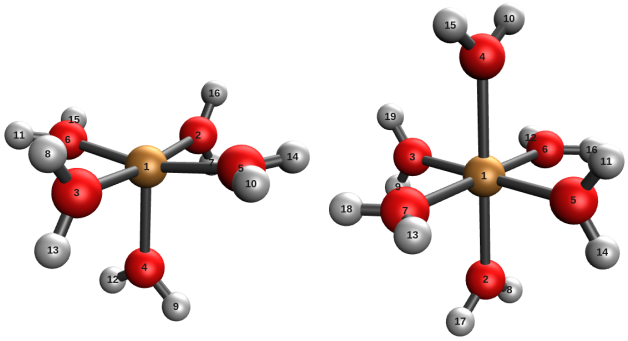

Figure 5.20: *Visualisations of the simple model complexes used to represent a desolvated (left) and solvated (right) system. Atoms are labelled numerically to keep track of binding site locations in later calculations.*

atoms, and pointed away from those bonded to their closest O atoms. Guest molecules CO_2 and CH_4 were separately built and optimised.

Each guest molecule in turn was then placed in selected possible binding sites adjacent to the the copper-water complexes, and the geometry of the whole guest-complex dimer was optimised. Binding energy following the optimisation, E_b, was calculated as:

$$E_b(guest) = E_{tot}(complex + guest) - E_{tot}(complex) - E_{tot}(guest) \qquad (5.2)$$

where *complex* refers to either the solvated or desolvated copper complex, *guest* refers to either CO_2 or CH_4 and E_{tot} is total energy. The likely binding sites attempted for the solvated form were: beside O in line with a long Cu-O bond, beside O in line with a short Cu-O bond, between two short Cu-O bonds and between a long and a short Cu-O bond. Binding sites attempted for the desolvated form were the same, with the addition of a site alongside the Cu atom. It was found that some attempted binding sites immediately collapsed to other locations on a single geometry optimisation. For others, the optimisation was followed by a slight (0.1 Å) nudge in the location of the guest and then further optimisation. This resulted in either collapse to a

different binding site or a small reduction or increase in binding energy within the same site. The labile nature of the binding locations is an indication that potential wells for CH_4 and CO_2 sorption in this complex are shallow and not well-defined in both the solvated and the desolvated form.

Taking a binding site as a guest position which was not immediately vacated on optimisation, various binding sites were found: two for CH_4 and the solvated complex, three for CH_4 and the desolvated complex, one for CO_2 and the solvated complex, and two for CO_2 and the desolvated complex. All binding locations are visualised in Figures 5.21 and 5.22 and the relevant binding energies are summarised in Table 5.4, with binding sites labelled numerically for convenience. Using the numbering system thus established, CO_2 desolvated site 2 and CH_4 desolvated site 2 and 3 collapsed after some number of nudge-optimisation procedures to CO_2 desolvated site 1 and CH_4 desolvated site 1 respectively. These more labile sites are labelled with a * in Table 5.4 and Figures 5.21 and 5.22. The binding sites thus established illustrate interactions between guest gases and the environment surrounding metal centres. However, the ability of the model complexes to illustrate direct interactions between guests and metal atoms is limited: when the guest molecules were placed directly adjacent to the Cu centres, they migrated to other binding locations on optimisation. The direct interactions observed are therefore between the guest gases and the coordinating water molecules, and the observed effects of increased solvation must be viewed in that context. Although in real MOF systems interactions of guests with more than just the metal atoms may be relevant, water molecules are unlikely to form such dominant parts of the coordinated system. It is additionally important to understand solvent effects on direct guest-metal interactions. Similar future modelling using alternative metal-containing systems may give insights in that area.

Stronger binding energies are seen for CO_2 in both the desolvated and the solvated form than for CH_4 in either form by a significant margin: the strongest CO_2 binding energy is almost double the strongest CH_4 binding energy. This

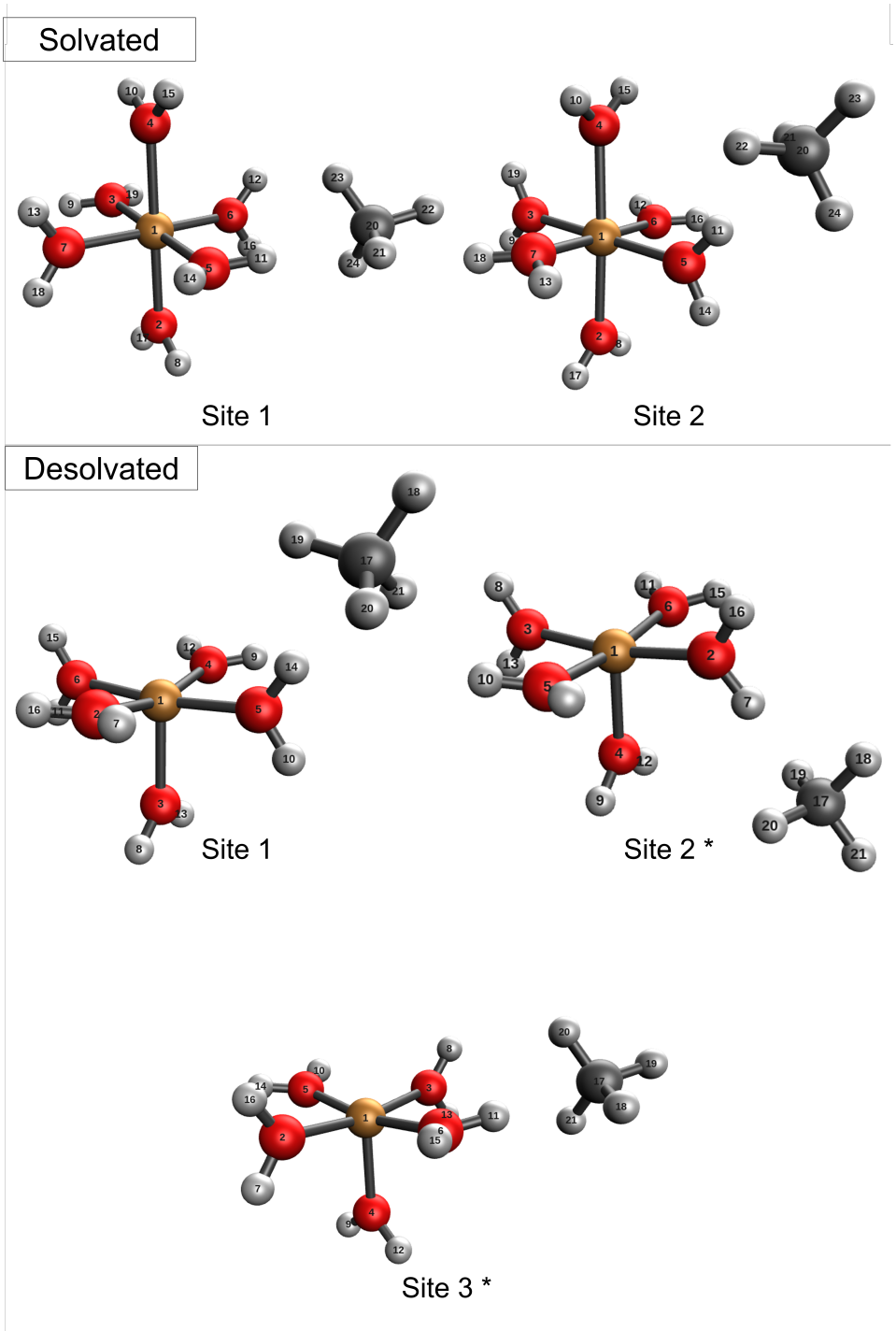

Figure 5.21: *Visualisations of the binding sites for CH_4 to the solvated and desolvated copper complexes. Sites labelled with a * were observed to collapse back to site 1 after some number of nudge-optimisation procedures.*

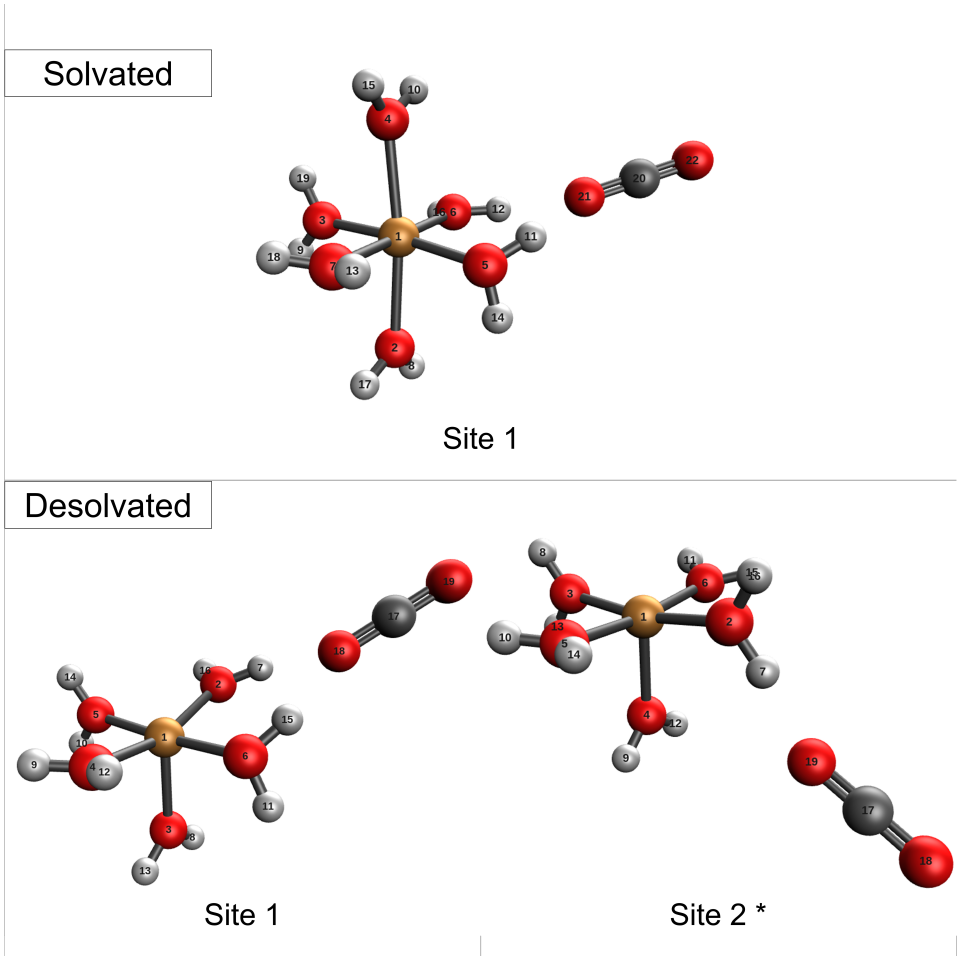

Figure 5.22: *Visualisations of the binding sites for CO_2 to the solvated and desolvated copper complexes. Sites labelled with a * were observed to collapse back to site 1 after some number of nudge-optimisation procedures.*

Table 5.4: Binding energies for the two guests in each binding site identified for both the solvated and the desolvated forms of the model copper complex.

Guest	Site	E_b /Hartree	E_b /eV
CO_2	Solvated 1	-0.01864	-0.51
CO_2	Desolvated 1	-0.02177	-0.59
CO_2	Desolvated 2 *	-0.01989	-0.54
CH_4	Solvated 1	-0.01026	-0.28
CH_4	Solvated 2	-0.01024	-0.28
CH_4	Desolvated 1	-0.01161	-0.32
CH_4	Desolvated 2 *	-0.01150	-0.31
CH_4	Desolvated 3 *	-0.01133	-0.31

is no surprise when the greater polarity of CO_2 compared to CH_4 is considered, nor when the high throughput results in Section 5.2 are examined, although in the latter the difference between CH_4 and CO_2 loading was measured and was close to tenfold.

For the model system studied, both CH_4 and CO_2 have lower binding energy in the desolvated form than in the solvated form: bound solvent has a negative effect on both binding energies. The difference between the strongest solvated binding energy and the strongest desolvated binding energy for CO_2 is 0.08 eV, while the same difference for CH_4 binding energy is 0.04 eV. The ratios of the strongest solvated to the strongest desolvated binding energy are 0.86 for CO_2 and 0.875 for CH_4. This shows, for this particular system, a stronger absolute negative solvent effect on CO_2 binding energy than CH_4, the difference in solvent effect brought about partially by the stronger total binding energy of CO_2.

It is interesting that the binding sites found do not for the chosen complex indicate a strong direct interaction of the guest molecules with the metal centre itself. Instead, binding sites which persist are in similar locations for the desolvated and the solvated forms. For CH_4, desolvated site 1 and 3 are both in similar locations to solvated site 1, although desolvated site 1 is in line with the Cu1-O5 bond and higher above the plane of the four short Cu-O bonds, while the location of site 3 is ill-defined and easily collapses. Desolvated site 2 is similar to solvated site 2. For CO_2, desolvated site 1 is in a similar position to solvated site 1, between two short Cu-O bonds, although again the desolvated site is higher above the plane of short Cu-O bonds. Desolvated site 2 is in a similar position to the CH_4 site 2s, between a long and a short Cu-O bond. In no case is a new binding site directly above the metal atom observed in the desolvated form, although the loss of the solvent molecule does allow the guest to migrate higher and closer to this position. The effect of solvent is therefore not in this case generally to block very strong binding sites, but to reduce the strength of binding in existing sites. A greater number of stable or metastable

CH_4 binding sites are seen in both forms than CO_2 sites. CH_4 sites which are not the lowest energy observed for their case are closer in energy to the lowest than is the similar CO_2 site, indicating that binding sites are better defined for CO_2 than for CH_4.

The results shown here give interesting new information about the nature of binding in solvated and desolvated metal complex-based structures. They show, for instance, the labile nature of binding sites around those structures, and an example of a case in which strong direct interaction with metal sites is not a primary contributor to interactions in the desolvated form but solvent still negatively influences binding energy. Further examination of a similar nature applied to various other metal complexes may give information about how generally these observations apply to metal complexes. A large range of possible metal complexes remain to be examined. Different geometries of copper complexes, such as the commonly occuring copper paddlewheel, may be examined, as well as other metals aside from copper (24 different metals appear in the dataset used in Section 5.2). Without changing either the metal or its initial geometry, additional charges and multiplicities may be investigated. Beyond the nature of short range attractions limited to the vicinity of the metal site, interactions in MOFs are influenced by a number of other factors, in particular ligand interactions and geometry. Further ab initio studies of interactions in full structures of solvated and desolvated MOFs would be instructive to elucidate the nature of all interactions. Cost increases with the use of large systems and the range of possible MOF structures to consider is even larger than the range of metal geometries and is already known to display diverse behaviour. However, more information is needed about the nature of interactions with solvated or desolvated MOFs, and the origins of the positive loading effects commonly seen in CH_4: geometry and ligands may play a large role here which has not yet been assessed in a systematic ab initio manner.

5.4 Conclusions

Over the course of this study, the effect on adsorption of residual solvent in MOF pores has been systematically assessed with the aid of a number of computational tools. Solvent may persist in MOF pores following its use in synthesis, and where it persists it may have a significant effect on gas uptake and selectivity. To advance understanding of this phenomenon, GCMC simulations were used first to explore uptake and selectivity of CH_4 and CO_2 in a small selection of MOFs with and without bound solvent in their pores, and then for high-throughput calculation of the same metrics for a curated dataset of hundreds of MOFs. Difference in uptake and selectivity afforded by solvent has been used to draw conclusions about the nature and magnitude of solvent effects in situations relevant to common industrial processes. Ab initio methods have been used for further exploration of solvent effects on binding energy.

Both positive and negative effects of solvent have been observed using GCMC simulations. Where solvent has a significant effect on CO_2 uptake, it is seen that the effect is almost exclusively negative under various conditions, and tends to be more negative at higher pressure. Uptake of CH_4 is more likely to be positively affected by solvent presence, though can be negatively affected also. Where it is positively affected, the effect tends to be more positive at higher pressure.

The relationship between geometrical properties of MOFs and the extent of solvent effect has also been assessed. A relationship between pore size and shape and solvent effect is seen, in which smaller pores are more likely to be affected strongly by solvent presence than larger pores. Other geometrical properties, surface area, volume and density, display similar but looser relationships with solvent effect. All iIthey cannot be used to make confident predictions for a given MOF, but they can be used to state general guidelines about the likelihood of significant solvent effect in different types of geometry.

They may additionally be able to combine to make more accurate predictions of solvent effects in MOFs using machine learning methods similar to those employed in Chapter 4.

The results of this work may be used to suggest whether and to what extent the presence of solvent is likely to be useful or detrimental in a MOF for a given application. They may also be used to inform future studies attempting to establish further statistical relationships between MOF properties and solvent effects.

Exploration of binding energy using ab initio DFT calculations on model metal complex structures has additionally been instructive. Binding energies of CO_2 and CH_4 to models representing a solvated and a desolvated copper centre have demonstrated cases of negative solvent effect on binding energy to the environment surrounding a metal centre, and have illustrated the labile nature of binding sites to this complex.

Chapter 6

Electron Beam Resilience of Two-Dimensional Metal Organic Frameworks

Up to this point, this work has been dominated by consideration of MOFs for gas adsorption properties. Various computational methods that may be used to obtain information about MOFs and their gas sorption processes have been applied and assessed. In this area, representative MOF structures generally called to mind tend to be three-dimensional. However, as discussed in section 1.1.2, 2D MOFs are emerging as a sub-class of considerable interest. Like 3D MOFs, they can display permanent porosity and take part in many of the processes commonly associated with the family, such as gas adsorption and catalysis. [57, 73] Their additional propensity to display conductivity and charge transport properties may lead to many interesting and unique applications, including in next-generation battery technology. [57, 74, 262] Two dimensional MOFs which display significant conductivity are commonly termed 2D c-MOFs. Their conductivity and other properties depend on fine control of synthesis conditions impacting structure, stacking mode, thickness and defect presence, which remains a challenge. [26]

In developing 2D MOFs for novel applications, as well as in developing 3D

MOFs, characterisation of structure is necessary and instructive. Structural determination of MOFs is commonly done by x-ray diffraction (XRD) methods [25, 74, 101]. Single crystal XRD (SC-XRD) is the preferred XRD method, able to give information on space groups, unit cell parameters, atomic coordinates, bond lengths and angles and packing modes. [74] However, it can be difficult to resolve positions of light atoms using XRD due to their comparatively weak scattering of x-rays, while disorder and interpenetration can lead to complex diffraction patterns and uncertainty, and structural determination requires an anticipated net topology. [25] The presence of residual solvent molecules (see chapter 5) additionally obscures determination of true MOF structures. [25, 101] Furthermore, it is often not possible to obtain crystals of sufficient size and quality for SC-XRD, and the alternative powder XRD (PXRD) is necessitated, which can less readily give detailed information about atomic positions.[25, 74]

Transmission electron microscopy (TEM) is a powerful alternative method of structural determination that can be useful for complex systems such as MOFs. It makes use of an electron beam (e beam) microscope and may be applied to smaller crystals than SC-XRD thanks to the strength of interaction of electrons with matter, which is higher than that of the equivalent interaction of x-rays with matter. [127] TEM has been used in a number of characterisation studies for MOFs with crystal structures inaccessible to XRD techniques. [263, 264] In particular, high resolution TEM (HR-TEM) is capable of obtaining images which can reach atomic resolution. It can give real-time pictures of structures and unambiguously identify specific crystallographic features and defects [127] where SC-XRD and in particular PXRD cannot. It additionally has the potential to directly make in situ observations of chemical transformation processes of MOF structures [264] and processes occurring within their pores. [125] In a similar vein, it has been used for structure modification: if conditions of electron beam exposure are carefully controlled, the electron beam may be a tool to induce reactions of interest and bring about desired properties, including adsorption properties. [125, 126]

The influence of electron beam radiation on the structure of a MOF, while in some cases a useful tool, is a double-edged sword more commonly known by its other edge. It is one of the major barriers to widespread use of TEM in MOF imaging. In order to obtain high-resolution images of any sample under normal conditions, a high energy electron beam is required, with voltages of 300 keV common and lower voltages of 80 keV or less also accessible. [131] In addition to imaging, the electron beam inevitably transfers energy to the sample with which it interacts, and the harsh conditions required for high image resolution can lead to transfer of sufficient energy to bring about significant damage up to total decomposition of the materials to be imaged. MOFs, unfortunately for those wishing to successfully image them, are generally considered to be particularly sensitive to electron beam damage. [127, 129, 130]

Understanding the origins of this beam sensitivity is a formidable task of high mechanistic and fundamental importance and requires knowledge of the different possible beam damage routes. Beam damage may have origins of various kinds. [133, 265, 266] Damage can be split into that caused by interaction of the electron beam with a nucleus (electron-nucleus damage, which is often related to elastic scattering) and that caused by interaction of the electron beam with the electron density of a sample (electron-electron damage, often related to inelastic scattering) Under electron-nucleus processes, the incident electron interacts directly with the Coulomb field of a sample nucleus. At sufficient scattering angles it can transfer to the nucleus kinetic energy of up to tens of eV (or larger in select cases). The transferred energy may be above some threshold for bond breaking or other damage. [265] Damage thus produced is known as knock-on damage and is discussed in greater detail in section 6.1. Under electron-electron scattering processes, the incident electron interacts with the surrounding atomic electrons and transfers sufficient energy to them to cause some damaging effect. [265] Electron-electron damage effects can be further split into various types. Beam induced heating may be significant if thermal conductivity is low. [127, 265] Electrostatic charging fol-

lows release of secondary electrons or entrapment of incident electrons in the system; the mechanism followed of these opposing options depends on sample thickness. [265] Electrons may also be promoted short of ionisation, and decay of these exciton states can be sufficient to cause atomic displacement. [265] Bond cleavage resulting from these electron-electron damage events is termed radiolysis.

Damage events may therefore be broadly split into the two categories knock-on damage and radiolysis. These two kinds of damage have previously been considered to be quite separate, although with structures susceptible to both mechanisms at the same beam energy observed. [266] In general, radiolysis dominates at low beam energies below 70 keV, and knock-on damage at high beam energies above 200 keV, though with appreciable contribution from radiolysis also. [266] This is related to the fact that in the energy range relevant to TEM accelerating voltages rate of radiolysis decreases with increasing beam voltage, and rate of knock-on damage increases. It has recently additionally been shown that the two types of pathway can behave synergistically, with radiolysis generating excited states with reduced knock-on thresholds in MoS_2, particularly at intermediate energies. [133, 267]. This complicates the already complicated landscape of efforts to reduce beam damage, particularly as certain advised methods for reducing knock-on damage (reducing beam voltage) conflict with those for reducing damage from radiolysis (increasing beam voltage). [127]

Several possible damage mechanisms are therefore available to MOFs, and it has been suggested that their high susceptibility to damage may originate from a number of sources, with relevance of both knock-on damage and radiolysis. [127] The metal atoms that form nodes may be ionised, inducing weakening of metal-ligand bonds. Metal atoms, particularly Cu and Ru, also appear to be susceptible to reduction to metal nanoparticles. The celebrated high surface areas of MOFs can be detrimental to beam resilience as they can lead to high surface energy and therefore significant release of energy on

structural collapse. [127] The contribution to damage that comes from organic ligands is likely to be similar to analogous damage occurring in organic molecules, which are notoriously beam-sensitive. [127] For example, susceptibility to knock-on damage depends heavily on mass, and it has previously been observed in coronene that H atoms can have an extremely high contribution to damage. [131] Further contributions may come from electrical conductivity, or lack thereof, and crystallinity. For electrical conductivity, MOFs are more commonly electrically insulating than electrically conductive, and electrically insulating materials are commonly more susceptible to beam damage than electrically conducting ones, because electrons promoted by the electron beam are unable to quickly relax to the ground state. [127, 268–270] For crystallinity, it is necessary to align the electron beam to crystallographic axes, and traditional methods to do this require extra time under the electron beam for the alignment and therefore a larger total electron dose. [129]

All the above mentioned factors together may contribute to beam sensitivity of a general MOF, but the reduced dimensionality of 2D MOFs complicates matters further by introducing factors which can both increase and decrease stability of samples. [271] In 3D structures, secondary electrons emitted by interaction of the beam with the electron density of the sample may themselves subsequently interact and cause further damage. This effect is greatly reduced in few-layer materials to the extent of being almost non-existent in atomically thin samples. Damage locations in 2D samples are also more isolated than those in 3D structures, reducing the chance of further reaction once a damage site is created. However, this reduction in neighbours can also contribute to damage susceptibility: outside of the bulk, a far greater proportion of atoms are a part of the surface, having more dangling bonds and lower thresholds for initial knock-on damage events than bulk atoms. An additional effect of reduced dimensionality which applies to MOFs is a possible increase in electrical conductivity. The common electrically insulating properties of 3D MOFs leave them particularly susceptible to radiolysis, but 2D MOFs are more likely

to be electrically conductive than their 3D counterparts, and so may lack this contribution to damage susceptibility.

Thus far, though its origins have been discussed, the actual event of damage has been treated here as a rather nebulous concept, covering some kind of structural change up to total decomposition that occurs following transfer of beam energy to a sample. Events falling under the full range of this definition may certainly occur, but a more useful classification of damage types is conducive to a focused discussion of those which are most relevant. In that spirit, observed damage includes the following types: atom displacement within a sample amounting to rearrangement, bond cleavage, and complete ejection of atoms or fragments. The first two, displacement and cleavage, can affect the structure of a MOF immediately following impact, but knock on damage of this kind can be expected to be repaired on the timescale of electron impacts [132, 268] and so is unlikely to result in permanent changes to the integrity of the sample and therefore unlikely to be relevant to decomposition of a sample. By contrast, the third kind of damage, complete ejection, involves removal of an atom or fragment from the vicinity of the rest of a sample. Such damage can generally be considered to be irreversible: there can be no recombining without some other source of the ejected material. Ultimately, therefore, subsequent electron impacts will be with an already damaged sample which has greater susceptibility to damage itself. Although amorphisation may be possible without ejection, ejection can be treated as a dominant precursor to full amorphisation or destruction of a crystalline MOF sample.

When permanent damage occurs, as it does readily in MOFs, it is a barrier to useful imaging; under standard imaging conditions MOFs are habitually fully amorphised before an image can be obtained. Under such circumstances, clearly TEM is less useful than standard XRD methods. However, advances in imaging technology can help to allow MOF samples to be imaged with atomic resolution in select cases. [129, 130] A number of possible routes have been explored. Carrying out TEM at very low temperatures can limit beam-induced

heating, but this is only one small contributor to damage. [127, 129] Reducing the energy of the electron beam can reduce the energy transferred to the sample and reduce knock-on damage, but a low energy electron beam is accompanied by low image resolution. In any case, certain light atoms display increased, rather than reduced, damage susceptibility under a low-energy electron beam. [131] The recent advance which has had the greatest contribution to MOF imaging quality has been the development of imaging methods requiring very low total electron dose. In general, samples possess some critical electron dose above which amorphisation occurs. If an image can be obtained with a low total electron dose, the sample is required to interact with fewer electrons and the chance of obtaining an image before amorphisation is increased. For the success of imaging with low doses, advanced machinery is needed which can detect very small numbers of electrons. The direct-detection electron-counting (DDEC) cameras first developed for biological systems [128] fulfil that brief, and have been combined with low dose rates to obtain high quality images of MOF systems. [130] The success of this method has been increased by development of methods for automatic alignment of the beam with crystallographic axes. [129] A significant drawback associated with these methods, however, is the high cost of DDEC cameras. They are not readily available and therefore cannot be used for every MOF characterisation problem.

With advanced imaging methods for MOFs selectively available, understanding of beam damage mechanisms is clearly desirable. In addition to the fundamental knowledge this understanding can furnish, it may direct efforts to synthesis of MOFs most likely to be resilient to beam damage in cases where structural elucidation is important but a DDEC camera is not accessible. However, such understanding is elusive: indeed, prediction of the electron beam stability of a MOF has been described as "basically impossible". [127]

Attempts to surmount this formidable task can involve use of computational methods to probe the state of a sample following e-beam interactions leading to specific damage events. Damage to various parts of samples pos-

sessing different features may be computationally examined, probing the relationship between these features and the extent of damage. The success of such an endeavour depends on the computational methods selected to model the system. Classical force fields of a general nature such as the UFF [4] and Dreiding [151] Lennard-Jones force fields used for large-scale modelling of guest interactions with MOF frameworks in chapters 3, 4 and 5 are certainly not sufficient. They are not parameterised for the bonded interactions that occur between framework atoms and are not set up for the breaking of bonds. More complex molecular force fields represent some improvement, as they make more complete consideration of bonding and many-body interactions, but are still unlikely to be sufficient to describe the complex nature of the bond-breaking interactions studied, and are not always set up appropriately for the system at hand.

Ab initio methods are required for a problem of this nature. Several ab initio methods exist with differing levels of complexity which may be used to model various beam damage events. Knock-on damage can be modelled far more readily than radiolysis, as it can in principle be treated by standard ab initio methods within the Born-Oppenheimer approximation, taking no account of electron dynamics. Density functional theory (DFT) is a common choice in Born-Oppenheimer chemical modelling. However, even allowing for neglect of electron dynamics, DFT results are dependent on the specifics of selected methods and are often not well-suited to modelling electronic properties of MOFs, as discussed in Chapter 1 and section 2.2.1. In particular, if pure DFT local LDA and GGA functionals are used, band gaps are systematically underestimated, [97–99] leading to an inaccurate description of the system. This is particularly true when the MOF is based on an open-shell metal which engenders high correlation. [96] Open shell metals may be seen in MOFs and may also occur as a result of damage events, meaning that unrestricted Kohn-Sham (pairs of electrons occupying different spatial orbitals) is necessary and that LDA and GGA functionals are unlikely to give a good description of the

system. Pure LDA and GGA methods occupy the lower rungs of the Jacob's ladder of DFT methods discussed in section 2.2.1. Progressing up the rungs of the ladder generally leads to increased accuracy but also to increased computational cost, leading to the existence of a trade-off relationship between cost and accuracy. The progression involves an increase in the amount of non-local information included in the approximation and can involve the use of hybrid functionals which describe exchange-correlation using a combination of DFT and some other method constrained by different limitations. Use of hybrid functionals can improve description of band gaps at increased cost, [98] but may be prohibitively expensive, particularly when used for dynamics and for solid state systems, [272] and limitations of the introduced method must also be considered. [169, 170] Beyond knock-on modelling, computational methods additionally exist by which electron dynamics can be included and radiolysis can be modelled, as has been done using Ehrenfest dynamics, [133] but the cost and complexity of methods to treat radiolysis is heavily increased compared to cost of methods to treat knock-on damage. [273]

Here, increased understanding of knock-on beam damage to MOFs possessing various structural features is pursued. The knock-on damage experienced by four selected 2D MOF structures is modelled using Born-Oppenheimer ab initio molecular dynamics (AIMD). MOFs selected for this analysis all display the same structural backbone and differ by some key features. By this setup, the influence of these key features on structural damage may be assessed. The selected 2D MOFs are: Cu_3(hexaimniobenzene)$_2$ (referred to here as HAB), Cu_3(hexahydroxybenzene)$_2$ (referred to as HHB-2), Cu_3(hexahydroxybenzene) (referred to as HHB-1) and Cu_3(benzenehexathiol) (referred to as BHT). Representations of portions of the periodic structure of each are illustrated in Figure 6.1. The four MOFs are all copper-based with structures based on a Kagome lattice and ligands of heteroatom-capped benzene. They differ most significantly in the heteroatom bonded to the metal. The MOF HAB contains NH groups, and is the only MOF of the four which contains hydrogen, so by its

inclusion the contribution of very light H atoms to damage under the conditions studied can be assessed. For damage to organic molecules at 80 keV, hydrogen was previously seen to be descisive. [131] Of the remaining three MOFs, HHB-2 and HHB-1 are both oxygen-containing, while BHT is sulphur-containing. The contribution of these different heteroatoms to damage can therefore be assessed, an important avenue to take as the effects of organometallic bonds on beam damage are not well studied. Mass is known to be strongly linked to beam resilience, so the twofold mass increase between O and S is a point of interest. In addition to atom types, the MOFs differ in density. Both HAB and HHB-2 are porous, while HHB-1 and BHT are denser, containing an extra ligand where the other two MOFs have pores. Direct comparison here can be made between HHB-2 and HHB-1, which are otherwise equivalent. A final important point of difference among the family of structures is electrical conductivity, which has already been mentioned as a feature which can affect damage susceptibility. [127, 268–270] Among the four selected MOFs, BHT is highly electrically conductive, with a measured electrical conductivity of 2,500 S cm^{-1}, HAB is somewhat electrically conductive (electrical conductivity = 13 S^{-1}) and HHB-2 and HHB-1 are electrically insulating (electrical conductivities of 7.3×10^{-8} S cm^{-1} and 2.6×10^{-2} S cm^{-1} respectively) (U. Kaiser; H. Qi, 2022, private communication, 29th March). Including structures with differing electrical conductivity allows assessment of the effects of electrical conductivity on beam damage. If effects of electrical conductivity are significant, it may be expected that the more electrically conductive MOFs display higher e beam resilience. The increased damage susceptibility of electrically insulating structures relates to radiolysis, which as discussed may not be modelled as readily as knock-on damage. Therefore, in cases where electrically insulating structures are not well-described by knock-on modelling, there is an indication of the electrically insulating nature influencing beam damage.

The focus of this work is on beam damage at 300 keV. This is a common energy at which to image structures. It is a sufficiently high energy that

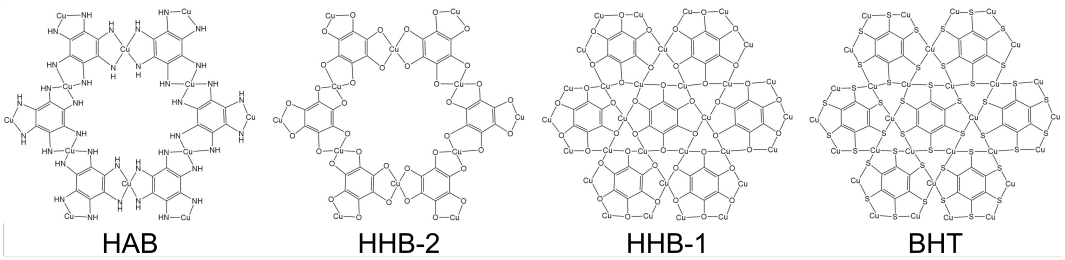

Figure 6.1: *Representations of portions of the periodic structures of all four MOFs selected for study in this chapter.*

knock-on damage is likely to be more dominant than radiolysis. [266] Damage susceptibility of all four MOFs has been measured experimentally at 300 keV and it has been found that BHT is by far the most resilient to damage, followed by HHB-1, HHB-2 and HAB in that order. (U. Kaiser; H. Qi, 2022, private communication, 4th July) Since the computational models used here only probe knock-on damage, they may capture effects relating to the nature of the atoms or to density, but not directly to electrical conductivity as the effects of electrical conductivity relate to radiolysis damage. Under an electron beam of 300 keV, radiolysis and electrical conductivity effects would not normally be expected to be significant for these types of materials.

6.1 Methods

Electron beam damage following impacts of a 300 keV TEM electron on each of the constituent atoms of the four MOFs was computationally simulated and mathematically examined to obtain insight into the damage pathways which dominate amorphisation processes in the MOFs. The interaction of each atom with the electron beam must be understood, and in this context the atom which experiences the electron impact may be referred to as the primary knock-on atom (PKA). The damage primarily considered here is damage leading to permanent ejection of either the PKA, some other atom, or some fragment which may or may not contain the PKA. Focus on only permanent ejection is justified by timescales as discussed in section 1.3. Considering the time

between electron impacts in TEM, the approximation is made that damage short of permanent ejection caused by an impact is repaired before any subsequent impact, achieving restoration of equilibrium, and thus restoration of crystallinity. [132, 268] Permanent ejection, however, is treated as a precursor to full amorphisation. Only direct knock-on damage is considered computationally: the methods make no consideration of electron dynamics, neglecting effects of electron-electron damage events. The impact of radiolysis compared to knock-on damage has previously been shown to decrease as beam energy increases, faster electrons having less time to induce excitation. [269, 270] Knock-on damage is expected to be the dominant pathway at a beam energy of 300 keV, and more generally for electrically conductive materials, in which excited electrons relax easily to the ground state between electron impacts. [268–270]

Knock-on damage may be described in terms of energy transfer as follows. During TEM, the electron beam has an energy of some E eV. During a knock-on scattering event, some T eV of kinetic energy is transferred to the nucleus of the PKA. Each atom may be described by an ejection threshold energy E_d, which is determined by the atom type, the environment in which it resides, and the angle of transferred energy. If the transferred energy T is greater than the relevant threshold E_d the atom can be ejected; otherwise the atom, though it may be disturbed, remains bonded to the sample. In order to determine E_d for the constituent atoms of the four MOFs, each atom in turn was assigned the role of the PKA. Ab initio molecular dynamics (AIMD) simulations (see Section 6.1.1) were conducted for each PKA at a series of transferred energies to determine whether permanent ejection occurs for the selected T. The threshold E_d for an ejection event is the lowest T at which that ejection event is observed.

Ejection cross-section, σ_d, for a PKA at a given electron beam energy describes the probability of a particular ejection event at that beam energy. It may be calculated from E_d and the beam energy E (see section 6.1.2). Ultimately, computational ejection cross-sections σ_d were here calculated for

each atom using E_d obtained from AIMD simulations (section 6.1.1). This done, atomic values of σ_d were combined using a weighted sum to obtain per-MOF values of σ_d (section 6.1.2). Cross-sections for individual events offer insight into the nature of beam damage processes in the MOFs. Per-MOF cross sections can be used to predict resilience of a MOF to knock-on damage, and can be compared to experimental values where available to give information about whether damage processes other than knock-on are likely to be significant for the MOF in question.

6.1.1 Ab Initio Molecular Dynamics: Computational Setup

Spin polarised ab initio molecular dynamics (AIMD) at the B3LYP/6-31G* level of theory was employed using the ORCA software package [274–276] to simulate the effects of energy transfer following electron impacts on the constituent atoms of the four MOFs and thus to obtain computational ejection thresholds E_d. The performance of a method varies depending on the system and chemical property to which it is applied; B3LYP/6-31G* is widely used, with good general applicability. [277] Indeed, applied to calculation of a range of chemical properties of organic systems, it is not atypical for 6-31G-type basis sets to equal or outperform more computationally demanding competitors, neither is it uncommon for only a small effect to be observed following expansion of the basis set by addition of diffuse functions (6-31+G* or 6-31++G*). [277] When it comes to the particular challenge of transition metal modelling, the errors associated with DFT methods are notably higher than for organic molecules. [278] Hybrid functionals such as B3LYP, while performing better than their non-hybrid counterparts for main group systems, [169, 170] come under criticism for their ability to model transition metal dimers [169] and, to a lesser extent, small metal-ligand complexes. [170] Among the transition metals, copper suffers from this issue comparatively little. [169, 170] However,

in terms of the applicability of the methods here selected to copper specifically, 6-31G-type basis sets have met some criticism regarding their ability to calculate the bond length of the copper dimer. [279] None of the four MOFs contains a Cu-Cu bond, and such bonds are observed rarely if at all over the course of the dynamics simulations. In previous comparison of a range of methods and basis sets for modelling transition metal-ligand complexes, [278, 280] B3LYP/6-31G** showed reasonable performance including for copper complexes. The 6-31G** basis set differs from 6-31G* only in the inclusion of polarisation functions on hydrogen atoms. Although by no means the most accurate computational method for copper-organic complexes, [280] B3LYP/6-31G* is used as an appropriate compromise between accuracy for the organic part of the system, accuracy for the metal, and the substantial cost of AIMD. Future in-depth comparison of the performance of alternative functionals and basis sets for the systems studied, including expansion to the 6-31G** basis set, would be valuable, though this would be a large computational undertaking.

Even employing a relatively modest level of theory, AIMD with large systems is prohibitively expensive: representative finite flakes were used in place of periodic structures, reducing cost to a computationally manageable level. The selected flakes are shown in Figure 6.2, with dangling bonds capped with hydrogen atoms. For convenience in discussion of the identity of impacted atoms and ejected fragments, atoms which at any point form part of an ejected fragment are labelled numerically, with impacted atoms labelled in red. All atoms of any given element in the MOFs are symmetrically equivalent in the periodic structure, so only one atom of each element was treated as the PKA in each flake. By necessity, atoms in the finite flakes are not all symmetrically equivalent. In cases where they are not, atoms whose environment best approximates bulk were selected. Initial tests of energy transfer to the flakes with no alteration resulted in significant distortion, rotation and translation that would not be possible in the periodic structures. Based on the tests, selected edge atoms of each flake were immobilised during simulations using an

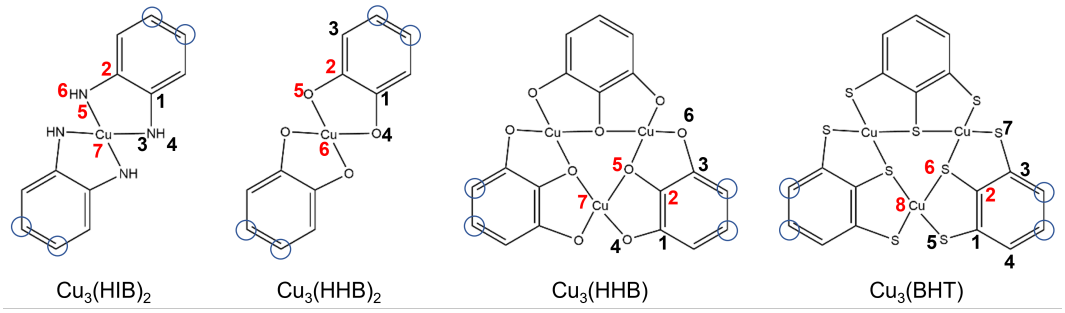

Figure 6.2: *The representative fragments used in simulations of each of the four MOFs. Atoms which form parts of ejected fragments over the course of the simulations are labelled numerically for convenience. Atoms whose labels are in red are the impacted atoms. Atoms which are circled in blue, and hydrogen atoms connected to them, are immobilised during the simulations.*

iterative constraint procedure available in the ORCA software package. [274–276] In Figure 6.2, atoms circled in blue and any hydrogen atoms connected to them are the immobilised atoms.

In order to simulate transfer of velocity to the atoms in the system, initial velocities must be assigned to the flake atoms prior to computations. Simulation is of a 2D material, so beam interaction is assumed to be perpendicular and velocity perpendicular to the plane of the flake is assigned to be equivalent to a selected transferred energy for the PKA, with zero velocity assigned for all other atoms and directions. Even for a 2D material, this remains an approximation, as the angle of transferred energy may vary experimentally. The system was allowed to propagate using molecular dynamics with timesteps of 0.5 fs until it was judged that any ejection or lack thereof was permanent, with ejection deemed permanent based on distance from the sample. This was repeated for a range of energies, with ejection of an atom or fragment, or non-ejection, observed in each case. The requisite precision specified is 0.1 keV beam energy.

6.1.2 Calculation of Cross Sections: Mathematical Details

Using the E_d for ejections determined following simulated impacts on examples of all flake atoms (with the exception of the capping H atoms which do not exist in the real structure), atomic and fragment ejection cross-sections were calculated as described by Skowron et al, [132] using the McKinley-Feshbach formula. [281] This formula arises from mathematical consideration of electron scattering as follows.

The electron scattering cross-section $\sigma(\theta)$ describes the probability of scattering of an incident electron of given velocity by an angle θ. It is based on the classical Rutherford electron scattering cross section, σ_R, (Equation 6.1) but expanded for accuracy. It is given in Equation 6.2, where z is nuclear charge, m_e is the mass of an electron, and β is the ratio of electron velocity to the speed of light c. [131]

$$\sigma_R(\theta) = \left(\frac{ze^2}{4\pi\epsilon_0 2m_e c^2}\right)\frac{1-\beta^2}{\beta^4}\csc^4\left(\frac{\theta}{2}\right) \tag{6.1}$$

$$\sigma(\theta) = \sigma_r\left(1-\beta^2\sin^2\left(\frac{\theta}{2}\right)+\pi\frac{ze^2}{\hbar c}\beta\sin\left(\frac{\theta}{2}\right)\left(1-\sin\left(\frac{\theta}{2}\right)\right)\right) \tag{6.2}$$

For a collision leading to elastic scattering, the energy transferred to a static atom T is given by Equation 6.3, where M is the mass of the atom, E is the e-beam energy, and again θ is the scattering angle.

$$T = \frac{2ME(E+m_e c^2)}{(M+m_e)^2+2ME}sin\left(\frac{\theta}{2}\right) \tag{6.3}$$

The maximum possible energy transferred, T_{max} is observed when complete backscattering occurs. That is, $\theta = 180°$, meaning $\sin(\theta/2) = 1$. T_{max} for energy elastically transferred to a static atom is given in Equation 6.4. Note that T_{max} (like T) depends on M and E, and is inversely proportional to atom

mass.

$$T_{max} = \frac{2ME(E + m_e c^2)}{(M + m_e)^2 + 2ME} \tag{6.4}$$

Rather than considering the PKA to be static as in equations 6.3 and 6.4, a closer match to experimental data has been shown to be available [282] if the distribution of possible initial velocities of the PKA due to lattice vibrations is taken into account. This distribution depends on the conditions of the system and, for the purposes of this method, on the model used to describe it. Here, the probability distribution of all velocities is obtained from the Maxwell-Boltzmann distribution (equation 6.5), where k_b is the Boltzmann constant and temperature is referred to as T_{mp} in order to differentiate it from the transferred energy T.

$$P(V_n, T_{mp})dV_n = \sqrt{\frac{M}{2\pi k_b T_{mp}}} exp\left(-\frac{MV_n^2}{2k_b T_{mp}}\right) dV_n \tag{6.5}$$

This velocity distribution is implemented by using an updated definition of T_{max} given by equations 6.6, 6.7 and 6.8. In these equations, T_{max}, E, M and c are the maximum transferred energy, the beam energy, the mass of the PKA and the speed of light as before, while V_n is the initial velocity of the PKA and E_n is the initial kinetic energy arising from it.

$$T_{max} = \frac{1}{2M}\left(r\left(r + \frac{2t}{c}\right) + \left(\frac{t}{c}\right)^2\right) \tag{6.6}$$

$$r = \frac{1}{c}\sqrt{E(E + 2mc^2)} + MV_n \tag{6.7}$$

$$t = \sqrt{(E + E_n)(E + 2mc^2 + E_n)} \tag{6.8}$$

These equations allow T_{max} to be calculated for a given initial velocity of the PKA. An average T_{max} for the distribution of initial velocities of a PKA

experiencing lattice vibrations is obtained by integration of equation 6.6 over the range of velocities in the distribution. This may be applied with velocity distributions other than the Maxwell-Boltzmann distribution used here, if suitable for the system in question. [132, 269] The Maxwell-Boltzmann distribution gives a good general description and does not require knowledge of any additional constants for the system.

To determine the ejection cross section, σ_d, the scattering cross-section must be expressed in terms of T rather than θ and integrated over the range of energies for which ejection occurs. That is, it is integrated in the range $E_d < T < T_{max}$. This results in an expression for σ_d with a complex dependence on T_{max} and E_d, given in Equation 6.9.

$$\sigma_d = 4\pi \left(\frac{ze^2}{4\pi\epsilon_0 2m_e c^2}\right)^2 \frac{1-\beta}{\beta^4} \left(\frac{T_{max}}{E_d} - 1 - \beta^2 \ln\left(\frac{T_{max}}{E_d}\right) + \pi\frac{ze^2}{\hbar c}\beta\left(2\left(\frac{T_{max}}{E_d}\right)^{1/2} - \ln\left(\frac{T_{max}}{E_d}\right) - 2\right)\right) \tag{6.9}$$

Since σ_d is proportional to T_{max}, it is inversely proportional to the mass of the atom: more energy is transferred to lighter atoms, increasing likelihood of ejection. This is somewhat at odds with the proportionality of σ_d to z, nuclear charge, but the mass dependency dominates at low e-beam energies.[131]

If ejection cross-sections for every relevant atom are known over a range of e-beam energies, conclusions can be drawn about which parts of a sample are likely to undergo knock-on damage at these energies. For example, it has been shown for coronene that H atoms have very large ejection cross-sections at low energies where heavier atoms' cross-sections are small. [131]

Many of the quantities required to calculate σ_d are known or calculable from knowledge about experimental setup. The threshold energy is not and is instead determined here from a series of AIMD simulations as described in

section 6.1.1.

Once computational atomic ejection cross sections for all atoms within a MOF are known, individual cross sections are combined into total ejection cross-sections for each of the four MOFs using a weighted sum over all atoms of the relevant type, as in Chamberlain et al. [283] In the work of Chamberlain et al, total cross-sections for damage are given per molecule. For the periodic structures used here, total cross sections are calculated per copper atom, which is equivalent to per empirical formula unit. Critical dose for ejection per copper atom, the inverse of the total ejection cross section per copper atom, is then obtained. This may be compared to experimental values of critical electron dose for amorphisation.

6.2 Results and Discussion

6.2.1 Total Cross Sections and Fragmentation Patterns

Ejection cross sections were determined for a number of ejection events relevant to the atoms of the four MOFs. Both atomic cross-sections and total per-Cu cross sections reveal interesting insights into their beam resilience.

The AIMD simulations reveal various fragment ejection patterns over the energy ranges considered. Patterns are complex and continuously evolve as transferred energy changes, including in ways which may not be expected. Observed events following beam impact can be categorised, and here particular distinction is made between three cases: no ejection, intermediate fragmentation pathways, and the primary ejection pathway. These are defined here as follows. First, at low transferred energy, there is no ejection. Then there are intermediate energies at which the PKA is disturbed but returns to the fragment. At these energies, there are two possibilities. The first is that no ejection is again observed; the second is that, following return of the PKA, a small fragment ejects. The small fragment can contain the PKA but does

Table 6.1: Computational cross sections per Cu atom at 300 keV for all four MOFs.

	All pathways σ_d /barn	Primary pathway σ_d /barn
HAB	468	445
HHB-2	289	220
HHB-1	179	114
BHT	167	164

not always. Such situations are referred to here as intermediate fragmentation pathways. Finally, as transferred energy continues to increase, there is a point above which the PKA ejects and does not return, referred to here as the primary ejection pathway. With the exception of the N impact in HAB, the primary ejection pathway involves atomic ejection of the PKA alone. For the N impact in HAB, the ejection is of the NH fragment.

Total per-Cu computational knock-on cross sections for the four MOFs are summarised in Table 6.1. Two sets of values are presented. In the first column, intermediate fragmentation pathways are included along with the primary ejection pathway, whereas in the second colum only the primary ejection pathway is considered. A smaller cross-section corresponds to a lower probability of permanent ejection and therefore higher resilience to a 300 keV electron beam. The difference between the cross section of the primary ejection pathway and the total cross section calculated from all pathways gives information on the importance of complex fragmentation patterns to the breakdown of the MOF.

Examining damage in each MOF in turn is instructive in illustrating more clearly the various fragmentation pathways observed and their changing nature with transferred energy, as well as which ejection pathways dominate damage generation.

HAB Damage Pathways

First, the N-containing MOF HAB is the only MOF with H atoms. Its computational cross-sections suggest it is the MOF least resilient to the electron beam, and it may be expected that the presence of H plays a part in this, as it has previously been shown that due to its very low mass, H may be lost

Table 6.2: Ejection thresholds and cross sections for e beam impacts on the atoms in the HAB flake. Using the system established in Figure 6.2, the atoms and fragments directly referred to are as follows. N: N5; NH: N5H6; C: C2; H: H6.

PKA	Ejected atom(s)	Transferred energy /eV	Approx. beam energy /keV	σ (300 keV) /barn	Total σ for PKA/barn
N	NH	11.33	67.8	46	46
C	Fragments	16.88	85.2	6	24
	C	20.16	100.4	18	
H	H	4.83	2.2	47	47
Cu	None	-	-	0	0

orders of magnitude more readily than main group atoms in organic molecules at 80 keV. [131] The dominance of H in damage pathways, however, reduces at increased beam energies and its contribution to metal-organic structures has not until now been assessed in detail at any beam energy. The origin of the low beam resilience and the contribution of H to that phenomenon must therefore be examined. There is additionally an appreciable difference between the all pathways cross section and the primary pathway cross section of HAB. Contribution of intermediate ejection pathways must therefore be addressed.

Figure 6.3 and Table 6.2 give further information about fragmentation patterns in HAB. Figure 6.3a displays number lines showing at which transferred energies different ejection or non-ejection events are observed following simulated impacts on the atoms of the HAB flake. Figure 6.3b and c show plots of cross section for various ejection events over a range of beam energies. Table 6.2 summarises the thresholds for the ejection events and the resulting cross sections. Table 6.2 includes information on the possible atoms or fragments that can eject following impact on each PKA and the ejection threshold of transferred energy for each fragment. From this, the table gives an approximate corresponding beam energy threshold, calculated from the transferred energy assuming a stationary PKA. It also gives a cross section for each ejected fragment, followed by a total cross section for ejection following impact on the PKA.

The number lines in Figure 6.3a cover N, C and H, but not Cu, and Figure

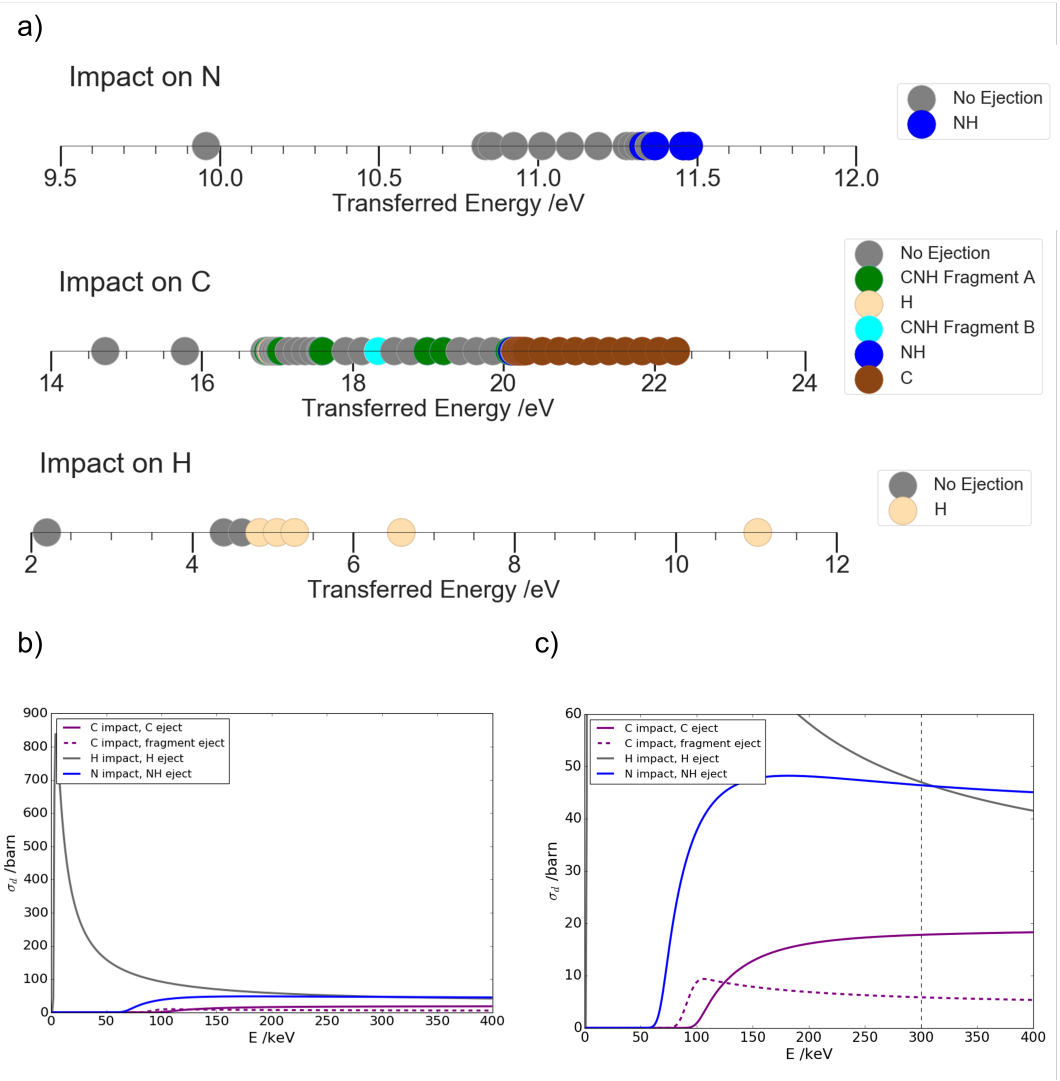

Figure 6.3: *Details of HAB ejection thresholds and cross sections. a) Number lines showing fragmentation events over a range of transferred energies. b) Plots of computational knock-on cross sections following impacts on each atom type split by fragmentation event. c) The same plots with reduced y-axis limits for clarity. Using the numbering system established in Figure 6.2, the ejected atoms and fragments are identified as follows. NH: N5H6, H: H6, C: C2, CNH fragment A: C2N5H6, CNH fragment B: C1N3H4.*

6.3b displays no line for impact on or ejection of Cu: impacts were attempted for all atom types, but no ejection was observed following impact on Cu for any T up to and including 13.52 eV, which corresponds to the maximum energy transferred to a stationary Cu atom by a 300 keV beam. It is clear from Figure 6.3a that only two possible outcomes were observed following impacts on N and H, and the two are cleanly separated: no ejection or the primary ejection pathway. These two pathways dominate damage at 300 keV, as seen in Figure 6.3c. Thus, H is involved in providing damage pathways which are important to the amorphisation of the HAB flake. It does not, however, dominate damage pathways to the same extent seen at 80 keV for organic molecules.

Impacts on C also lead to relevant damage pathways, as illustrated by the solid and dashed purple lines in Figure 6.3b and c. Unlike those on N and H, impacts on C may lead to a range of outcomes as modelled by the current setup, with thresholds not always clearly defined. A fuller picture would be more complicated still, with the angle of transferred energy further influencing damage pathways. The lowest transferred energy at which ejection is observed is 16.88 eV, and at this energy the CNH fragment is ejected. Since this is not the C atom alone, this is an example of an intermediate fragmentation pathway. Simply taking this value as the threshold for damage by intermediate fragmentation pathways implies that all events following transfer of energy between 16.88 eV and the primary ejection threshold of 20.16 eV fall into this category. Under such a framework, a small but appreciable contribution of such events to damage at 300 keV is observed, as seen by the dashed purple line in Figure 6.3b and c.

The fragmentation patterns observed during the simulations are more complicated than this view. For T above 16.88 eV but below 20.16 eV, both intermediate fragmentation pathways and non-ejection events occur. Non-ejection events cover a greater proportion of this energy range than intermediate fragmentation pathways, and where ejection does occur it is often over a very small range of T, too small to readily assign an ejection threshold. Thus, dam-

age caused by intermediate fragmentation pathways, although clearly relevant within that transferred energy range, contributes less to amorphisation than implied by the dashed purple lines in Figure 6.3b and c. The 'all pathways' value in Table 6.1 underestimates resilience to some degree by ignoring some pathways by which no permanent damage occurs, while the 'primary pathway' value in Table 6.1 overestimates resilience by neglecting observed intermediate fragmentation pathways. A true computational cross section within the approximations used lies between the 'all pathways' and 'primary pathway' values in Table 6.1. The difference between the two gives a measure of the extent to which intermediate fragmentation pathways affect stability.

The significance of intermediate fragmentation pathways cannot, however, be fully determined by the difference between an all pathways cross section and a primary pathway cross section. It is instead also affected by the relative dominance of ejection and non-ejection events within the intermediate fragmentation pathways range. In the case of HAB, although the difference between the all pathways cross section and the primary pathway cross section is fairly large, the proportion of cases of non-ejection at energies above the lowest ejection threshold is also large: the true computational cross-section is likely to lie closer to the primary pathway cross-section than to the all pathways cross section.

HHB-2 Damage Pathwyas

The relative proximity to the primary pathway cross-section displayed by HAB is not seen for all of the MOFs. For example, Figure 6.4 and Table 6.3 give fragmentation details for HHB-2 equivalent to those given in Figure 6.3 and Table 6.2 for HAB. HHB-2 is the less dense of the two O-containing MOFs, which has the largest difference of the four between its all pathways cross-section and its primary pathway cross-section.

Simulations show that the energy of a 300 keV electron beam is sufficient to cause permanent damage to the HHB-2 flake following impacts on all three

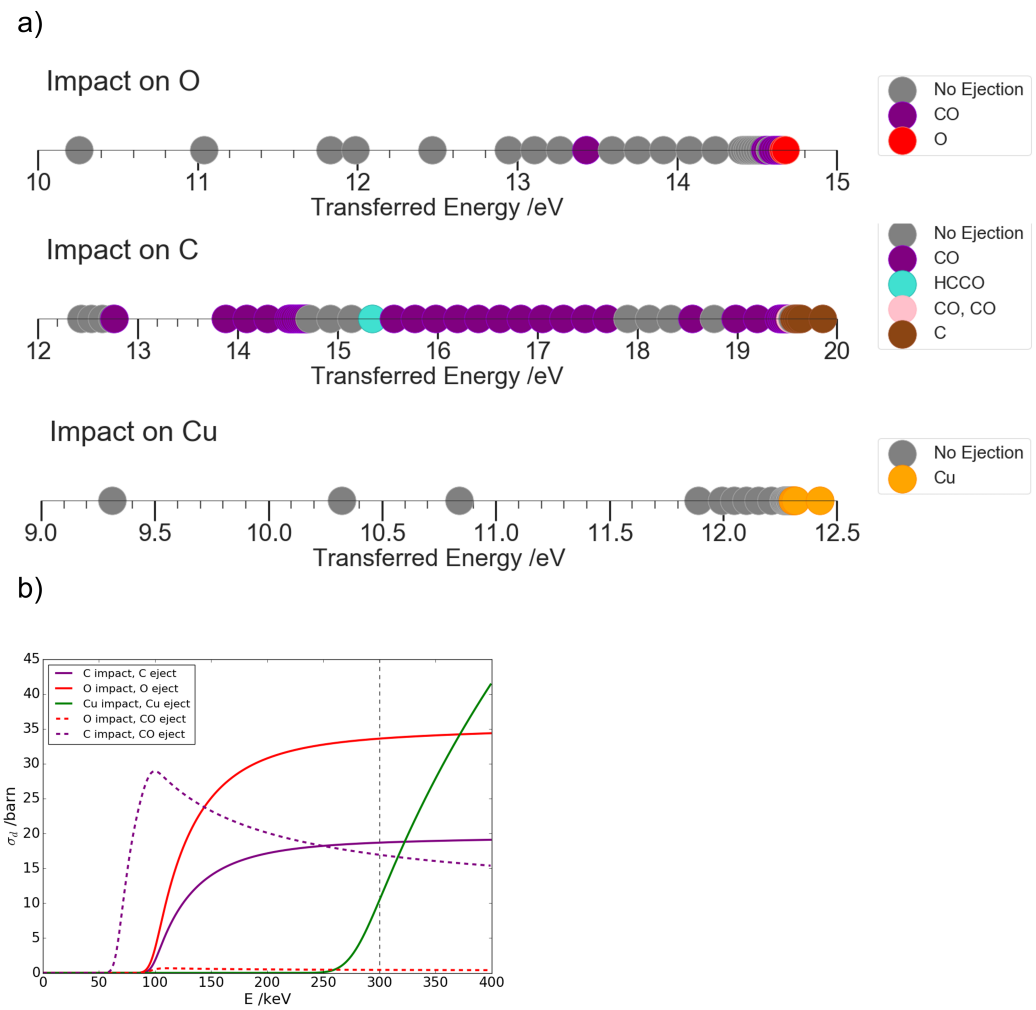

Figure 6.4: *Details of HHB-2 ejection thresholds and cross sections. a) Number lines showing fragmentation events over a range of transferred energies. b) Plots of computational knock-on cross sections following impacts on each atom type split by fragmentation event. Using the numbering system established in Figure 6.2, the ejected atoms and fragments are identified as follows. O: O5, C: C2, CO: C2O5, HCCO: C2C3HO5, CO, CO: C2O5 and C1O4.*

Table 6.3: Ejection thresholds and cross sections for e beam impacts impacts on the atoms in the HHB-2 flake. Using the system established in Figure 6.2, the atoms and fragments referred to are as follows. O: O5; C: C2; CO: C2O5.

PKA	Ejected atom(s)	Transferred energy /eV	Approx. beam energy /keV	σ (300 keV) /barn	Total σ for PKA/barn
O	CO	14.55	96.9	0.4	34
	O	14.66	97.6	34	
C	CO	12.76	65.6	17	36
	C	19.57	97.7	19	
Cu	Cu	12.31	277.8	10	10

atom types it contains. Impact on Cu leads to only non-ejection or the primary ejection pathway, loss of Cu; impact on O leads primarily to either non-ejection or the primary ejection pathway, loss of O, but with some evidence of pathways in which CO can also be lost. For the C impact, as for HAB, the story is different. The primary ejection pathway involves loss of C atoms, but intermediate fragmentation pathways are also significant. The energy range in which intermediate fragmentation pathways are found is dominated not by non-ejection as in HAB but by CO ejection. Intermediate fragmentation pathways not only cover a larger proportion of the energy range than the equivalent pathways in HAB, but the pathways are also better defined, consisting primarily of ejection of one fragment type (CO). In this case, the true cross section from intermediate fragmentation pathways is closer to that implied by the dashed purple line in Figure 6.4b than the equivalent value is to the equivalent lines for HAB in Figure 6.3b and c. Similarly, the true total cross section will be closer to the 'all pathways' cross section in Table 6.1 than the 'primary pathway' cross section.

The significance of intermediate fragmentation pathways also has implications regarding the relative susceptibility of the atoms of HHB-2 to damage at 300 keV. The single largest cause of damage at 300 keV is loss of O by the primary ejection pathway following impact on O, shown by the solid red line in Figure 6.4b. The contribution to damage of intermediate fragmentation pathways following O impacts is, however, negligible, shown by the dashed red line in Figure 6.4b. Non-negligible contributions are observed from damage following impact on Cu (green line), but only from the primary ejection pathway. Significant contributions are observed following impact on C. In this case the primary ejection pathway (solid purple line) and intermediate fragmentation pathways (dashed purple line) have contributions of similar size, which combine to give a larger total cross-section following impact on C than on O.

The contribution of Cu to damage in HHB-2 is an interesting phenomenon that is relevant to imaging of metal-containing structures such as MOFs at high

beam energy. Cu possesses a very high mass in comparison to the other atoms considered, and heavy atoms are generally considered to be less susceptible to damage pathways than lighter atoms. The high mass leads to low transferred energy for a given beam energy (see section 6.1.2) but the bond is weak and the atom easily displaced. This causes no ejection to be seen until a very high beam energy, but once a sufficient beam energy is reached, cross section increases steeply. For HHB-2, the contribution of Cu to damage is still small relative to the other elements at 300 keV, but at slightly higher beam energies, while cross sections of the other elements flatten, loss of copper becomes the dominant damage pathway.

HHB-1 Damage Pathways

Fragmentation pathways for HAB and HHB-2 have thus far been shown to be complex, particularly for impacts on C atoms. For the denser of the two O-containing MOFs, HHB-1, the situation is more complex still. The total cross section for HHB-1 (Table 6.1) is smaller than the total cross section for either HAB or HHB-2, and if only the primary ejection pathway is considered it has the smallest total computational cross section of the four MOFs. A relatively very high proportion of the damage caused to the HHB-1 flake in simulations proceeded via intermediate fragmentation pathways. Figure 6.5 displays number lines (a) and cross section plots (c) for the HHB-1 flake, as well as a cartoon representation of the ejection events observed following C impact, with atoms labelled according to the numbering scheme established in Figure 6.2. Table 6.4 gives the corresponding threshold and cross section values.

Again, impacts on carbon lead to a several different knock-on fragmentation pathways. For C impacts, the range of transferred energies at which fragmentation occurs is large, comparable to the equivalent range for HHB-2. The lowest energy transferred to C at which any fragment is ejected is 13.49 eV, and the primary ejection threshold is 18.64 eV. Several different events

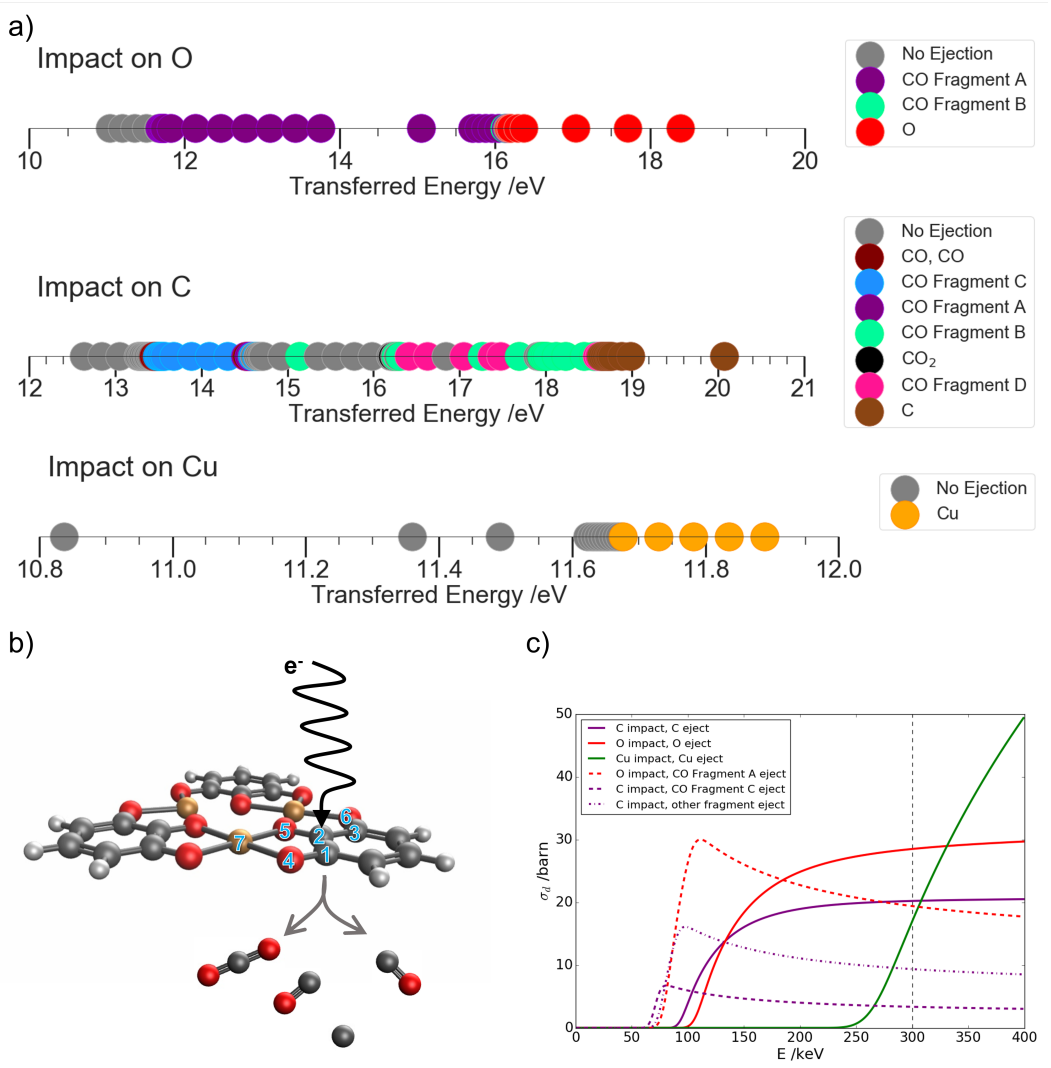

Figure 6.5: *Details of HHB-1 ejection thresholds and cross sections. a) Number lines showing fragmentation events over a range of transferred energies. b) Cartoon representation of ejections following impacts on C, with all atoms involved in ejection events labelled numerically. c) Plot of computational knock-on cross sections following impacts on each atom type, split by fragmentation event. Using the numbering system established in b), the ejected atoms and fragments are identified as follows. O: O5; C: C2; Cu: Cu7; CO fragment A: C2O5; CO fragment B: C1O4; CO fragment C: C3O6; CO fragment D: C2O4; CO_2: O4C2O5. Where two CO fragments eject, they are fragments A and C.*

Table 6.4: Ejection thresholds and cross sections for e beam impacts on the atoms in the HHB-1 flake. Using the system established in Figure 6.2, the atoms and fragments directly referred to are as follows. O: O8; CO (A): C5O8; CO (B): C4O7; CO (C): C6O9; CO (D): C5O7; C: C5; Cu: Cu7.

PKA	Ejected atom(s)	Transferred energy /eV	Approx. beam energy /keV	σ (300 keV) /barn	Total σ for PKA/barn
O	CO (A)	11.69	79.1	19	48
	O	16.18	106.8	29	
C	CO (C)	13.49	69.1	3	33
	CO (other)	14.53	74.1	9	
	C	18.64	93.4	20	
Cu	Cu	11.68	266	17	17

are observed at T larger than 13.49 eV but smaller than 18.64 eV. Where any fragment is ejected within this range it is almost always a CO fragment, but the atoms in the CO fragment do not always originate in the same positions relative to the PKA. The region includes well-defined energy ranges at which a single CO fragment ejects, overlapping energy ranges at which different CO fragments may eject, and significant energy ranges at which non-ejection is observed. With several reasonably well-defined events occurring above the lowest ejection threshold, it is possible to define multiple intermediate ejection thresholds for the individual pathways observed, as shown by the two dashed purple lines in Figure 6.5c. Pathways for which thresholds have been defined include non-ejection, which occurs over a well-defined energy range. All events for which thresholds were defined contributed to the all pathways total cross section, with the non-ejection energy range acting to reduce its value.

Meanwhile, impacts on the metal of HHB-1 engender no intermediate fragmentation pathways. However, the unique damage behaviour of Cu seen for HHB-2 is seen again for HHB-1.

Impacts on the heteroatom of HHB-1 lead to behaviour different from that seen following impacts on the heteroatoms of the other three MOFs. An intermediate fragmentation pathway, ejection of CO, occurs following impact on O for an energy range on a much larger scale than any seen for the heteroatoms of the other MOF flakes. Only one significant intermediate pathway is observed

and the thresholds for the intermediate pathway and the primary pathway are quite well-defined. The intermediate ejection pathway following O impact alone has a very large contribution to damage on the MOF, comparable to the contribution of the primary ejection pathway for impact on C, as shown by the dashed red line in Figure 6.5c.

The large and complex transferred energy range occupied by intermediate fragmentation pathways following impact on C, along with the prominence of CO ejection following impact on O, together translate to the large difference between the 'all pathways' and the 'primary pathway' values for HHB-1 in Table 6.1. The knock-on behaviour of carbon and oxygen in the HHB-1 model illustrates very pertinently the existence of structures for which disordered fragmentation pathways occuring at lower energy than the primary ejection pathway cannot be ignored. The largest contribution to damage in the MOF at 300 keV, as shown in Figure 6.5c, is ejection of O as the primary ejection pathway following O impact. The other two primary ejection pathways are also significant contributors to damage at 300 keV, with Cu ejection contributing more to damage than in HHB-2 or HAB. However, in addition to these primary pathways, there are three distinct intermediate pathways. If combined, these contribute notably more to damage at 300 keV than any of the three individual primary pathways. Overlooking these pathways for HHB-1 would be highly erroneous and suggest much greater stability of the MOF than observed when they are considered.

BHT Damage Pathways

The final MOF, and the most resilient to knock-on damage when all possible fragmentation pathways are considered, is the dense, S-containing BHT. Details of the simulated impacts on the MOF are given in Table 6.5 and Figure 6.6, with number lines illustrating ejection events in Figure 6.6a, a plot of beam energy against cross section of individual events in Figure 6.6b and a cartoon representation of possible events with relevant atoms numbered according to

Table 6.5: Ejection thresholds and cross sections for e beam impacts on the atoms in the BHT flake. Using the system established in Figure 6.2, the atoms referred to are as follows. S: S6; C: C2; Cu: Cu8.

PKA	Ejected atom(s)	Transferred energy /eV	Approx. beam energy /keV	σ (300 keV) /barn	Total σ for PKA/barn
S	S	13.10	164.5	45	45
C	Fragments	18.96	94.4	2	20
	C	19.79	98.7	18	
Cu	Cu	10.27	238.9	37	37

the previously established scheme in Figure 6.6c.

Impacts on the S and the Cu of BHT, following the pattern established for the other MOFs, lead again to only two possible outcomes, which are cleanly separated: no ejection or the primary ejection pathway. Impacts on C may lead to different outcomes, with thresholds again not always clearly defined. As seen before, most prominently for the C impact in HAB, some events are observed only for a very small transferred energy range, too small to be assigned their own ejection threshold. This includes an appreciable energy range at which non-ejection occurs, meaning neither the 'all pathways' nor the 'primary pathway' value in Table 6.1 will exactly reflect a true computational cross section, although the difference between the two values for BHT is small so both values will be close to a true computational value.

The contributions of individual fragmentation pathways and PKAs to total cross section for BHT at 300 keV can give mechanistic insight. The small contribution of intermediate fragmentation pathways to damage in comparison to primary ejection pathways is evident in the dashed purple line in Figure 6.1b. Although the ejected fragments from BHT are normally CS, larger fragments have been observed and the CS fragments do not always contain the impacted C atom, and additionally do not always contain a C and S that were adjacent in the original flake. Comparing PKAs, impacts on the heavier atoms, Cu and S, have the largest contribution to damage at 300 keV. It is telling that neither of these large contributors experiences intermediate fragmentation pathways at all. The significance of Cu at 300 keV is particularly high for BHT compared

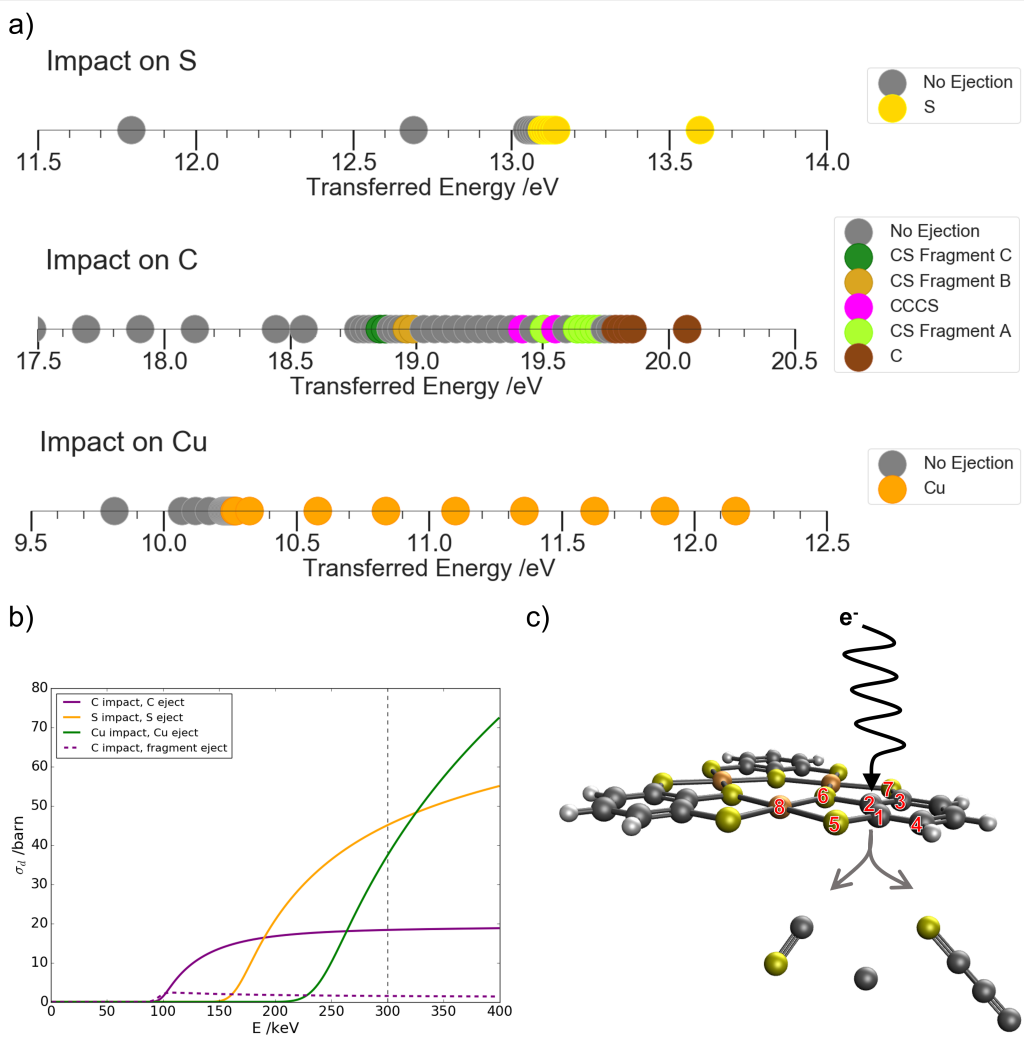

Figure 6.6: *Details of BHT ejection thresholds and cross sections. a) Number lines showing fragmentation events over a range of transferred energies. b) Cartoon representation of ejections following impacts on C, with all atoms involved in ejection events labelled numerically. c) Plot of computational knock-on cross sections following impacts on each atom type, split by fragmentation event. Using the numbering system established in b) and in Figure 6.2, the ejected atoms and fragments are identified as follows. S: S6; C: C2; Cu: Cu8; CCCS: C2C4C1S5; CS fragment A: C1S5; CS fragment B: C2S5; CS fragment C: C3S7.*

to the other three MOF flakes. This is the most extreme case among the four MOFs of the observed unique damage behaviour of metals in complexes.

The beam energy of 300 keV which has here been under focus is a commonly used beam energy in TEM imaging and has previously been a focus for MOFs, [129] but lower beam energies may also be applied, and may themselves be useful tools in reducing beam damage. In general in Figures 6.3a and b, 6.4b, 6.5c and 6.6b, the heavier atoms, Cu and to an extent S possess higher ejection thresholds with cross sections whose gradient is still steep at 300 keV, while the lighter atoms their associated fragmentation pathways possess lower ejection thresholds and less steep cross section gradients at 300 keV. As beam energy reduces, the cross sections of the heavier atoms reduce more quickly than those of the lighter atoms. This means that BHT, whose damage pathways are dominated by Cu and S ejection, reduces more rapidly in total cross section as beam energy reduces than the other three MOFs. At energies close to 100 keV, computationally predicted knock-on damage in BHT is very low. By contrast, several of the fragmentation pathways for lighter atoms in the four MOFs still possess high cross sections at 100 keV, in some cases higher than at 300 keV. This is particularly true of H ejection in HAB, which displays the characteristic rapid increase in cross section at low beam energy. For all four MOFs, intermediate fragmentation pathways are more significant contributors to damage at lower energy that at higher energy. This is even true of BHT: although the total cross section at low beam energy is low for this MOF, a larger fraction of it originates from intermediate fragmentation pathways.

The significance of intermediate fragmentation pathways in knock-on beam damage events has been clearly illustrated. An important consequence of this significance is that discovering the lowest transferred energy at which the primary ejection event occurs should not be assumed to be sufficient to obtain a total cross-section for ejection following impact on that atom. Similarly, the lowest transferred energy at which ejection of any fragment occurs does not always give a full description of ejection cross section. Instead, the various

possible fragmentation events should be assessed in detail for a given system and beam energy. This is the more pertinent for lower beam energies and for impacts on lighter atoms.

6.2.2 Comparison of Cross Sections to Experimental Critical Electron Dose

Discussion up to this point has focused on knock-on damage events which are initiated by transfer of energy from an electron beam to the nucleus of an impacted atom. These events can be described by simulation of energy transfer in the form of velocity to atoms in Born-Oppenheimer dynamics simulations. However, they constitute only a part of the damage that can be done to a material by an electron beam. Damage may additionally be caused by electron-electron interaction and radiolysis, [133, 268–270, 273] which is not captured by the methods applied here. Further, the periodic nature of MOF materials may lead to long-range effects on the nature and extent of damage which are not well-modelled by small flakes.

Experimental observation of beam damage encompasses every kind of effect. Comparison of computational measures of damage susceptibility to equivalent experimental measures can therefore give an indication of the proportion of damage in a material that is caused by short-range knock-on effects, and to what extent the damage has other origins. An entirely direct comparison, however, is elusive for a number of reasons. For one, computational critical doses are commonly calculated per molecule by a weighted sum over constituent atoms, [283] but the equivalent for periodic structures of 2D MOFs is less clear. For another, computational simulations and macroscopic experiments cannot measure exactly the same thing. Experimentally, critical electron dose for amorphisation is measured. Computational simulations, on the other hand, are used to determine cross sections for permanent ejection, which is treated as a precursor to amorphisation but is not exactly the same event. Additionally,

Table 6.6: Computational critical doses for permanent ejection per Cu atom calculated in this work, and experimental critical doses for amorphisation (U. Kaiser; H. Qi, 2022, private communication, 4th July), all at 300 keV for all four MOFs.

	Experimental Critical dose /e Å^{-2}	All pathways Critical dose /e Å^{-2}	Primary pathway Critical dose /e Å^{-2}
HAB	4.45×10^3	2.14×10^5	2.25×10^5
HHB-2	7.80×10^3	3.46×10^5	4.55×10^5
HHB-1	8.88×10^3	5.60×10^5	8.74×10^5
BHT	2.48×10^5	5.98×10^5	6.09×10^5

uncertainties are inherent in experimental imaging of 2D MOFs which cannot be overcome. The thickness of a sample, for example, cannot be controlled with good precision, but can have a significant effect on damage.

Experimental critical electron doses for amorphisation of samples of the four MOFs at 300 keV were provided by the research group of Professor Ute Kaiser, Ulm University (U. Kaiser; H. Qi, 2022, private communication, 4th July). For comparison to these values, critical doses for permanent ejection were obtained as the inverses of the computational per-Cu cross sections as described in section 6.1.2. These computational critical electron doses for ejection are given alongside experimental critical electron doses for amorphisation in Table 6.6. Like Table 6.1, Table 6.6 contains values calculated with consideration of all possible fragmentation pathways and values calcualted with consideration of only the primary ejection pathway.

The total magnitude of the computational critical dose for BHT agrees quite well with experimental values. For the other three MOFs, a qualitative comparsion shows a similar level of agreement: the order of the all pathways computational critical doses is the same as the order of the experimental critical doses. However, the computational critical dose values are two orders of magnitude higher than the experimental ones. This is likely a result of phenomena not captured by the modelling conditions. [268–270] Of particular importance may be the effects of electrical conductivity. It has been experimentally determined that the MOF BHT is far more electrically conductive

than the other three MOFs (U. Kaiser; H. Qi, 2022, private communication, 29th March) and it has previously been shown that electrically conductive materials are less susceptible to damage caused by radiolysis than electrically insulating materials due to the ability of electrons in electrically conductive materials to relax rapidly once excited by the imaging beam. [270] This would appear to point towards radiolsyis causing the dominant portion of damage in the electrically insulating MOFs. However, the contribution of radiolysis to damage in electrically insulating materials reduces with increasing beam energy, so at 300 keV may not in fact be as large as suggested by the comparison of experimental and computational critical doses. Further work is therefore needed to determine the precise origins of this difference and establish to what extent direct quantitative comparison between the computational and experimental cross sections is possible.

6.2.3 Relationship Between Threshold and Cross Section

In addition to examination of specific examples of materials, the propensity of different elements to be involved in knock-on damage pathways may be examined. Each atom in a material possesses some ejection threshold, as discussed in previous sections. This threshold varies depending on the identity of the atom, and also on the environment in which it resides. The atom may also, as has been shown in section 6.2.1, possess different thresholds for various damage events.

Threshold established, the likelihood of damage at a particular beam energy is described by the cross section, which has a complicated relationship with the threshold as described in Equation 6.9 in section 6.1.2. Examining this relationship for a range of ejection thresholds of relevant atoms can give interesting insight into the behaviour of elements under an electron beam and the sensitivity of cross sections to changes in ejection threshold. With that

in mind, a plausible range of hypothetical E_d between 7 eV and 35 eV was selected. For each of the atoms in the four MOFs other than H (C, Cu, N, O, and S), hypothetical cross sections over the E_d range were calculated. Since the E_d here are hypothetical thresholds for some damage event, no separate consideration is made of primary and intermediate fragmentation pathways, nor needs to be. The hypothetical cross sections are plotted against beam energy in Figure 6.7, which also contains a plot of hypothetical ejection thresholds against their corresponding hypothetical cross sections at 300 keV.

Cross sections are larger for smaller damage thresholds and smaller for larger damage thresholds. Variation of cross section with damage threshold over different beam energies depends on the atom in question. The three lighter atoms, C, O and N display similar behaviour. Their cross section has large variation with energy threshold at low beam energies, where the lower thresholds display peaks of cross section and the higher thresholds remain flat. Variation is smaller at higher beam energies, where the peaks have flattened. For the heavier Cu atom, all considered transferred energies correspond to beam energies of 150 keV or higher; at lower beam energies all cross sections are zero and there is no variation at all. As beam energy increases, variation increases and continues to increase over the range of beam energies considered. The S atom, which is of intermediate mass, displays behaviour reflecting aspects of both cases.

The changing behaviour of damage with ejection threshold for all five atoms at one beam energy of 300 keV, is compared in the final plot of Figure 6.7. Where total magnitude of the damage threshold is smaller, the variation displayed by cross section as it changes is larger for all atoms. That is, going from a damage threshold of 7 eV to a damage threshold of 8 eV produces a larger reduction in cross section than going from a damage threshold of 25 eV to a damage threshold of 26 eV. The three lighter atoms display similar behaviour, all reducing in cross section at a similar rate. The cross section of the heaviest atom Cu reduces much faster and is zero for damage thresholds

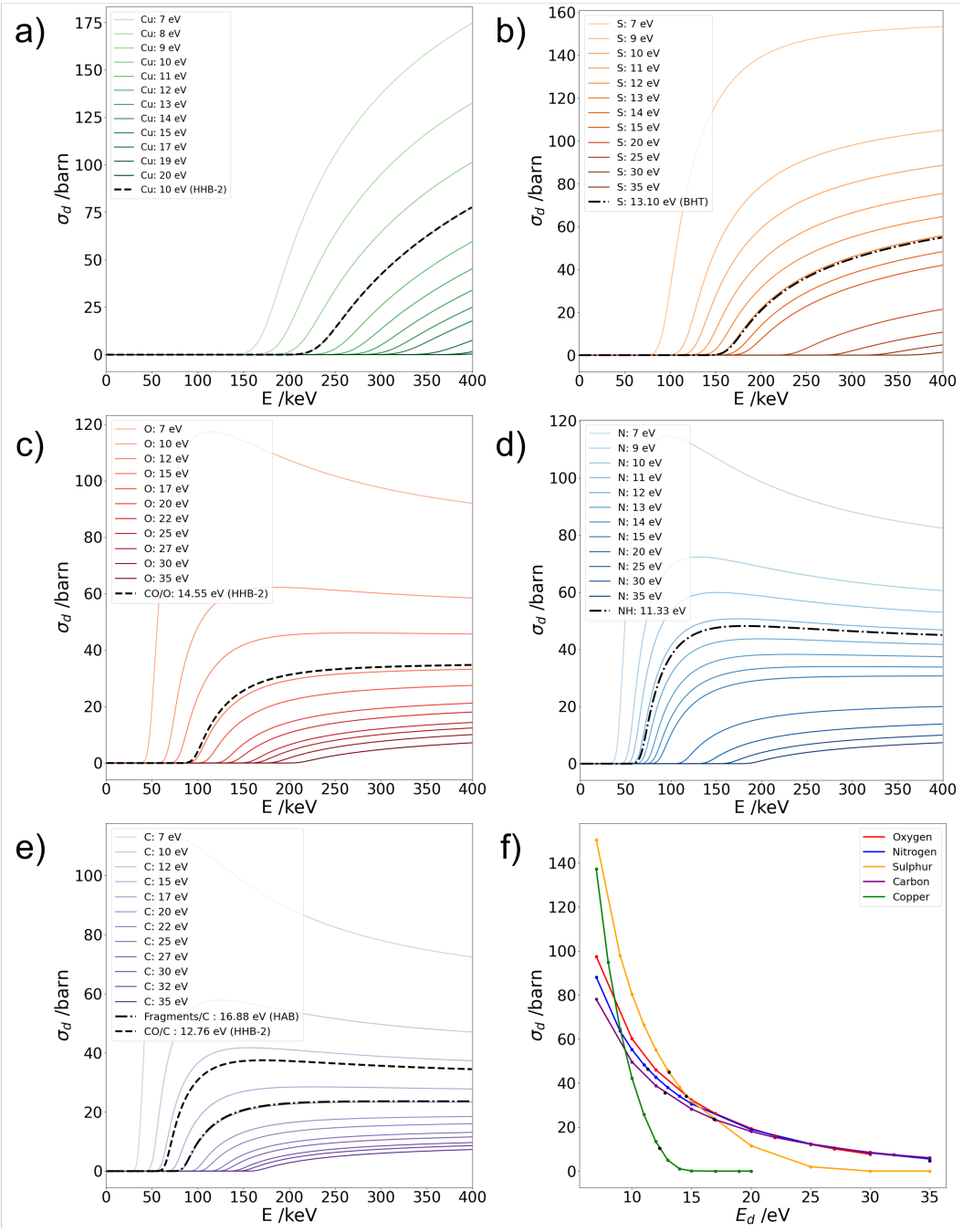

Figure 6.7: *(Plots (a-e) of beam energy against cross section for a selected range of hypothetical ejection thresholds for five elements in the 2D MOFs under consideration, along with some calculated thresholds for context (black dashed lines). Also a plot (f) of ejection threshold against cross section at 300 keV for each of the elements.*

corresponding to beam energies larger than 300 keV. The behaviour of S again falls in between the two cases.

Of the five atoms, the cross sections of Cu and S are the most sensitive to changes in ejection threshold at 300 keV, followed by the lighter atoms in decreasing order of mass. Damage thresholds from 7 eV to 35 eV result in ejection cross sections between 0 barn and 160 barn for Cu and S, and in ejection cross sections between 5 barn and 100 barn for the lighter atoms. While useful for indicating the pattern of behaviour, this range of damage thresholds represents a wide spread the limits of which are unlikely to be reached in practical situations where permanent ejection is the damage under consideration. For context, the smallest and largest ejection thresholds observed in the simulations of the four MOFs are: for Cu, 10.27 eV and above 13.52 eV; for S, 13.10 eV (only one threshold), for O 11.69 eV and 14.55 eV, for N 11.33 eV (one threshold) and for C 12.76 eV and 18.96 eV. These thresholds correspond to a much smaller range of cross sections than the range considered in Figure 6.7 and it is unlikely that a practically accessible change in atomic environment would lead to any permanent ejection threshold for these atoms as low as 7 eV or as high as 35 eV. Among a more practically accessible range of damage thresholds, the cross section range is much lower. Atomic environment, while clearly having an effect on cross section, does not seem to have a drastic one. Interestingly, the accessible range of knock-on ejection cross sections does not appear to be large enough to facilitate critical electron doses on the order of 10^{-3} as observed experimentally for the three electrically insulating MOFs, again indicating that knock-on damage is unlikely to account for all damage in the these structures.

6.3 Conclusion

Ab intitio molecular dynamics simulations have been used to determine transferred energy thresholds for knock-on ejection of the constituent atoms of a

family of four structurally similar 2D MOFs. The McKinley Feshbach equation has been used to determine computational cross-sections for individual atomic and fragment ejection events, and hence total computational critical electron doses of amorphisation for the MOFs.

The importance of different and unexpected fragmentation pathways to knock-on damage has been demonstrated. It has been shown that description of knock-on damage purely in terms of direct ejection of the PKA is insufficient in many cases. Instead, intermediate pathways of fragmentation must be considered which occur below the threshold for clean ejection of the PKA. It has also been shown that total knock-on ejection thresholds calculated based on the lowest energy at which any fragment is ejected following impact on a given PKA are insufficient, due to the possibility of non-ejection events occurring above these thresholds.

It has been indicated that knock-on damage is not a sufficient model to describe ejection patterns in all MOFs under a 300 keV electron beam. In particular, computational critical electron doses calculated using only knock-on damage models display qualitative agreement with experimental values, but the quality of quantitative agreement depends on whether the MOF in question is electrically insulating or electrically conductive, with poor quantitative agreement shown by electrically insulating MOFs. Electrically insulating materials have previously been shown to be particularly susceptible to damage resulting from a radiolysis mechanism in which promoted electrons are unable to relax quickly, leading to the persistence of excited states which are prone to ejection. [270] Previous studies have implied that this phenomenon is prevalent at beam energies lower than 300 keV, but the work performed here suggests that at 300 keV there is some damage mechanism other than knock-on ejection that must be considered.

Consideration has additionally been made of the relationship between beam energies and cross sections for damage events concerning various relevant atoms. Cross sections of damage to heavier atoms are shown to be more

sensitive to changes in ejection threshold than cross sections for damage to lighter atoms. Meanwhile, the maximum possible effect of atomic environment for a given atom type on knock-on cross section is shown not to be large enough for orders of magnitude swings in critical electron dose.

General conclusions may be drawn from this work about prudent practice for imaging MOFs, and in particular 2D MOFs, using TEM. MOFs which contain C and other atoms to which it can strongly bond may be particularly susceptible to complex intermediate fragmentation pathways, and so should be treated with particular care. Meanwhile, the presence of hydrogen can be a significant contributor to damage so should be avoided where possible. This is particularly true at low beam energies, but can still be relevant at higher energy. Selection of appropriate beam energy is also relevant more generally: light atoms display a peak in susceptibility to ejection at some beam energy below 200 keV, but display relatively stable behaviour at higher energies, but imaging metal complexes with a high energy electron beam should only be done with great care, as knock-on damage susceptibility increases rapidly once the ejection threshold is passed. Another particular conclusion is that when imaging electrically insulating MOFs, it is likely that their electrically insulating nature may account for the bulk of the damage that they incur. This may potentially be countered by combining them with some electrically conductive material for the imaging process.

Chapter 7

Imaging and Classical Modelling of Metals Beyond Lennard-Jones

It has been seen in chapter 6 that TEM imaging is an important part of the structural determination of novel materials, MOFs included, as well as of understanding material behaviour. In addition to the TEM imaging of 2D MOFs already seen, the technique can be used to elucidate structures of and interactions within other materials of interest, including other low-dimensional materials. An example is metal nanoclusters, which are famous for their catalytic behaviour. [22] TEM can be used to probe (and promote) cluster growth, defects and the structural changes available to a cluster which are necessary for catalysis. [120] It was additionally seen in chapter 6 that computational modelling can be a vital tool in understanding the processes occurring during an imaging event. In addition to aiding with understanding of beam damage as demonstrated in chapter 6, computational modelling can help to establish the energetic favourability of proposed structures and to increase understanding of structure-property relationships as well as of beam-influenced growth mechanisms and reactions. [284–286] All of this is subject to the use a suitably accurate model. DFT methods were seen to be useful for modelling

imaging events in chapter 6, and useful for modelling interactions relating to material behaviour in chapter 5. However, they have also been seen to be limited, including by cost, and they are difficult to apply to large systems; their application on a very large scale is unfeasible. On top of this, they can pose challenges with the treatment of metals (see Chapter 2).

Classical methods can be judiciously applied on a scale inaccessible to ab initio methods. It was seen in chapters 3, 4 and 5 that standard non-bonded Lennard-Jones parameters can be useful for large scale modelling, in particular of gas sorption events in porous systems. However, standard parameters are not always sufficient for modelling the diverse chemical landscape occupied by MOFs, other commonly imaged systems, and other systems with useful functionality. For one thing, parameters from the UFF [4] and Dreiding [151] force fields that have been used in this work are incredibly general. [287] The UFF, for example, was parameterised for every atom in the periodic table, but contains only limited modifications for atoms in different chemical environments. A consequence of this, which has already been mentioned, is that these standard force fields are not always able to accurately describe interactions involving the metal atoms which can be important to MOF and cluster behaviour. [84, 287]

The setups used in previous chapters for, and commonly applied to, large-scale MOF modelling involve only two kinds of interactions, a Coulomb's law treatment of electrostatics and Lennard-Jones modelling of van der Waals interactions. They are only sufficient for non-bonded interactions and cannot always describe the interactions involving open metal sites, which may be stronger than the interactions for which the force fields were parameterised. Non-bonded parameters lack the ability to account for chemical bonds and flexibility in MOFs, which can be highly relevant to their behaviour. [76] The full, rather than non-bonded, UFF contains terms which account for the stretching, bending, torsion and inversion of a chemical bond, but it is still not widely transferable and cannot account for every type of system relevant to

functional materials and imaging. Metallic bonding of nanoclusters is not well described by the molecular bond. Further, use of the pairwise Lennard-Jones potential does not involve treatment of the many-body nature of interactions. Instead, pairwise interactions are computed and summed, which cannot fully account for the chemical environment of a given atom, a description which is particularly insufficient for metal systems, being unable to reproduce several of their key properties. [140]

Therefore, although the combination of the Lennard-Jones potential and Coulomb's law can be a useful approximation for making predictions, particularly in a high-throughput manner, and ab initio methods can be useful for more detailed calculations on a smaller scale, there is a clear place for other descriptions of functional materials and systems to be imaged. This applies to a range of systems including, as alluded to, both MOFs and atomic solids and clusters. Methods that can be applied on a force field timescale but give a more comprehensive and widely applicable description of interactions than Lennard-Jones are desirable. Various possible force field methods with greater complexity than Lennard-Jones may be pursued, although if fitted to specific systems or types of systems they are not always readily transferable. Machine learned potentials are a popular emerging family of methods which aim for, among other things, transferrability, [114, 288] while there has been extensive development of classical force fields suited specifically for MD of biologcial systems (most notably the AMBER [289] and CHARMM [290] collections of programs), and bonded force fields have been parameterised to describe flexibility [291] and non-standard interactions [287] in MOFs.

In this chapter, the application of classical force fields beyond Lennard-Jones potential is explored. As a result of the lack of transferability of force fields, the focus of exploration is on force fields which can offer a strong treatment of one particular kind of bonding which is not described by pairwise non-bonded terms. The bonding considered is metallic bonding and this is done in the context of modelling metal clusters. These are a class of mate-

rial famous for their catalytic behaviour, [22] with the rare metals platinum and palladium showing particular relevance. They display extensive catalytic properties, being well known for their use in the catalytic converter, [27] and other processes including organic synthesis. [22] The specifics of cluster, rather than bulk, behaviour are of high relevance to catalysis, as metal nanoclusters can have increased catalytic efficiency compared to bulk metals, thanks to increased area in contact with the substrate. Where MOFs, which have been the focus of the majority of this work, are particularly famous for their diversity, nanoclusters display less: a pure metal cluster has no variety available to it in terms of components, and even if an alloy is considered, the number of possible components is increased to a handful, rather than many thousands. Although this reduces their range of possible properties compared to MOFs, it increases the accessibility of their modelling by more complex force fields. It is size and geometry that account for a large amount of the relevant properties of nanoclusters. As a result of their important properties, metal clusters are also common candidates for imaging.

Force field methods which successfully treat metallic bonding can aid in description of clusters as catalytically relevant materials. Focus on these materials and metallic bonding is not directly applicable to MOFs, in which the interactions of metal centres are predominantly not with other metal atoms. However, clusters made up of metal atoms are a relevant part of the context of MOFs for which metal atoms clearly play an important role. Both Lennard-Jones and ab initio methods have been used throughout this work to model the behaviour of the metal centres of MOFs; the addition of further classical force fields to model metal atoms in other environments represents attempts to build a more comprehensive picture of the modelling of metal atoms in general. Meanwhile, hybrid force fields which make use of contributions specific to metal atoms as well as other kinds of bonding to model MOF interactions are not inconceivable.

With the focus on metallic bonding, the category of force-fields pursued

in this chapter is those whose forms have been commonly parameterised for atomic solids, which include group 14 solids, noble gases and most notably metal systems. [140] For convenience they will here be referred to as atomic solid force fields. The definition of this category appears somewhat loose, and indeed force fields with different kinds of functional forms can fit into it. They are generally united, however, by the fact that they routinely make some many-body consideration: by the nature of their bonding, most atomic solids and certainly bulk and cluster metals cannot be treated by only pairwise interactions. [140] There exist a number of force fields which have been developed to treat metals and other atomic solids in bulk form, which can broadly be grouped into those which account for interactions beyond pairwise by some treatment of the total environment of the solid, and those which account for interactions beyond pairwise by some low-order truncation of a many-body expansion. [140] As a widely-employed example of the first kind, embedded atom models (EAMs) [141–143] such as the Sutton-Chen potential, [292, 293] model pairwise interactions explicitly and include a second term for the energy required to embed an atom in the electron density caused by all other atoms (within a cutoff). Challenges are associated with EAMs, such as their lack of treatment of the angular dependence of electron density, [140] but they can be modified to improve accuracy and applicability. Modifications can include introduction of angle dependence and increase of the number of interactions considered explicitly. [143, 294, 295] The many-body Gupta potential is an example comprising a repulsive pairwise and an attractive many-body term, [296, 297] while the Murrell-Mottram (MM) potential is a $2 + 3$ body potential comprised of a pairwise term and a 3-body term. [144]

Parameters of empirical potentials are fitted to reference data, which is usually experimental. Because of the nature of the data available, the fitting of atomic solid force fields is almost always to bulk data, meaning their applicability to other kinds of systems is not guaranteed, although is an important factor to consider. In the publication which introduced the MM potential, for

example, the authors discussed the general motive of finding potentials with validity for surfaces and clusters in addition to bulk, as well as the need for bulk potentials to be tested for their applicability to non-bulk systems. [144] When it comes to metal nanoclusters, systems of a range of sizes with varying proximity to bulk are relevant to catalytic activity. Common sizes of Pt cluster in fuel cells, for example, are around 2-5 nm, [143] and highest catalytic activity is observed at 1-3 nm (10^2-10^3 atoms). [22] Examining the translation of bulk force field performance to clusters of these reduced sizes is a necessary part of the exploration in this chapter. Performance can be validated against experimental data and, for sufficiently small clusters, the results of ab initio modelling.

Atomic solid force fields have previously been used on a number of occasions for metal cluster modelling. Considering the example of the metals Pt and Pd, a study by Ignatov et al [22] parameterised Gupta and Sutton-Chen force fields intended to be suited to cluster modelling by fitting to density functional theory (DFT) cluster data instead of to bulk experimental data. In terms of turning force fields parameterised for the bulk to non-bulk cases, a pertinent example is the fitting of the MM potential for the same metals by Cox. [6] As usual, this fitting made use of bulk data (lattice spacings, cohesive energies, vacancy formation energies, elastic constants, mass and phonon frequencies) with no explicit consideration of non-bulk information. However, extensive consideration of surface as well as bulk information was made in the testing of the fit, [6] and the parameterisation has subsequently been used in cluster studies. [8]

In addition to representing metals outside of the bulk, there are other desirable capabilities for a force field to have in order to enjoy widespread practical use. Specifically, a force field that is not limited to monatomic systems is certainly more useful than one that is, while ability to model processes of interest as well static properties is beneficial. Some atomic solid force fields have been used previously in this context. EAMs, for example, may be used

to model adsorption processes of metals on metals, [298] and from their outset they included parameterisation for H impurities, which were studied in the context of bulk and surface adsorption. [141] They were later expanded to allow treatment of alloy systems, [142, 299] which could then also be studied with H impurities. [299] Alloy treatment involves specifying the relevant parameters for each chemical species and combination of chemical species in the system. [299] The Gupta potential is an example which has seen use in modelling nanoalloy clusters, [297] while it has been speculated that within the MM formalism it may be possible to apply combining rules with some level of generality to elements of alloys in a similar vein to the Lennard-Jones Lorentz-Berthelot mixing rules used frequently in this work. [140] Of particular interest for imaging simulations is the ability to model metal systems as they interact with the support on which they are imaged, which is commonly carbon. Aside from use in imaging simulations, metal-C systems also have applications themselves in catalysis. Jeong and Lee studied these, presenting modified EAM potentials able to model Pt-C and Pd-C systems with MD to further understand the relationship between the metals and the substrate. [143]

Here, the performance of a selected atomic solid potential, the MM potential, for aspects of modelling metal nanoclusters of various sizes, is explored. The two metals Pd and Pt are used for study, with particular focus on Pt. The parametersiaton for both is taken from Cox. [6] The potential has been shown to be able to account for distortions in the bulk [144] and the parameterisation has additionally been used for Pt and Pd surface distortions, [6] and applied to clusters. [8] All three cases, particularly cluster modelling, are further addressed here. The MM potential is applied to reference systems using FORTRAN code developed by Tailor et al, [7] and modified as part of this work where relevant for desired functionality. The performance of the code and necessary modifications are validated by comparison to published data and the performance of the potential for clusters of a range of sizes is validated against

literature ab initio results. The potential is then used to explore binding energy and stability of a series of clusters relevant to experimental observations of Pt cluster growth. This gives insight both into the capabilities of the potential and the properties of Pt clusters. The force field exploration examines how size and geometry of a cluster can dictate its behaviour and stability.

7.1 Methods

Classical modelling in this chapter makes use of the Murrell-Mottram (MM) potential. [144] Relevant MM parameters for both Pt and Pd are taken from the fitting of Cox, [6] and the code used to apply the potential was developed by Tailor et al. [7] Both code and potential are validated against literature data, including both experimental data [6] and ab initio data from a range of sources. [9, 22, 300–304]

7.1.1 The Murrell-Mottram Potential

The Murrell-Mottram potential is a semiempirical 2 + 3 body potential. It is based on a many-body expansion of potential energy, truncated after the three-body term (Equation 7.1).

$$V = \sum_i \sum_{j>i} V_{ij}^{(2)} + \sum_i \sum_{j>i} \sum_{k>j} V_{ijk}^{(3)} \tag{7.1}$$

Here, $V_{ij}^{(2)}$ is the two-body term and $V_{ijk}^{(3)}$ is the three-body term. Definitions for both the two-body and the three-body terms follow, beginning with the two-body term which is defined in equation 7.2.

$$V_{ij}^{(2)} = -D(1 + a_2\rho_{ij}) \exp(-a_2\rho_{ij}) \tag{7.2}$$

Here, D, which represents the dissociation energy of the diatomic, [144] and a_2 are MM parameters to be determined during fitting, and ρ_{ij} is given by equation 7.3 where r_e, the diatomic equilibrium distance, [144] is another

parameter to be fitted and r_{ij} is the distance between atoms i and j.

$$\rho_{ij} = \frac{r_{ij} - r_e}{r_e} \tag{7.3}$$

The three-body term is expressed as a polynomial in symmetry coordinates Q_1, Q_2 and Q_3, defined in terms of ρ_{ij}. This definition is given in Equation 7.4.

$$\begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \end{pmatrix} = \begin{pmatrix} \sqrt{1/3} & \sqrt{1/3} & \sqrt{1/3} \\ 0 & \sqrt{1/2} & -\sqrt{1/2} \\ \sqrt{2/3} & -\sqrt{1/6} & -\sqrt{1/6} \end{pmatrix} \begin{pmatrix} \rho_{ij} \\ \rho_{jk} \\ \rho_{ki} \end{pmatrix} \tag{7.4}$$

The three-body term uses these symmetry coordinates as in Equation 7.5.

$$V_{ijk}^{(3)} = D(c_0 + c_1 Q_1^2 + c_2 Q_1^2 + c_3(Q_2^2 + Q_3^2) + c_4 Q_1^3 + c_5 Q_1(Q_2^2 + Q_3^2) + c_6(Q_3^3 - 3Q_3 Q_2^2))\,\mathrm{sech}(a_3 Q_1) \tag{7.5}$$

Again, D (diatomic dissociation energy) is specific to the material in question, along with the parameters a_3 and c_{1-6}. It is possible to extend the polynomial to a higher degree, which would use more c_i parameters, but the six displayed are those used in the Cox fittings of Pt and Pd which are applied here. [6] Thus, the MM parameters are D, r_e, a_2, a_3 and c_{1-n}. D and r_e are scaled to the cohesive energy and the lattice constant of the reference solid. The non-polynomial factor in Equation 7.5 ($\mathrm{sech}(a_3 Q_1)$) is a damping function which sends the three-body term to zero at the dissociation limit. It is possible for this damping function to take forms other than sech, [7] such as exp and tanh, but Cox selected sech as a superior function. The parameters for Pt and Pd derived from Cox, which are used throughout this chapter, are given in Table 7.1.

The cutoff within which interactions between atoms are considered is also

Table 7.1: MM parameters derived by Cox [6] for Pt and Pd, which are used here throughout this chapter.

Coefficient	Pt	Pd
a_2	8.5	7
a_3	9	10.2
D /eV	1.613	0.946
r_e	2.699	2.667
c_0	0.244	0.197
c_1	-0.429	-0.221
c_2	5.814	6.516
c_3	-2.581	-0.435
c_4	1.268	10.273
c_5	-7.386	-14.543
c_6	5.401	4.463

relevant as it determines the spatial range covered by the potential, which has implications for both accuracy and cost. The cutoff used by Cox is $2.25 \times r_{nn}$ where r_{nn} is the nearest-neighbour distance. This leads to inclusion of five shells.

7.1.2 Validation of the Code and Tuning of the Setup

The code of Tailor et al [7] has functionality for energy minimisation and vibrational spectroscopy simulation using a single user-specified set of MM parameters. Prior to using the code to explore the MM potential, its performance for the relevant systems was verified against known previous MM results. The Tailor code [7] and the Cox parameters [6] were used to reproduce computational bulk and surface Pt data from Cox [6] and Pt cluster data from a theoretical study by Lloyd and Johnston, [8] which used the same potential. The code was written using the exp damping function for the three-body term (equation 7.5), so it was necessary to replace this with sech and its derivative. Any geometry optimisations used a convergence threshold of 1×10^6 eV Å^{-1}.

In line with the purpose of this exploration of the MM potential, quantities relevant to energetic interactions between particles are considered. From the

work of Cox, [6] the relevant bulk quantity is cohesive energy, E_{coh}. This is the difference between the energy per atom of the cluster and the energy of a free atom; for these purposes, energy is a function of only interactions between atoms so E_{coh} is simply the energy per atom of the cluster. In the MM potential the scaling parameter D ensures that E_{coh} is reproduced. For the purposes of the verification, bulk Pt is considered, for which the experimental E_{coh} to which Cox fitted was 5.84 eV.

Reproduction of this value by the code of Tailor et al [7] was tested. The code enforces cutoff differently to the implementation used by Cox. [6] Some finite system or periodic unit cell is defined and interactions are computed for every atom in the system with all other atoms, or their closest periodic image. A cutoff exactly equivalent to the one implemented by Cox [6] cannot be defined within this setup, so a series of cutoffs defined within the system used were tested against the reference E_{coh}. Pt unit cells of varying sizes were constructed and periodic boundary conditions were applied to the unit cells to calculate unrelaxed energy and relaxation energy (E_r), optimise the geometry and calculate relaxed energy of the total system. From here, unrelaxed and relaxed E_{coh} were calculated. Energy values for the unit cell sizes are given in Table 7.2. Note that the number of atoms N is not equal to the product of the number of layers in the x, y and z directions because of the nature of the Pt unit cell, which is face centred cubic: there are not the same number of atoms in each layer. Note also that a number of the relaxation energies are negative, in cases where relaxed energy is less negative than unrelaxed energy. Negative relaxation energy should not be seen, but this is down to the precision of the method. The starting structure was generated to represent the bulk solid, and was already very close to a relaxed structure. The difference is very small and within the available precision a negative relaxation energy is seen.

Using a very small unit cell of $4 \times 4 \times 4$ layers, E_{coh} is not well reproduced. This unit cell is significantly smaller than the cutoff distance used to fit the potential. [6] Interactions occurring at separations larger than the unit cell size

Table 7.2: Unrelaxed and relaxed total and cohesive energies in eV for a range of Pt unit cell sizes, calculated in this work using the MM potential with the fitting of Cox [6] and the code of Tailor et al. [7] Rel. = relaxed, unrel. = unrelaxed

Unit cell layers ($x \times y \times z$)	N	E_{tot} (unrel.)	E_{coh} (unrel.)	E_{tot} (rel.)	E_{coh} (rel.)	E_r /%
$4 \times 4 \times 4$	32	-178.05	5.5641	-177.01	5.5316	-0.58521
$6 \times 6 \times 10$	180	-1047.6	5.8201	-1047.7	5.8204	0.00601
$8 \times 8 \times 8$	256	-1494.3	5.8370	-1494.1	5.8364	-0.00891
$14 \times 14 \times 12$	1176	-6864.5	5.8371	-6863.7	5.8365	-0.01124

but smaller than the original cutoff distance are not be correctly accounted for. The reproduction of E_{coh} shows a marked increase with unit cell size and E_{coh} is reproduced quite faithfully with a cell of $8 \times 8 \times 8$ or larger.

Testing pure bulk quantities is indispensable for a robust verification, but it is necessary to depart from the bulk to explore modelling of the full range of metal behaviour and to approach cluster properties which are of particular relevance to catalysis and imaging. A useful step away from bulk is consideration of surface energies, E_S, which measure the excess energy of atoms in surface environments compared to bulk. Their treatment, including treatment of surface reconstructions, by the MM potential has been accurate in the work of Cox which introduced the Pt and Pd fittings used here. [6] The Cox surface energies are here used in tests of the function of the code outside the bulk. Cox presented E_S for the (111), (100), and (110) surfaces of Pt, as well as for two Pt surface reconstructions which are observed experimentally for a handful of metals, Pt included. As examples, the implementation of Tailor et al [7] is used to calculate surface energies for the (100) and (111) Pt surfaces for comparison to the values given by Cox. [6] Surface energy is modelled using a slab which is periodic in the x and y directions, and has two identical z surfaces of the appropriate type. It is calculated using Equation 7.6.

$$E_S = \frac{NE_{coh} - V}{2A} \tag{7.6}$$

Table 7.3: Unrelaxed and relaxed total (eV) and (100) surface (meVÅ^{-2}) energies for two unit cell sizes, calculated using the MM potential.

Unit cell layers ($x \times y \times z$)	N	E_{tot} (unrelaxed)	E_S (unrelaxed)	E_{tot} (relaxed)	E_S (relaxed)
$8 \times 8 \times 8$	256	-1419.2	154.80	-1420.6	151.90
$14 \times 14 \times 12$	1176	-6634.4	153.16	-6637.0	151.40
$14 \times 14 \times 11$ (Cox)[6]	1008				152.28

Table 7.4: Unrelaxed and relaxed total (eV) and (111) surface (meVÅ^{-2}) energies, calculated using the MM potential.

Unit cell layers ($x \times y \times z$)	N	E_{tot} (unrelaxed)	E_S (unrelaxed)	E_{tot} (relaxed)	E_S (relaxed)
$6 \times 8 \times 9$	864	-4803.2	187.95	-4875.4	131.30
$7 \times 8 \times 9$ (Cox)[6]	1078				131.90

In equation 7.6, E_{coh} is the bulk cohesive energy, N is the number of atoms in the slab, V is the total energy of the slab and A is the surface area of one z face. The factor of two accounts for the fact that two faces composed of surface atoms are modelled.

To calculate surface energy, Cox used 11 layers with 7×7 atoms per layer for the (100) surface and 9 layers with 7×8 atoms per layer for the (111) surface. Double-sided slabs containing $8 \times 8 \times 8$ layers and $14 \times 14 \times 12$ layers were used here for the (100) surface, and $6 \times 8 \times 9$ layers were used for the (111) surface. The E_{coh} from the $8 \times 8 \times 8$ bulk case was used in equation 7.6. Unrelaxed and relaxed total and surface energies for the (100) and (111) surfaces are given in Tables 7.3 and 7.4 respectively. For both surfaces, E_S is well reproduced. The use of the larger slab for the (111) surface makes very little difference to the calculated surface energy.

With surface and bulk properties examined, modelling of finite clusters can be considered. The ability to model clusters of various sizes is pertinent to the range of relevant materials. Here, the setup implemented in the code of Tailor

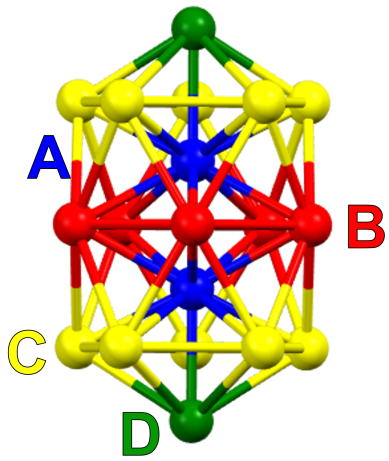

Figure 7.1: *Visualisation of the Pt DI modelled both in this chapter and in the work of Lloyd and Johnston. [8] Atoms are coloured according to symmetry equivalence. Blue: A; red: B; yellow: C; green: D.*

et al [7] was tested for reproduction of cluster data from the work of Lloyd and Johnston [8] in which calculations were carried out on small Pt clusters at the few-atom level. The clusters used were a Pt_{19} double icosahedron (DI) and the Pt_{18} and Pt_{17} incomplete DIs (IDIs) resulting from removal of one or two atoms from it. The Pt_{19} and Pt_{18} calculations are reproduced here. The DI structure is shown in Figure 7.1, with atoms coloured according to symmetry equivalence (i.e. an atom is the same colour as any other atoms which are symmetrically equivalent to it). There are four groups of symmetry-equivalent atoms, labelled A-D in Figure 7.1 for reference.

Since there are four symmetry-equivalent groups, four different 18-atom structures can be made by removal of one atom A-D from the DI. These were all considered by Lloyd and Johnston [8] and are in turn considered here. The DI structure was built and optimised using the MM potential implemented in the code of Tailor et al, [7] then for each group of atoms A-D in turn, one atom was removed and the structure optimised again. In each case, unrelaxed binding energy, E_b^u, relaxed binding energy, E_b^r and relaxation energy, ΔE_b^r were calculated. These quantities are defined in Equations 7.7 and 7.8 in which V_{clus} is the total energy of the cluster and N is the number of atoms it

Table 7.5: E_b^u, E_b^r and ΔE_b^r (eV) calculated by Lloyd and Johnston [8] for the Pt_{18} IDIs and equivalent values calculated here for the Pt_{18} IDIs and for the Pt_{19} DI.

		Pt_{19} DI	Pt_{18} DI−A	Pt_{18} DI−B	Pt_{18} DI−C	Pt_{18} DI−D
	Unrelaxed		-64.9206	-65.2536	-66.3012	-66.357
	Relaxed		-67.3092	-65.403	-66.9114	-66.5838
Ref [8]	N	19	18	18	18	18
	E_b^u		3.6067	3.6252	3.6834	3.6865
	E_b^r		3.7394	3.6335	3.7173	3.6991
	ΔE_b^r		0.1327	0.0083	0.0339	0.0126
	Unrelaxed	-70.949	-64.8995	-65.2283	-66.2737	-66.3313
	Relaxed	-71.5605	-67.2902	-65.3787	-66.8915	-66.5568
This	N	19	18	18	18	18
work	E_b^u	3.7342	3.6055	3.6238	3.6819	3.6851
	E_b^r	3.7663	3.7383	3.6321	3.7162	3.6976
	ΔE_b^r	0.0322	0.1328	0.0084	0.0343	0.0125

contains. Binding energy is equivalent to the cohesive energy used by Cox. [6]

$$E_b = \frac{-V_{clus}}{N} \tag{7.7}$$

$$\Delta E_b^r = E_b^r - E_b^u \tag{7.8}$$

The E_b^u, E_b^r and ΔE_b^r values calculated by Lloyd and Johnston [8] are displayed in Table 7.5, along with values for equivalent quantities calculated here for verification. The calculated values are very close to those given by Lloyd and Johnston, [8] and each quantity follows the same order among the different IDIs. Visually, all optimised structures display distortion from the starting geometries of the optimised DI with one atom removed. For each structure, the distortion observed following the optimisation applied here is very close to that observed following the optimisation of Lloyd and Johnston. [8] This is seen in images of the optimised structures, which are are given in Figure 7.2 and can be compared to equivalent images presented in reference [8].

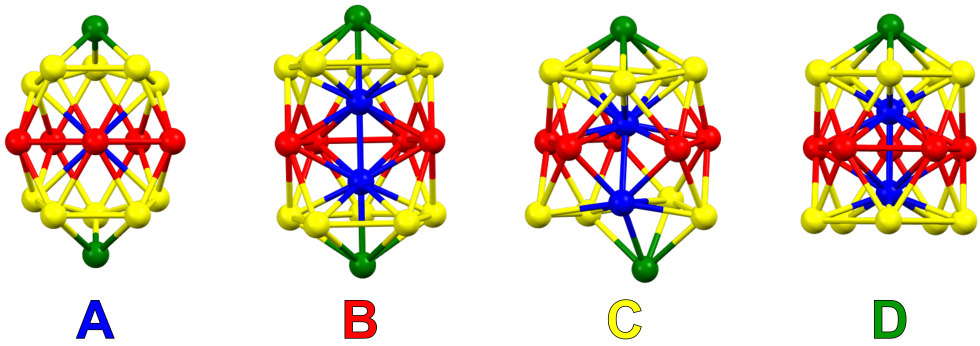

Figure 7.2: *Visualisation of the optimised IDIs resulting from application of the optimisation function in the code of Tailor et al [7]. Structures are labelled A, B, C and D according to which atom was removed prior to optimisation.*

7.1.3 Validation of the Murrell Mottram Potential Parameters

With the confirmation that the code of Tailor et al [7] functions as expected, and the establishment of a suitable setup with appropriate cutoffs and damping function, it must also be ascertained that the potential satisfactorily models cluster behaviour. The Pt fit of Cox was validated against experimental bulk and surface data available at the time (1998). [305] The experimental surface energy used was determined at the melting point of the element in question using the surface tension of the liquid. It was extrapolated to 0 K to obtain the experimental value used by Cox, 155.40 meV $Å^{-2}$. [6, 305] Due to experimental limitations, this is an isotropic energy over all surfaces and not specific to individual surfaces.

Experimental limitations persist: it remains the case today that no experimental data could be found for individual Pt surfaces, although the most up to date experimental average value over all surfaces is now 147.9 meV $Å^{-2}$. [306] There also exist a number of computational studies which have used DFT methods to obtain surface energies for individual surfaces, although the difficulty in modelling metals using DFT results in surface energies heavily dependent on computational setup, so published values occupy a large range. [9] Some of the most notable studies of this nature are databases of multiple

Table 7.6: Computational literature surface energies (meV $Å^{-2}$) for Pt surfaces. The row entitled range refers to the range of literature values identified by Tran et al. [9]

	(111)	(100)	(110)
Tran [9]	92.8	115	105
Range [9]	91.8 [301]-147[304]	139 [303]-170 [300]	142 - 182[302]
Skriver [304]	147	155	-
Vitos [300]	144	171	176
Galanakis [302]	144	165	182
Da Silva [301]	91.8-140	-	-
Yoo [303]	113	139	142
Cox[6]	132	152	162
This work	131	152	-

surface energies for tens of elements presented by two sets of authors, Vitos et al [300] in 1998 followed by the more comprehensive database of Tran et al [9] in 2016. Both authors calculated surface energy using DFT calculations and made comparison to other published computational surface energies where available. The work of Tran et al [9] includes an extensive literature search which illustrates the range of published computational values for surfaces of several elemental solids. Surface energy values are given in Table 7.6 to illustrate the range. The surface energies calculated by Tran for the (111), (100) and (110) surfaces are included, [9] along with the literature surface energy ranges found in the same work and values for each surface from the publications which gave the extremes of these ranges, including that of Vitos et al, [300] in chronological order. The row referring to the work of Da Silva et al [301] contains a range of values for the (111) surface because the work presented multiple values calculated by different methods. The reference for the lower bound energy of the (110) surface recorded by Tran et al [9] is omitted, as the work referenced Vitos et al [300] for the value and it is not found in the Vitos et al study.

MM surface energy values calculated by Cox [6] and in this work are also shown in Table 7.6 for comparison. The Cox values, which are accurately

reproduced here, fall well within the ranges of ab initio values presented and appear to increase in the same order as the ab initio values. Although the wide range of ab initio values highlights their variation with selected method and the difficulty of using them as benchmarks, it also shows the usefulness of the classical MM potential. The potential is cheaper than ab initio methods and in many cases one ab initio surface energy varies from another more than from the MM surface energy. The specific setup of an ab initio method may introduce more error than use of this classical force field. The values presented by Cox and in this work also match well with the isotropic experimental value.

Strong performance for surface energy is a fundamentally important part of a successful potential, and is also promising for its ability to calculate cluster data. In establishing the method's performance for a range of metal systems, it is also useful to verify cluster performance directly, for which both Pt and Pd were considered, again using the potential parameters fitted by Cox. [6] A large selection of Pt clusters from Pt_6 to Pt_{148}, and Pd clusters from Pd_5 to Pd_{309} were examined. Each selection of structures underwent geometry optimisation using the MM Pt and Pd parameters fitted by Cox, and the resulting energies are compared to literature cluster data. For Pt, literature data was taken from the work of Ignatov et al [22] in which a modified Sutton-Chen potential was fitted to DFT data and applied in a global minimum search for Pt nuclearities up to 150. For Pd, the literature data was taken from a study by Nava et al [23] which made direct DFT calculations.

No attempt at a global minimum search was made here for either Pt or Pd. Instead, optimised structures from the two literature studies [22, 23] were taken as starting geometries, a single optimisation carried out for each, and binding energy calculated. This lacks the robustness of a global optimisation, but has a lower cost, and since the literature structures are taken as initial geometries, the local minima found are expected to be close to them. In Figure 7.3, relative binding energies are plotted for Pt and Pd using both the MM potential and the literature studies. [22, 23] In each case, the structure with

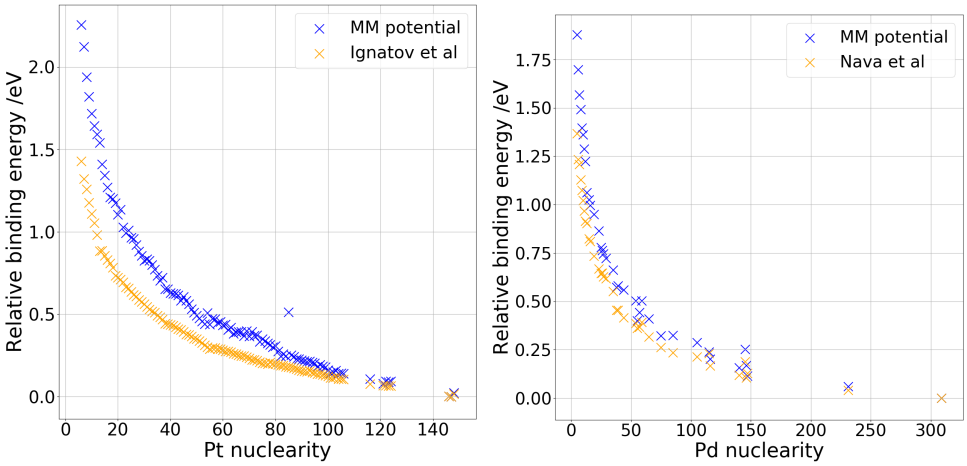

Figure 7.3: *Plots of nuclearity against relative binding energy for Pt and Pd according to the MM potential, with equivalent literature data for comparison. Left: Pt, literature data from Ignatov et al[22]; right: Pd, literature data from Nava et al.[23]*

the lowest binding energy is fixed at zero. Relative binding energies calculated using the MM potential follow the same trend as the literature data with little exception, and are of a similar magnitude. Agreement becomes stronger as cluster size increases.

Since no global minimum search was undertaken and optimised structures are not examined in detail, there is no guarantee that the same structures are found once the Sutton-Chen and DFT minima are optimised using the MM potential. However, there is a clear general agreement in energy trends according to nuclearity, suggesting that the MM potential has a comparative performance to other computational methods.

7.2 Exploration of Model Clusters Using the MM Potential

With the performance of the setup established in relation to literature computational and experimental data, the MM potential was used to study interactions in a system representing an experimentally observed cluster of some hundreds of atoms which has not yet been studied computationally. A system of this size is expensive to model with ab initio calculations, so this is an

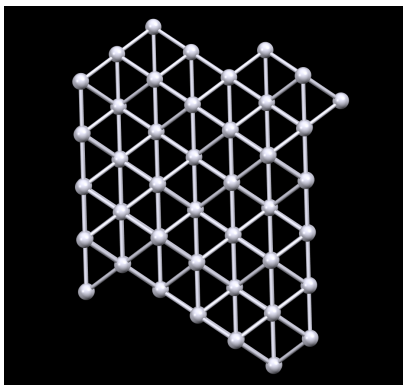

Figure 7.4: *Visualisation of the Pt cluster studied in this chapter after 609s of e-beam irradiation.*

example of a case where a model with classical computational cost is beneficial.

The cluster selected (visualised in Figure 7.4) was grown on a carbon support under the influence of an 80 keV electron beam and imaged using transmission electron microscopy (TEM), with an atom counting method [307] applied to estimate its structure. The particular configuration was taken from a single frame of the growth process, 609 s after initialisation of radiation and was estimated to contain 199 atoms. (U. Kaiser; K. Cao, 2020, private communication, 6th October). Local geometry resembles bulk Pt, but the external shape of the cluster is not regular. Atom counting was affected by the tilt of the structure; the number of atoms in each column and their relative vertical positions on the C support were unable to be determined exactly by experiment. By modelling the selected cluster and possible deviations from the experimental estimation, the MM potential can be used in an experimentally relevant environment that takes into account the kinetic disorder present under the circumstances of real cluster growth.

MM calculations were carried out for an approximated structure of the cluster to probe two different conditions relating to the cluster growth process. First, the lability of surface atoms was examined by energy calculations of various perturbations to the surface of the cluster to give indication of the likelihood of structural variation during growth. Second, the arresting effect of the carbon support was probed by geometry optimisations in which the

positions of atoms in layers close to the support in the real structure were constrained.

7.2.1 Surface Atom Migration

The Pt cluster was probed by calculation of interaction energy following atom migrations. The structure of the cluster in the x and y directions is readily available from the experimental image. The number of atoms in each of the 42 columns in the z direction is less clear, but has been estimated along with the positions of atoms in each column using image contrast. (U. Kaiser; K. Cao, 2020, private communication, 6th October) The Pt columns are represented in Figure 7.5 (left), each with a label indicating the estimated number of atoms that the column contains. The structure with these recorded column populations is hereafter referred to as the original cluster. It is subjected to well-defined surface perturbations and the resulting perturbed clusters are analysed and compared. This allows assessment of the sensitivity of the MM potential to small changes in geometry as well as assessment of how the energy of the estimated structure compares to other similar clusters. It was convenient to assign labels to the columns to keep track of perturbations, so Figure 7.5 (right) displays a representation of the same structure with index labels assigned to each stack.

The perturbations applied to the cluster are translations of the uppermost atoms in a given column (highest z coordinate). A series of perturbations are separately applied, with two migrations individually made from each stack to result in 84 new perturbed clusters. The migrations are of the top atom of each stack to each of the two stacks with adjacent indices in turn. That is, for n from 0 to 41 the top atom of stack n migrates to stack $n-1$ and, separately, to stack $n+1$. The indices adjacent to stack 0 are taken to be 1 and 41.

The migrated atom must be assigned coordinates in its destination stack. Its x and y coordinates are simply the x and y coordinates of the other atoms in

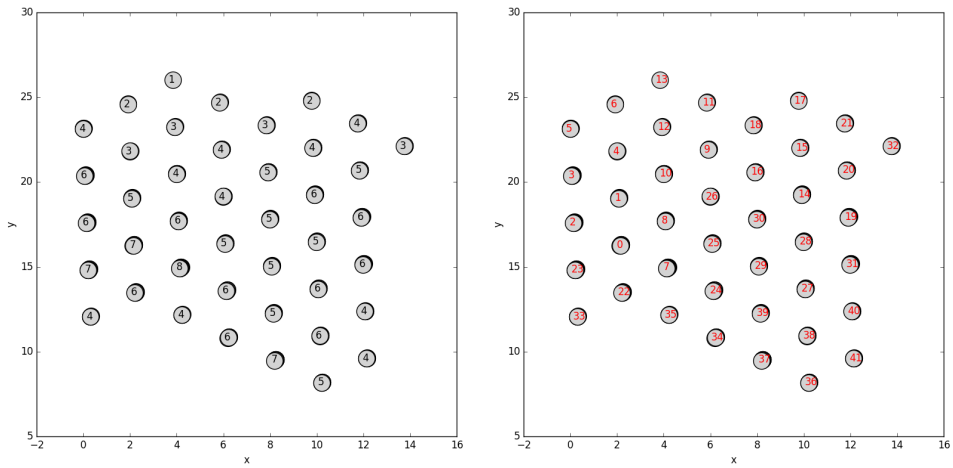

Figure 7.5: *Representations of the structure of the Pt cluster after 609s of irradiation. Left: numbers are amount of Pt atoms in each stack (black text); right: numbers are indices assigned to stacks for convenience (red text).*

the destination stack. In cases where the destination stack contains more than one atom, the destination z-coordinate is the sum of the z-coordinate of the previous top atom in the destination stack and the average z-separation in that stack. One stack in the original cluster contains only one atom, and in the case of this being the destination stack, the destination z-coordinate was assigned as the sum of the z-coordinate of the atom in the stack and 2.774 Å, the average z-separation for the whole cluster. With a cluster of 42 atoms, it would have been possible to apply several other migrations: one separate migration from every stack to every other stack would amount to $42 \times 41 = 1,722$ total migrations, and migrations of multiple atoms at once could also be considered, resulting in several more possibilities. The 82 migrations applied here were chosen as a convenient way to consider two moves from each stack without imposing conditions on the moves, and resulted in a varied distribution of migrations.

In order to make comparison to the original cluster, its energy must be known. Using the MM potential, the initial energy of the original cluster was calculated as -943.394 eV; the energy was minimised and the final potential energy calculated as -966.120 eV. The unrelaxed and relaxed structures are shown in Figure 7.6. Note that the relaxed cluster has been distorted in a way that may not be feasible for the cluster grown on a carbon support as in

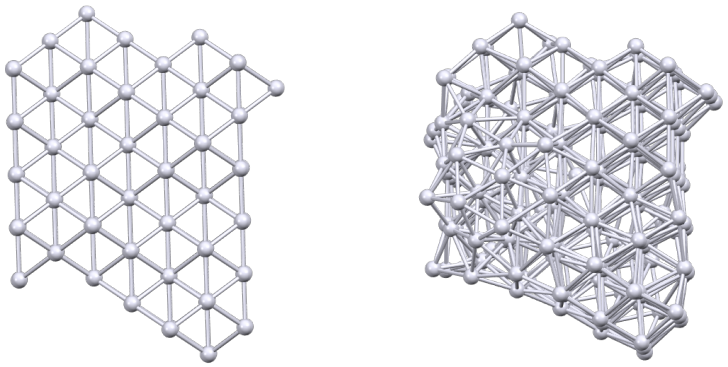

Figure 7.6: *Visualisations of the original cluster obtained from experiment. Left: unrelaxed, MM energy = -943.394 eV; right: relaxed using the MM potential, MM energy = -966.120 eV.*

experiment, although the shape is recognisable as the grown cluster.

The 84 perturbed structures were also subjected to MM geometry optimisation. The highest and lowest value of relaxed, unrelaxed and relaxation energy observed among the 84 perturbed structures are compared with equivalent values for the original cluster in Table 7.7 (all three energy values are given for each structure). Table 7.7 also displays are the percentiles of the unrelaxed, relaxaed and relaxation energy values of the original cluster. That is, the percentage of the data set whose energy lies below that of the original cluster (eg. the 10^{th} percentile is the number below which 10% of the data falls). The comparison of the unrelaxed, relaxed and relaxation energies of the orignal cluster with the distributions of these values among all of the perturbed structures are illustraed in histograms in in Figure 7.7

The unrelaxed and relaxation energies of the orignal cluster are low in comparison to the rest of the data, with only 8.33% of unrelaxed structures possessing a lower energy, and 17.86% of the structures having a lower relaxation energy. The relaxed energy of the original cluster is higher by comparison with the perturbed structures, with 64.29% of perturbed structures possessing lower relaxed energies. During real cluster growth, atom positions display closer resemblance to those of the unrelaxed than the relaxed structures. Under these unrelaxed circumstances, the geometry of the original structure (as

Table 7.7: Unrelaxed, relaxed and relaxation energy (eV) of the original cluster, and of the clusters possessing the highest (**bold**) and lowest (*italics*) unrelaxed, relaxed and relaxation energies in the set of 84 migrations. All energy calculations and relaxations are performed using the MM potential.

	Unrelaxed energy /eV	Relaxed energy /eV	Relaxation energy /eV
3 to 4	*-945.396*	-957.504	*12.109*
Lowest unrelaxed	*(lowest)*		*(lowest)*
4 to 3	**-936.819**	-971.416	34.597
Highest unrelaxed	**(highest)**		
16 to 17	-942.110	*-979.851*	37.742
Lowest relaxed		*(lowest)*	
22 to 21	-938.955	**-956.709**	17.754
Highest relaxed		**(highest)**	
3 to 4	*-945.396*	-957.504	*12.109*
Lowest relaxation	*(lowest)*		*(lowest)*
25 to 24	-938.321	-977.522	**39.200**
Highest relaxation			**(highest)**
Original cluster	-943.394 (percentile 8.33)	-966.120 (percentile 64.29)	22.726 (percentile 17.86)

estimated by image contrast) is among the most stable studied. The original cluster requires comparatively little structural perturbation to reach an energy minimum as defined by the optimisation threshold, but other structures which undergo more perturbation to reach a minimum find lower local minima.

Only seven structures have a lower unrelaxed energy than the original cluster. Unrelaxed, relaxed and relaxation energies for these clusters are given in

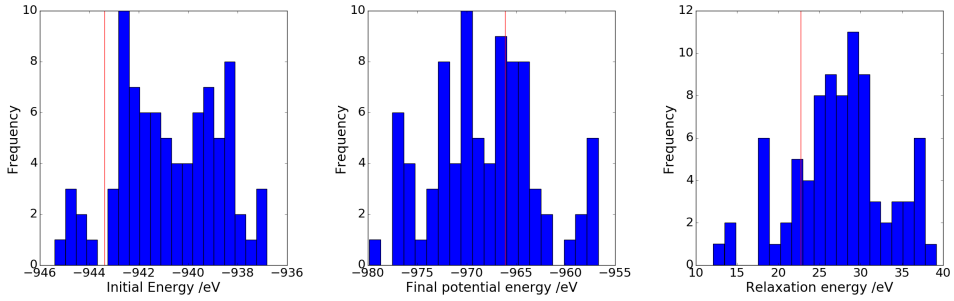

Figure 7.7: *Histograms showing the number of structures in different energy ranges. Left: unrelaxed structures, red line = unrelaxed energy of the original cluster; centre: relaxed structures, red line = relaxed energy of original cluster; right: relaxation energies, red line = relaxation energy of original cluster.*

Table 7.8: Unrelaxed, relaxed and relaxation energy (eV) of the seven structures from among the 84 studied with a lower unrelaxed energy than the original cluster and determined using the MM potential.

	Unrelaxed energy /eV	Relaxed energy /eV	Relaxation energy /eV
27 to 28	-943.729	-961.837	18.108
34 to 35	-944.294	-958.562	14.268
36 to 35	-944.605	-970.640	26.035
37 to 36	-944.155	-975.795	31.640
3 to 4	-945.395	-957.504	12.108
41 to 40	-944.699	-968.547	23.848
5 to 4	-944.713	-962.908	18.194
Original cluster	-943.394 (percentile 8.33)	-966.120 (percentile 64.29)	22.726 (percentile 17.86)

Table 7.8. It is interesting to note that almost all of these structures also have a lower relaxation energy than the original cluster. This is fairly intuitive given that the unrelaxed energies are low: they were already close to a local minimum before relaxation. It is also observed that many of the relaxed energies for these structures are higher (less negative) than that of the original structure, some quite considerably.

The structures that possess the highest and lowest unrelaxed, relaxed and relaxation energies are examined in greater detail. Each is shown in Figure 7.8, which includes a representation of the number of atoms in each stack and a visualisation of the relaxed structure. Some optimised structures are distorted to the point of being barely recognisable as modifications of the original cluster, while in some cases parts of the relaxed cluster display strikingly little distortion from the unrelaxed cluster. The range of levels of perturbation suggest a shallow energy landscape as defined by the MM potential with several local minima close in energy to each other. The clusters for which little distortion is seen are the 3 to 4 and 22 to 21 cases. The 3 to 4 case has low unrelaxed and relaxation energies; it is intuitive that a low level of distortion goes hand in hand with a low relaxation energy. The 22 to 21 case has a high relaxed energy. A large part of it seems to have been unable to rearrange to

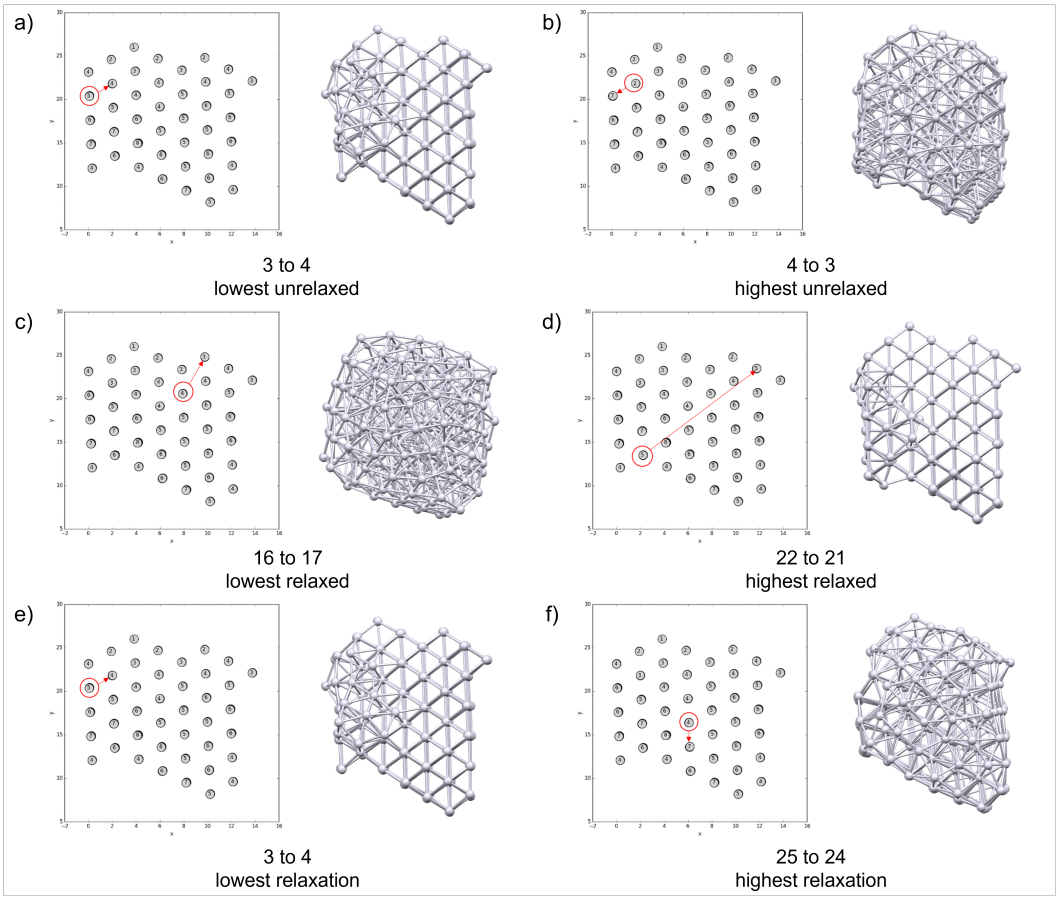

Figure 7.8: *Representations of the perturbed structures with the highest and lowest relaxed, unrelaxed and relaxation energy. In each case left; plot of the migration; right: visualisation of the optimised structure. Note that 3 to 4 appears twice as it has the lowest unrelaxed and relaxation energy.*

find a more favourable configuration.

In comparison to the extremes illustrated in Figure 7.8, the original cluster (Figure 7.6) visually displays an intermediate amount of distortion when relaxed. The data presented lend some support to the experimental estimation of the cluster's structure by suggesting that the structure suggested is reasonably favourable. It has a relatively low unrelaxed energy, and does not need to undergo high levels of distortion to reach a local minimum in the way that some perturbed structures do. Its existence under experimental conditions appears feasible. However, during optimisation it undergoes a notable level of distortion, including in those parts of the structure which are not on the surface and may in reality be held fixed on the C support. This limits the usefulness of conclusions about the experimental observations drawn directly

from the optimisation, but may be addressed by further study of the clusters using constrained optimisation to prevent distortion in lower layers which exhibit less freedom.

The 84 perturbations considered here lead to data of rough statistical quality, with a number of data points which is reasonably large, but not large enough to paint a complete statistical picture of the various small perturbations possible in the cluster. It would be possible to expand the study by considering all possible migrations of a single atom on the surface, leading to 1,722 perturbations. Other aspects of the cluster may also be examined. For example, it may be instructive to consider the relationship between total or average coordination number of cluster atoms and the modelled interaction energy.

7.2.2 Constraints Applied to Cluster Layers

To address the impact of the carbon support, a first approximation can treat selected atoms within the structure as fixed on the support and therefore frozen to any translation. This requires constrained optimisation, the capacity for which was not included in the code of Tailor et al. [7] The code was edited as part of this work to implement constrained optimisation by assigning a force of zero to selected atoms at every step of the optimisation procedure. The implementation was tested using example small Pt clusters and then used to constrain lower layers of the original Pt cluster and observe the effects of the constraints on optimisation. Successive layers were constrained and the resulting geometries and energies of the relaxed structures examined.

The experimentally estimated structure of the imaged cluster is rough on both surfaces in the z-direction, so there are different ways that a layer could be defined. One is that the bottom layer consists of the atom with the lowest z-coordinate in each stack, the next layer the next lowest and so on. This would result in all of the lowest layers having the same number of atoms, but each

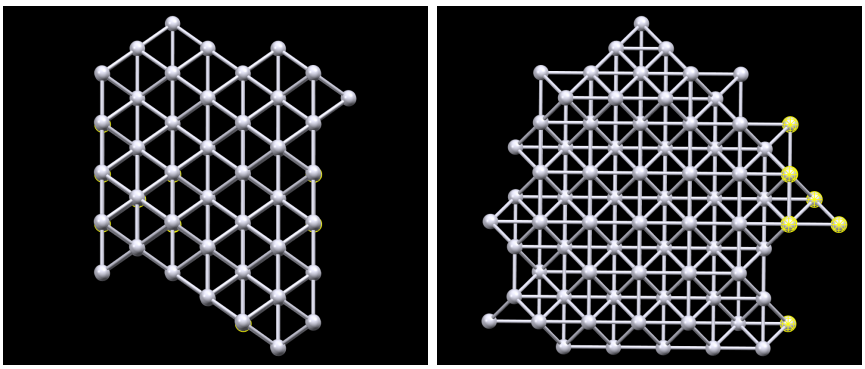

Figure 7.9: *Visualisations of the original cluster with the atoms in the first three layers highlighted in yellow to illustrate the definition of layers used here. Left: view down the z-axis; right: 90°rotation about the y-axis from original viewing angle.*

layer containing atoms with different z-coordinates. Another is that each layer consists of atoms with the same z-coordinates (within a tolerance), resulting in layers with different numbers of atoms, each defined by a horizontal slice through the material. The second definition is used, and is illustrated for clarity in Figure 7.9, in which the first three layers (the three layers with the lowest z-coordinates) of the cluster are highlighted in yellow.

By this definition, the cluster has 15 layers. Each layer was frozen in turn, cumulatively, and the cluster was optimised at each stage. That is, the first layer (lowest z-coordinate) was frozen and the cluster optimised, then the process was repeated with the first and second layers frozen, then the first, second and third, and so on. This resulted in 14 separate optimisations (with the final layer all atoms are frozen so there is no optimisation, only a single-point energy calculation). The changing relaxed energy with changing number of layers frozen, and number of atoms frozen, is displayed in Table 7.9 and Figure 7.10.

At first glance there may appear to be an issue here in that when one layer is frozen the relaxed energy of the cluster as calculated by the MM potential is -979.015 eV, which is more than 10 eV lower than that of the unconstrained cluster, -966.120 eV. In general the lowest achievable energy of an unconstrained minimisation is lower than or equal to the lowest achievable minimum of a constrained optimisation. In this case, the constrained

Table 7.9: Relaxed energies of the Pt cluster with successive layers frozen as calculated using the MM potential.

Number of layers frozen	Number of atoms frozen	Relaxed energy /eV
0	0	-966.120
1	1	-979.015
2	2	-964.979
3	10	-961.901
4	21	-955.914
5	39	-956.465
6	54	-955.151
7	76	-951.466
8	95	-951.161
9	118	-947.078
10	137	-946.827
11	159	-949.683
12	174	-946.885
13	189	-944.033
14	196	-934.623
15	199	-943.394

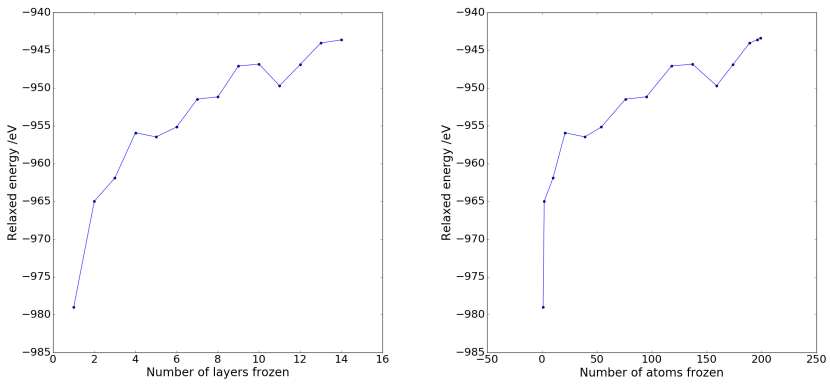

Figure 7.10: *Plots of the degree of atom constraint against the relaxed energy of the Pt cluster as calculated using the MM potential. Left: x-axis is number of layers frozen; right: x-axis is number of atoms frozen.*

cluster in question has one atom frozen, so the expectation may be that the two relaxations would result in the same structure with the atoms shifted in absolute but not relative position, and give the same energy. It may appear somewhat counter-intuitive that the constrained optimisation would result in a lower energy than the free one, but it must be remembered that the minimisation attempted here is local and not global. It has already been seen that the Pt_{199} cluster has several local minima in the vicinity of the optimised geometry with only fairly shallow separation. Two relevant minima are observed here, one at -966.120 eV, the other at -979.015 eV. The -966.120 minimum is easier to access in the free relaxation but becomes harder to access when one atom is frozen: the other atoms would have to move significantly to reach the conformation, so it is replaced as the most accessible minimum by the lower -979.015 eV minimum.

To test this explanation, an unconstrained optimisation was performed starting from the final geometry of the constrained optimisation in which the first layer had been frozen. The resulting relaxed energy was -979.015 eV, the same as for the constrained optimisation, and the resulting atomic positions are very similar with slight variations. This supports the idea that the constrained optimisation found a genuine local minimum. The existence of multiple close-lying local minima is promising for catalysis applications, [288] although it additionally highlights the sensitivity of the optimisation to small changes in calculation setup. A more robust analysis of the stability of the selected cluster under different conditions would involve constrained global optimisation, [22, 288] a large undertaking for a cluster of this size but possibly fruitful in understanding of the behaviour of Pt clusters under imaging conditions and of the sensitivity of the MM potential.

For all optimisations in which more than one layer was constrained, constraints led to energies less negative than the unconstrained optimisation. Results follow a pattern whose nature is generally not surprising: as more atoms are frozen, the relaxed energy increases up to the unrelaxed energy of -943.394

eV as the cluster is no longer able to reach lower energy conformations. In some places the energy drops when an additional layer is frozen. The reason for this is likely similar to those discussed above in terms of accessibility of close-lying local minima.

Individual relaxations are visualised in Figure 7.11. All visualisations are from the same two angles as those in Figure 7.9. In the interest of space, not all of the 14 structures are shown. From these images, the number of layers which must be constrained for the distortion of the cluster to resemble surface effects can be deduced. Bulk distortion becomes insignificant and surface distortion dominates when 5-7 layers of atoms are frozen (7-9 layers are free to move). For less than this, significant lower layer distortion is observed; for more, the surface is highly constrained and its motion may not represent surface migrations. This deduction can be useful in further considerations of the most appropriate constrained methods to use when modelling surface effects on this metal nanocluster and when modelling cluster behaviour on solid supports in general.

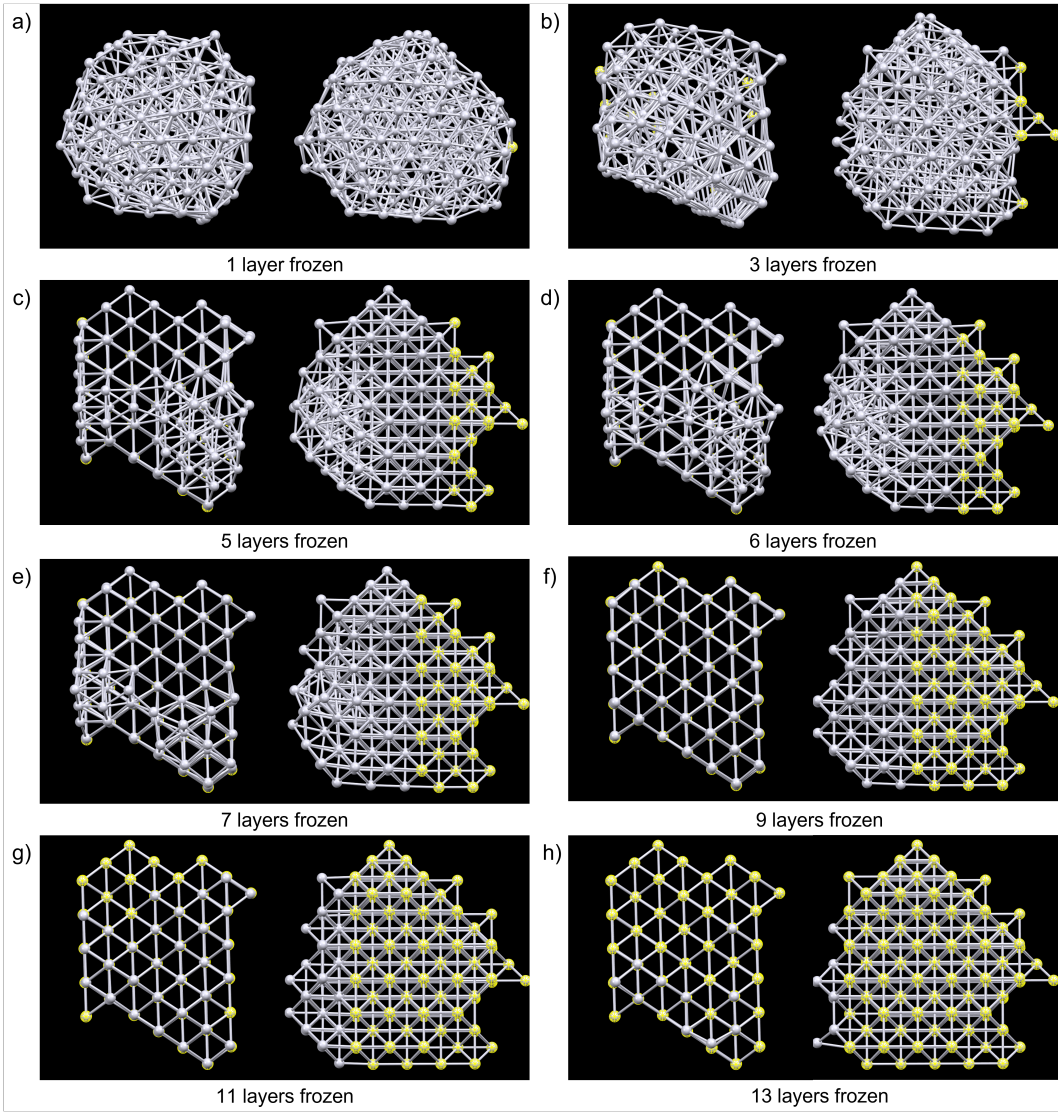

Figure 7.11: *Visualisations of the relaxed structure along the z-axis and along a 90° rotation from the z-axis following a constrained optimisation with increasing numbers of layers of atoms frozen. Atoms highlighted in yellow are frozen in each case.*

7.3 Conclusion

Over the course of this chapter, the modelling of metallic clusters using the many-body atomic solid MM potential has been explored in detail with a focus on Pt clusters. The modelling setup used, as well as the potential and its parameterisation, were robustly tested, before being turned to a real cluster whose structure has been estimated experimentally but which has not yet been modelled computationally. The potential was shown to reproduce literature experimental and ab initio data to a good degree of faithfulness. Indeed, the deviation of Pt MM surface energy from a given ab initio surface energy is less than that of certain ab initio surface energies calculated using different methods from each other.

Calculation of cluster energies following migration of selected atoms within the structure has shown a shallow potential energy landscape. The thermodynamics of the energy landscape suggest that atoms on the surface of a cluster appear readily able to change position, though kinetics have not been assessed in this work. This is in itself an interesting observation from the point of view of cluster growth processes. Coupled with the uncertainty associated with experimental atom counting, it suggests that the estimated nuclearity of the specific structure in question may be the true experimental nuclearity but is not guaranteed to be.

The possible effects of a C substrate on the freedom of a cluster and the implications of this for cluster modelling were assessed using constrained geometry optimisations. It was seen that for the 15-layer cluster considered, constraint of 5-7 layers of atoms resulted in a modelled situation in which surface atoms were free to rearrange to an extent, but the cluster could not rearrange in a bulk sense. This number range may be useful for future modelling aiming to examine surface distortions of clusters bound to supports.

There is still far more that may be attempted in terms of using the MM potential to model clusters. It may be explored for modelling of alloys, which

would greatly increase its scope, or employed within large scale modelling regimes such as MC or MD, allowing it to give information about the statistical behaviour of metal systems.

Chapter 8

Conclusions

Over the course of this work, properties and behaviour of metal organic frameworks within the wider context of advanced functional materials in general have been examined using a range of selected computational modelling and statistical methods. Computational methods may be highly useful in predicting and understanding properties of functional materials, but are limited by inherent approximations and do not always display full agreement with experimental results. The limitations and approximations associated with selected methods have been discussed during this work. Several of the properties examined have related to the celebrated and highly promising gas uptake abilities of MOFs. Their behaviour under imaging conditions has also been addressed, as well as modelling of small metal clusters in the context of imaging under similar conditions. Understanding of the behaviour of metal clusters widens understanding of metal behaviour as a whole, of which the behaviour of MOF metal centres is a part, while metal clusters may also be used in conjunction with MOFs for certain applications.

First, the aim of finding materials which can efficiently and effectively perform the difficult but important separation of xenon and krypton gases was pursued (chapter 3). This is an area in which porous materials and particularly MOFs show promise, and in this context a family of MOF structures, the MFM family (MFM-126-128 and MFM-136-138), was selected to study based

on multiple considerations. For one thing, trends in structural properties were identified in the literature and resemblance between the structural properties of the MFM MOFs and those of top-performing published MOFs was seen. For another, the MFM MOFs have already been experimentally synthesised and shown to display other promising functionality, and it was desirable to establish the MFM family as multifunctional materials.

Geometrical properties of the MFM MOFs were calculated using methods based on Voronoi networks, and gas sorption was simulated using GCMC simulations combined with classical force field methods. The study was additionally supplemented by computational calculation of infinite dilution Henry constants and heat of adsorption of the gases within each of the MOFs, and by initial analysis of temperature effects on uptake. Useful Xe and Kr uptake and separation properties were indicated by the simulations, and Xe/Kr separation may be added to the properties of this multifunctional family of MOFs.

The search for materials with ever improved Xe/Kr separation potential continues. In the case of the family of MOFs studied in this work, there is more that could be done to probe their Xe/Kr uptake properties. With the indication given here of their usefulness for the separation, it would be instructive for the properties to also be measured experimentally; computational predictions of strong separation performance give useful indications, but can be overpredictions, so experimental confirmation is necessary once well-performing MOFs are identified. There is additionally scope to further probe their properties computationally. It is likely that the MOFs display some level of flexibility to admit guest molecules slightly larger than the limiting diameters of their pores. Using simulations which account for flexibility would allow the effects of flexibility on Xe and Kr uptake to be examined.

The use of MOFs for gas separations was further studied in the context of biogas upgrading (chapter 4). Biogas is composed primarily of CH_4 and CO_2, and upgrading its composition to achieve greater efficiency involves separation of the two gases as well as removal of low-concentration impurities leading to

higher purity of methane. The search for materials which can achieve this separation is important from the perspective of sustainability, and MOFs can be promising separation candidates. The search for the most useful MOFs from among the many possible structures benefits from computational screening of a large volume of data, which itself can benefit from machine learning methods to improve efficiency. This is central to the approach taken in this study.

The study made use of a high-throughput screening of a dataset of some thousands of published MOF structures which have been synthesised experimentally and carefully curated to address potential structural issues. Hypothetical MOFs which may display issues with synthesisability were not included in the dataset. Uptake of CO_2 and CH_4 under conditions relevant to biogas upgrading was modelled for each MOF in the curated dataset using GCMC and classical force fields, and biogas upgrading potential was measured by a metric which accounts for the trade-off between uptake and selectivity. Potentially useful MOFs were identified and their structures examined, and the analysis of well-performing MOFs for uniting structural features revealed that MOFs composed of 2D layers with narrow gaps between layers are a useful category to pursue.

Computational screenings based on force field and Monte Carlo methods can become cumbersome when large numbers of materials are modelled. Therefore, machine learning models were trained on the data resultant from the biogas screening to enable biogas upgrading properties to be predicted using geometrical and infinite dilution data, which is cheaper to calculate than uptake at working pressures. The models were trained on the curated data taken from experimentally synthesised MOFs in an effort to minimise the volume of problematic structures used as training data. The performance of the models was analysed in detail using an external dataset of hypothetical MOFs as test data. Models, in particular random forest models, were able to perform well on training and validation data, and were able to predict many aspects of test data well, but had lower performance for certain test MOFs. This over-

fitting suggests room for improvement in aspects of the models. It appears to be related to differences in distributions of structural properties between the training set and the test set.

In general, machine learning models are becoming a very powerful tool in the search for MOFs with properties tuned to particular applications. They are likely to be used increasingly as a component of high-throughput screenings and experience continued development in quality and efficiency in the coming years. They may also be instructive for assessing the usefulness of identified MOFs, for example in predicting synthesisability or synthesis routes. For the particular models trained as part of this work, there is scope for improvement by supplementing the training set with data on MOFs with structural features which are underrepresented within it to better represent the diversity of structures available to MOFs. The differences between the curated dataset of real MOFs used in training and the curated dataset of hypothetical MOFs used as test data additionally highlight the differences between real and hypothetical MOFs, and raise questions about potential improvement to the useful category of databases which are composed of hypothetical MOFs.

Further questions relating to the precise structures of MOFs were then addressed in the form of examination of the effect of residual solvent on gas sorption properties in MOFs (chapter 5). Experimentally, solvent molecules may remain in MOF pores following synthesis and activation procedures. It is not always known in computational studies whether or how much solvent must be stripped from a structure in order to most accurately model an equivalent experimental structure. Furthermore, in both experimental and computational studies, the effect that solvent molecules may have on behaviour, including gas uptake properties, is not fully understood. This includes the questions of whether there is a significant effect, how significant the effect is, and whether the effect is positive or negative (for gas uptake, that is whether solvent causes more or less gas to be adsorbed).

In this work, some of these questions were addressed via a high-throughput

screening of a second curated dataset of experimental MOFs. Focus remained on the important and chemically different gases CO_2 and CH_4. GCMC simulations and classical force fields were used to simulate uptake of both gases in a desolvated and a solvated form. The difference in uptake between desolvated and solvated forms was measured and used as a metric for solvent effect. Solvent effect was analysed at different pressures and gas compositions, and the effect of MOF geometry on solvent effect was assessed. Ab initio calculations were then applied to small model systems to probe the chemical roots of solvent effect. For both gases, a small solvent effect was seen for some MOFs, but a larger effect for others. Where there was a large effect on CO_2 adsorption, it was almost exclusively a negative effect, while solvent effect on CH_4 adsorption could be positive or negative depending on the MOF. Geometry appeared to be a factor influencing solvent effect, with MOFs with small pore sizes being far more likely to experience large solvent effect than those with large pores.

This study has cemented the possible significance of solvent effects on gas uptake in MOFs and has highlighted the fact that solvent effects can be positive in certain situations, as well as providing some initial guidelines of which MOFs are most likely to experience strong solvent effects in either direction. It is to be hoped that solvent effects on MOFs will continue to be probed under a wider range of conditions to add to the base of knowledge about the situations in which solvent effect is positive, negative, strong or weak. It may be possible to train useful machine learning models to predict significance and direction of solvent effect based on information about a MOF including structural information. In studies of new MOFs, there is merit to consideration of whether solvent effects are likely to be significant, and if so whether it is particularly important to eliminate solvent molecules by robust methods, or whether it is desirable to promote their presence.

Turning from a close focus on gas sorption properties of MOFs, this work also probed techniques for imaging their structures in the form of TEM (chapter 6). This is a very useful method for obtaining structural information which

can be applied to crystals unsuitable for other conventional imaging methods. It generates direct images of the structure of a material, even to the point of atomic resolution, and has the additional advantage of being able to image dynamic processes, which it may also promote. However, it is hampered by damage that the high-energy electron beam can cause to imaged materials: damage to which MOFs are particularly susceptible. The origins of this damage and their relation to structural properties relevant to a selection of MOF structures were studied in this work.

The study involved simulation of knock-on damage events in flakes of relevant MOF structures by ab initio molecular dynamics methods based on Born-Oppenheimer DFT. The use of such methods amounts to an approximation for a number of reasons. Limitations of DFT methods and relative advantages and disadvantages of different functionals in modelling electronic properties of MOFs were discussed; methods were selected in the context of these limitations. Damage was simulated for a range of energies transferred to each relevant atom, and this was used to determine thresholds for ejection of atoms or fragments from the different MOFs. From these, ejection cross sections, which define the probability of ejection at a given beam energy, were calculated. It was seen that the ejection events which may occur can follow complex fragmentation patterns, and are not always simply atomic ejection of the impacted atom. Knock-on damage was compared between the different MOFs and conclusions drawn about the susceptibility of atoms and bonds to knock-on damage. Total computational damage metrics were also compared to experimental damage metrics to draw conclusions about which damage events are caused primarily by knock-on damage and which involve significant contribution from other factors not captured by the ab intio dynamics simulations applied. Damage to the highly conductive MOF BHT was seen to be well-described by purely knock-on events, but the three more insulating MOFs appeared to experience other significant contributions.

The complex fragmentation patterns observed in this study merit future

consideration in beam damage studies in general, and indicate that models of damage as purely atomic ejection are often unlikely to be sufficient. Meanwhile, given the indication that damage to the insulating MOFs is not described by only knock-on models it would be instructive to further study these structures to elucidate other damage mechanisms. It has previously been seen that damage caused by excitation of electrons, radiolysis, can be far more significant for insulating materials than for conductive ones. However, this generally appears to apply for much lower beam energies than those focused on in this study, and the specifics of effects on the selected MOFs at higher beam energy are not known.

The behaviour of materials in the context of electron beam imaging was further studied in this work with models of metal clusters aiming to further understand general behaviour of metal atoms in different forms, with a primary focus on platinum (chapter 7). Metal clusters have relevance in a number of applications, most notably catalysis, and may also be combined with MOF structures to improve the catalytic processes they undergo. Clusters can be grown on carbon supports and simultaneously imaged by irradiation of metals under high-energy electron beam. The particular sizes and geometries of cluster obtained by the growth process are very relevant to the behaviour of metal clusters, so it is important to understand interactions between metal atoms in different geometries.

Modelling metal clusters requires careful consideration of computational techniques: they cannot be accurately described by the cheaper 2-body potentials which are often sufficient for gas uptake modelling in MOFs, while use of ab initio simulations is limited by size and by the difficulty that common DFT methods have in describing the multireference character of metal atoms. Therefore, this work involved careful assessment of the performance of a selected 2+3 body potential, the Murrell-Mottram potential and its parameterisation for Pt and Pd in modelling clusters of various sizes. Following verification, the potential was used to examine an experimentally estimated

geometry of an imaged platinum cluster, how its energy is influenced by small structural perturbations, and the interaction of the cluster with the carbon support, as approximated by constrained optimisation. The thermodynamic landscape occupied by the cluster was seen to be populated by shallow wells allowing lability of atoms on its surface.

Modelling of metal clusters is a very active field of research, which involves modelling their structures and their catalytic properties by energetic and kinetic methods. With the Murrell-Mottram potential verified and used here to make interesting observations about an example of a Pt cluster, there is scope for it to be further used in cluster modelling. In its current form it can be useful for modelling structure, stability, and thermodynamic processes in pure Pt and Pd metals, along with other metals for which it has been parameterised. Further development may enable it to be used for systems composed of more than one element, initially alloys.

The questions addressed throughout this work have covered a range of MOF behaviour and properties, in particular relating to gas uptake applications for which MOFs are well-known. They have also covered representation of MOF structures from experimental and computational perspectives, how the two may be reconciled, the effect of structural defects (eg. solvent molecules) on MOF behaviour, and study of the ways in which MOF and other material structures may be imaged and better understood. Several different computational techniques have been used to address these questions, highlighting the range and versatility of computational methods and the situations to which they can be applied, while the approximations inherent to computational modelling and the resulting limitations have also been discussed.

Bibliography

[1] J. D. Humby, O. Benson, G. L. Smith, S. P. Argent, I. Da Silva, Y. Cheng, S. Rudić, P. Manuel, M. D. Frogley, G. Cinque, L. K. Saunders, I. J. Vitórica-Yrezábal, G. F. Whitehead, T. L. Easun, W. Lewis, A. J. Blake, A. J. Ramirez-Cuesta, S. Yang, and M. Schröder, "Host-Guest Selectivity in a Series of Isoreticular Metal-Organic Frameworks: Observation of Acetylene-to-Alkyne and Carbon Dioxide-to-Amide Interactions," *Chem. Sci.*, vol. 10, pp. 1098–1106, jan 2019.

[2] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *The Molecular Theory of Gases and Liquids*. New York: Wiley, 2nd ed., 1964.

[3] O. Talu and A. L. Myers, "Reference Potentials for Adsorption of Helium, Argon, Methane, and Krypton in High-Silica Zeolites," *Colloids Surf., A*, vol. 187-188, pp. 83–93, aug 2001.

[4] A. K. Rappe, C. J. Casewit, K. S. Colwell, A. Goddard III, W, and W. M. Skiff, "UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations," *J. Am. Chem. Soc.*, vol. 114, no. 25, pp. 10024–10035, 1992.

[5] J. Glover and E. Besley, "Modelling Sorption and Diffusion Behaviour in Porous Solids," in *Computer Simulation of Porous Materials: Current Approaches and Future Opportunities* (K. Jelfs, ed.), ch. 4, pp. 122–214, London: Royal Society of Chemistry, 2022.

[6] H. Cox, "Potential-Energy Functions for Platinum and Palladium Solids and their Application to Surfaces," *Surf. Sci.*, vol. 397, pp. 374–381, feb 1998.

[7] P. M. Tailor, R. J. Wheatley, and N. A. Besley, "An Empirical Force Field for the Simulation of the Vibrational Spectroscopy of Carbon Nanomaterials," *Carbon*, vol. 113, pp. 299–308, mar 2017.

[8] L. D. Lloyd and R. L. Johnston, "Theoretical Analysis of 17-19-Atom Metal Clusters using Many-Body Potentials," *J. Chem. Soc. Dalton Trans.*, pp. 307–316, jan 2000.

[9] R. Tran, Z. Xu, B. Radhakrishnan, D. Winston, W. Sun, K. A. Persson, and S. P. Ong, "Data Descriptor: Surface Energies of Elemental Crystals," *Scientific Data*, vol. 3, p. 160080, sep 2016.

[10] D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Chen, B. Smit, J. B. Parise, M. Haranczyk, and P. K. Thallapally,

"Metal–Organic Framework with Optimally Selective Xenon Adsorption and Separation," *Nat. Commun.*, vol. 7, p. 11831, jun 2016.

[11] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr, "Advances, Updates, and Analytics for the Computation Ready, Experimental Metal Organic Framework Database: CoRE MOF 2019," *J. Chem. Eng. Data*, vol. 64, no. 12, pp. 5985–5998, 2019.

[12] Y. S. Bae, B. G. Hauser, Y. J. Colón, J. T. Hupp, O. K. Farha, and R. Q. Snurr, "High Xenon/Krypton Selectivity in a Metal-Organic Framework with Small Pores and Strong Adsorption Sites," *Microporous Mesoporous Mater.*, vol. 169, pp. 176–179, mar 2013.

[13] U. Mueller, M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndt, and J. Pastré, "Metal–Organic Frameworks—Prospective Industrial Applications," *J. Mater. Chem*, vol. 16, pp. 626–636, feb 2006.

[14] S. Pawsey, I. Moudrakovski, J. Ripmeester, L.-Q. Wang, G. J. Exarhos, J. L. C. Rowsell, and O. M. Yaghi, "Hyperpolarized 129 Xe Nuclear Magnetic Resonance Studies of Isoreticular Metal-Organic Frameworks," *J. Phys. Chem. C*, vol. 111, no. 16, pp. 6060–6067, 2007.

[15] J. A. Greathouse, T. L. Kinnibrugh, and M. D. Allendorf, "Adsorption and Separation of Noble Gases by IRMOF-1: Grand Canonical Monte Carlo Simulations," *Ind. Eng. Chem. Res.*, vol. 48, pp. 3425–3431, apr 2009.

[16] P. Ryan, O. K. Farha, L. J. Broadbelt, and R. Q. Snurr, "Computational Screening of Metal-Organic Frameworks for Xenon/Krypton Separation," *AIChE Journal*, vol. 57, pp. 1759–1766, jul 2011.

[17] D. W. Breck, *Zeolite Molecular Sieves: Structure, Chemistry and Use.* Malabar, FL: R.E. Krieger, 1984.

[18] J. Glover and E. Besley, "A High-Throughput Screening of Metal–Organic Framework Based Membranes for Biogas Upgrading," *Faraday Discuss.*, vol. 231, pp. 235–257, oct 2021.

[19] C. Altintas, G. Avci, H. Daglar, A. Nemati, V. Azar, I. Erucar, S. Velioglu, and S. Keskin, "An Extensive Comparative Analysis of Two MOF Databases: High-Throughput Screening of Computation-Ready MOFs for CH_4 and H_2 adsorption," *J. Mater. Chem. A*, vol. 7, no. 16, pp. 9593–9608, 2019.

[20] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, and R. Q. Snurr, "Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals," *Chem. Mater.*, vol. 26, pp. 6185–6192, nov 2014.

[21] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, and D. Fairen-Jimenez, "Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future," *Chem. Mater*, vol. 29, no. 7, pp. 2618–2625, 2017.

[22] S. K. Ignatov, A. G. Razuvaev, A. S. Loginova, and E. Masunov, "Global Structure Optimization of Pt Clusters Based on the Modified Empirical Potentials, Calibrated using Density Functional Theory," *J. Phys. Chem. C*, vol. 123, p. 19, 2019.

[23] P. Nava, M. Sierka, and R. Ahlrichs, "Density Functional Study of Palladium Clusters," *Phys. Chem. Chem. Phys.*, vol. 5, pp. 3372–3381, aug 2003.

[24] H. Li, L. Li, R.-B. Lin, W. Zhou, Z. Zhang, S. Xiang, and B. Chen, "Porous Metal-Organic Frameworks for Gas Storage and Separation: Status and Challenges," *EnergyChem*, vol. 1, p. 100006, jul 2019.

[25] A. Sturluson, M. T. Huynh, A. R. Kaija, C. Laird, S. Yoon, F. Hou, Z. Feng, C. E. Wilmer, Y. J. Colón, Y. G. Chung, D. W. Siderius, and C. M. Simon, "The Role of Molecular Modelling and Simulation in the Discovery and Deployment of Metal-Organic Frameworks for Gas Storage and Separation," *Mol. Simul.*, vol. 45, no. 14-15, pp. 1082–1121, 2019.

[26] M. Ko, L. Mendecki, and K. A. Mirica, "Conductive Two-Dimensional Metal–Organic Frameworks as Multifunctional Materials," *Chem. Commun*, vol. 54, pp. 7873–7891, jul 2018.

[27] M. Omrani, M. Goriaux, L. Jean-Soro, and V. Ruban, "Platinum Group Elements in Atmospheric PM_10 Particles and Dry Deposition in France," *Environ. Sci. and Pollution Research*, vol. 28, pp. 33231–33240, 2021.

[28] N. Baig, I. Kammakakam, and W. Falath, "Nanomaterials: a Review of Synthesis Methods, Properties, Recent Progress, and Challenges," *Mater. Adv*, vol. 2, pp. 1821–1871, 2021.

[29] Z. Hu and M. P. Srinivasan, "Mesoporous High-Surface-Area Activated Carbon," *Micropor. Mesopor. Mater.*, vol. 43, pp. 267–275, may 2001.

[30] S. Cavenati, C. A. Grande, and A. E. Rodrigues, "Adsorption Equilibrium of Methane, Carbon Dioxide, and Nitrogen on Zeolite 13X at High Pressures," *J. Chem. Eng. Data.*, vol. 49, pp. 1095–1101, jul 2004.

[31] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, and A. I. Cooper, "Separation of Rare Gases and Chiral Molecules by Selective Binding in Porous Organic Cages," *Nat. Mater.*, vol. 13, pp. 954–960, jul 2014.

[32] C. S. Diercks and O. M. Yaghi, "The Atom, the Molecule, and the Covalent Organic Framework," *Science*, vol. 355, no. 6328, pp. 924–952, 2017.

[33] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, "Large-Scale Screening of Hypothetical Metal–Organic Frameworks," *Nat. Chem.*, vol. 4, pp. 83–89, nov 2012.

[34] Y. Kinoshita, I. Matsubara, T. Higuchi, and Y. Saito, "The Crystal Structure of Bis(adiponitrilo)copper(I) Nitrate," *Bull. Chem. Soc. Jpn.*, vol. 32, pp. 1221–1226, nov 1959.

[35] B. F. Hoskins and R. Robson, "Infinite Polymeric Frameworks Consisting of Three Dimensionally Linked Rod-like Segments," *J. Am. Chem. Soc.*, vol. 111, no. 15, pp. 5962–5964, 1989.

[36] O. M. Yaghi and H. Li, "Hydrothermal Synthesis of a Metal-Organic Framework Containing Large Rectangular Channels," *J. Am. Chem. Soc*, vol. 117, pp. 10401–10402, 1995.

[37] S. S. Chui, S. M. Lo, J. P. Charmant, A. G. Orpen, and I. D. Williams, "A Chemically Functionalizable Nanoporous Material $[Cu_3(TMA)_2(H_2O)_3]_n$," *Science*, vol. 283, pp. 1148–1150, feb 1999.

[38] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi, "Design and Synthesis of an Exceptionally Stable and Highly Porous Metal-Organic Framework," *Nature*, vol. 402, pp. 276–279, nov 1999.

[39] B. F. Abrahams, B. F. Hoskins, and R. Robson, "A New Type of Infinite 3D Polymeric Network Containing 4-Connected, Peripherally Linked Metalloporphyrin Building Blocks," *J. Am. Chem. Soc*, vol. 113, pp. 3606–3607, 1991.

[40] J. Long, O. Yaghi, J. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S. T. Nguyen, and J. T. Hupp, "Metal-organic Framework Materials as Catalysts," *Chem. Soc. Rev.*, vol. 38, pp. 1450–1459, 2009.

[41] L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. Van Duyne, and J. T. Hupp, "Metal-Organic Framework Materials as Chemical Sensors," *Chem. Rev.*, vol. 112, pp. 1105–1125, 2012.

[42] P. Horcajada, R. Gref, T. Baati, P. K. Allan, G. Maurin, P. Couvreur, G. Erard F Erey, R. E. Morris, and C. Serre, "Metal-Organic Frameworks in Biomedicine," *Chem. Rev.*, vol. 112, pp. 1232–1268, 2012.

[43] P. Z. Moghadam, A. Li, X. W. Liu, R. Bueno-Perez, S. D. Wang, S. B. Wiggin, P. A. Wood, and D. Fairen-Jimenez, "Targeted Classification of Metal–Organic Frameworks in the Cambridge Structural Database (CSD)," *Chem. Sci.*, vol. 11, pp. 8373–8387, aug 2020.

[44] L. J. Murray, M. Dincă, and J. R. Long, "Hydrogen Storage in Metal–Organic Frameworks," *Chem. Soc. Rev.*, vol. 38, no. 5, p. 1294, 2009.

[45] B. Chen, N. W. Ockwig, A. R. Millward, D. S. Contreras, and O. M. Yaghi, "High H_2 Adsorption in a Microporous Metal-Organic Framework with Open Metal Sites," *Ang. Chem. Int. Ed.*, vol. 44, no. 30, pp. 4745–4749, 2005.

[46] M. Ding, R. W. Flaig, H.-L. Jiang, and O. M. Yaghi, "Carbon Capture and Conversion using Metal-Organic Frameworks and MOF-Based Materials," *Chem. Soc. Rev*, vol. 48, pp. 2783–2828, 2019.

[47] A. R. Millward and O. M. Yaghi, "Metal-Organic Frameworks with Exceptionally High Capacity for Storage of Carbon Dioxide at Room Temperature," *J. Am. Chem. Soc*, vol. 127, no. 51, pp. 17998–17999, 2005.

[48] X. Y. Chen, H. Vinh-Thang, A. A. Ramirez, D. Rodrigue, and S. Kaliaguine, "Membrane Gas Separation Technologies for Biogas Upgrading," *RSC Adv.*, vol. 5, pp. 24399–24448, mar 2015.

[49] J. Marshall and A. C. Bird, "A Comparative Histopathological Study of Argon and Krypton Laser Irradiations of the Human Retina," *Br. J. Opthamol.*, vol. 63, no. 10, pp. 657–668, 1979.

[50] D. Banerjee, C. M. Simon, S. K. Elsaidi, M. Haranczyk, and P. K. Thallapally, "Xenon Gas Separation and Storage Using Metal-Organic Frameworks," *Chem*, vol. 4, pp. 466–494, 2018.

[51] R. D. Sanders, D. Ma, and M. Maze, "Xenon: Elemental Anaesthesia in Clinical Practice," *Br. Med. Bul.*, vol. 71, pp. 115–135, 2005.

[52] C. Degueldre, R. Dawson, I. Cooley, and E. Besley, "Fission Gas Released from Molten Salt Reactor Fuel: the Case of Noble Gas Short Life Radioisotopes for Radiopharmaceutical Application," *Med. Novel Technol. Devices*, vol. 10, p. 100057, jun 2021.

[53] S. Yeganegi, M. Gholami, and V. Sokhanvaran, "Molecular Simulations of Adsorption and Separation of Acetylene and Methane and their Binary Mixture on MOF-5, HKUST-1 and MOF-505 Metal–Organic Frameworks," *Mol. Simul.*, vol. 43, no. 4, pp. 260–266, 2017.

[54] Y. Ji, L. Ding, Y. Cheng, H. Zhou, S. Yang, F. Li, and Y. Li, "Understanding the Effect of Ligands on C2H2 Storage and C_2H_2/CH_4, C_2H_2/CO_2 Separation in Metal-Organic Frameworks with Open Cu(II) Sites," *J. Phys. Chem. C.*, vol. 121, pp. 24104–24113, nov 2017.

[55] T. Ghanbari, F. Abnisa, and W. M. A. Wan Daud, "A Review on Production of Metal Organic Frameworks (MOF) for CO2 Adsorption," *Sci. Total Environ.*, vol. 707, p. 135090, mar 2020.

[56] M. Zeng, Y. Xiao, J. Liu, K. Yang, and L. Fu, "Exploring Two-Dimensional Materials toward the Next-Generation Circuits: From Monomer Design to Assembly Control," *Chem. Rev.*, vol. 118, no. 13, pp. 6236–6296, 2018.

[57] Y. Li, H. Liu, L. Dai, C. Wang, J. Lv, X. Meng, A. Dong, B. Wang, and P. Li, "2D Titanium Catecholate Metal-Organic Frameworks with Tunable Gas Adsorption and Ionic Conductivity," *J. Mater. Chem. A*, vol. 11, p. 9136, 2023.

[58] K. Khan, A. K. Tareen, M. Aslam, R. Wang, Y. Zhang, A. Mahmood, Z. Ouyang, H. Zhang, and Z. Guo, "Recent Developments in Emerging Two-Dmensional Materials and their Applications," *J. Mater. Chem. C.*, vol. 8, no. 2, pp. 387–440, 2020.

[59] B. C. Brodie, "On the Atomic Weight of Graphite," *Philos. Trans. R. Soc. Lond.*, vol. 149, no. 149, pp. 249–259, 1859.

[60] J. D. Bernal and W. Bragg, "The Structure of Graphite," *Proc. R. soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character*, vol. 106, no. 740, pp. 749–773, 1924.

[61] P. R. Wallace, "The Band Theory of Graphite," *Phys. Rev.*, vol. 71, p. 622, may 1947.

[62] H. P. Boehm, R. Setton, and E. Stumpp, "Nomenclature and Terminology of Graphite Intercalation Compounds," *Carbon*, vol. 24, no. 2, pp. 241–245, 1986.

[63] H. Boehm, A. Clauss, G. Fischer, and U. Hofmann, "Surface Properties of Extremely Thin Graphite Lamellae," *Proceedings of the Fifth Conference on Carbon*, vol. 1, pp. 73–80, 1962.

[64] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric Field in Atomically Thin Carbon Films," *Science*, vol. 306, pp. 666–669, oct 2004.

[65] A. Chodos, "October 22, 2004: Discovery of Graphene," oct 2009.

[66] A. Geim and A. Chodos, "Many Pioneers in Graphene Discovery," jan 2010.

[67] X. Lu, M. Yu, H. Huang, and R. S. Ruhoff, "Tailoring Graphite with the Goal of Achieving Single Sheets," *Nanotechnology*, vol. 10, no. 3, pp. 269–272, 1999.

[68] C. Berger, Z. Song, T. Li, X. Li, A. Y. Ogbazghi, R. Feng, Z. Dai, N. Alexei, M. E. H. Conrad, P. N. First, and W. A. De Heer, "Ultrathin epitaxial graphite: 2D electron gas properties and a route toward graphene-based nanoelectronics," *Journal of Physical Chemistry B*, vol. 108, no. 52, pp. 19912–19916, 2004.

[69] R. B. Little, "Mechanistic Aspects of Carbon Nanotube Nucleation and Growth," *J. Clust. Sci.*, vol. 14, no. 2, pp. 135–185, 2003.

[70] A. J. Van Bommel, J. E. Crombeen, and A. Van Tooren, "LEED and Auger Electron Observations of the SiC(0001) Surface," *Surf. Sci.*, vol. 48, no. 2, pp. 463–472, 1975.

[71] S. Shaybanizadeh and A. Najafi Chermahini, "Fabricating Boron Nitride Nanosheets from Hexagonal BN in Water Solution by a Combined Sonication and Thermal-Assisted Hydrolysis Method," *Ceram. Int.*, vol. 47, no. 8, pp. 11122–11128, 2021.

[72] Q. Fu, J. Han, X. Wang, P. Xu, T. Yao, J. Zhong, W. Zhong, S. Liu, T. Gao, Z. Zhang, L. Xu, and B. Song, "2D Transition Metal Dichalcogenides: Design, Modulation, and Challenges in Electrocatalysis," *Adv. Mater.*, vol. 33, no. 6, pp. 1–24, 2021.

[73] P. Tshuma, B. C. E. Makhubela, C. A. Ndamyabera, S. A. Bourne, and G. Mehlana, "Synthesis and Characterization of 2D Metal-Organic Frameworks for Adsorption of Carbon Dioxide and Hydrogen," *Frontiers in Chemistry — www.frontiersin.org*, vol. 8, p. 581226, 2020.

[74] J. Liu, X. Song, T. Zhang, S. Liu, H. Wen, and L. Chen, "2D Conductive Metal–Organic Frameworks: An Emerging Platform for Electrochemical Energy Storage," *Ang. Chem. Int. Ed.*, vol. 60, pp. 5612–5624, mar 2021.

[75] A. Schneemann, R. Dong, F. Schwotzer, H. Zhong, I. Senkovska, X. Feng, and S. Kaskel, "2D Framework Materials for Energy Applications," *Chem. Sci.*, vol. 12, no. 5, pp. 1600–1619, 2021.

[76] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer, "Flexible Metal-Organic Frameworks," *Chem. Soc. Rev.*, vol. 43, no. 16, pp. 6062–6096, 2014.

[77] K. Barcus and S. M. Cohen, "Free-Standing Metal-Organic Framework (MOF) Monolayers by Self-Assembly of Polymer-Grafted Nanoparticles," *Chem. Sci.*, vol. 11, no. 32, pp. 8433–8437, 2020.

[78] J. Duan, Y. Sun, S. Chen, X. Chen, and C. Zhao, "A Zero-Dimensional Nickel, Iron-Metal-Organic Framework (MOF) for Synergistic N_2 Electrofixation," *J. Mater. Chem. A*, vol. 8, no. 36, pp. 18810–18815, 2020.

[79] Y. S. Wei, M. Zhang, R. Zou, and Q. Xu, "Metal-Organic Framework-Based Catalysts with Single Metal Sites," *Chem. Rev.*, vol. 120, no. 21, pp. 12089–12174, 2020.

[80] B. Widom, "Some Topics in the Theory of Fluids," *J. Chem. Phys.*, vol. 39, pp. 2808–2812, 1963.

[81] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk, "Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials," *Micropor. Mesopor. Mater.*, vol. 149, pp. 134–141, feb 2012.

[82] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, "RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials," *Mol. Simul.*, vol. 42, pp. 81–101, jan 2016.

[83] M. Fischer, J. R. Gomes, and M. Jorge, "Computational Approaches to Study Adsorption in MOFs with Unsaturated Metal Sites," 2014.

[84] M. Fischer, F. Hoffmann, and M. Fröba, "New Microporous Materials for Acetylene Storage and C_2H_2/CO_2 Separation: Insights from Molecular Simulations," *ChemPhysChem*, vol. 11, pp. 2220–2229, jul 2010.

[85] E. Haldoupis, J. Borycz, H. Shi, K. D. Vogiatzis, P. Bai, W. L. Queen, L. Gagliardi, and J. I. Siepmann, "Ab initio derived force fields for predicting CO2 adsorption and accessibility of metal sites in the metal-organic frameworks M-MOF-74 (M = Mn, Co, Ni, Cu)," *Journal of Physical Chemistry C*, vol. 119, no. 28, pp. 16058–16071, 2015.

[86] I. Cooley, L. Efford, and E. Besley, "Computational Predictions for Effective Separation of Xenon/ Krypton Gas Mixtures in the MFM Family of Metal-Organic Frameworks," *J. Phys. Chem. C*, vol. 2022, p. 57, 2022.

[87] T. Le, V. C. Epa, F. R. Burden, and D. A. Winkler, "Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties," *Chem. Rev.*, vol. 112, pp. 2889–2919, may 2012.

[88] M. Z. Aghaji, M. Fernandez, P. G. Boyd, T. D. Daff, and T. K. Woo, "Quantitative Structure–Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO_2 Working Capacity and CO_2/CH_4 Selectivity for Methane Purification," *Eur J. Inorg. Chem.*, vol. 2016, pp. 4505–4511, sep 2016.

[89] B. J. Sikora, C. E. Wilmer, M. L. Greenfield, and R. Q. Snurr, "Thermodynamic Analysis of Xe/Kr Selectivity in Over 137 000 Hypothetical Metal-Organic Frameworks," *Chem. Sci.*, vol. 3, no. 7, pp. 2177–2396, 2012.

[90] T. Düren, L. Sarkisov, O. M. Yaghi, and R. Q. Snurr, "Design of New Materials for Methane Storage," *Langmuir*, vol. 20, pp. 2683–2689, mar 2004.

[91] L. Li, L. Guo, Z. Zhang, Q. Yang, Y. Yang, Z. Bao, Q. Ren, and J. Li, "A Robust Squarate-Based Metal–Organic Framework Demonstrates Record-High Affinity and Selectivity for Xenon over Krypton," *J. Am. Chem. Soc.*, vol. 141, pp. 9358–9364, jun 2019.

[92] G. Garberoglio, A. I. Skoulidas, and J. K. Johnson, "Adsorption of Gases in Metal Organic Materials: Comparison of Simulations and Experiments," *J. Phys. Chem. B*, vol. 109, no. 27, pp. 13094–13103, 2005.

[93] D. Nazarian, J. S. Camp, Y. G. Chung, R. Q. Snurr, and D. S. Sholl, "Large-Scale Refinement of Metal-Organic Framework Structures Using Density Functional Theory," *Chem. Mater.*, vol. 29, no. 6, pp. 2521–2528, 2017.

[94] R. Anderson and D. A. Gómez-Gualdrón, "Large-Scale Free Energy Calculations on a Computational Metal-Organic Frameworks Database: Toward Synthetic Likelihood Predictions," *Chem. Mater.*, vol. 32, no. 19, pp. 8106–8119, 2020.

[95] F. X. Coudert and A. H. Fuchs, "Computational Characterization and Prediction of Metal-Organic Framework Properties," *Coord. Chem. Rev.*, vol. 307, pp. 211–236, 2016.

[96] H. Jiang, "First-Principles Approaches for Strongly Correlated materials: A Theoretical Chemistry Perspective," *Int.J. Quant. Chem.*, vol. 115, no. 11, pp. 722–730, 2015.

[97] J. P. Perdew, "Density Functional Theory and the Band Gap Problem," *Int. J. Quant. Chem.*, vol. 28, no. 19 S, pp. 497–523, 1985.

[98] P. Borlido, J. Schmidt, A. W. Huran, F. Tran, M. A. Marques, and S. Botti, "Exchange-Correlation Functionals for Band Gaps of Solids: Benchmark, Reparametrization and Machine Learning," *Npj Comput. Mater.*, vol. 6, no. 1, p. 96, 2020.

[99] A. Morales-García, R. Valero, and F. Illas, "An Empirical, yet Practical Way to Predict the Band Gap in Solids by Using Density Functional Band Structure Calculations," *J. Phys. Chem. C*, vol. 121, no. 34, pp. 18862–18866, 2017.

[100] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*. New York: John Wiley and Sons Inc, 2 ed., 2004.

[101] F. Gándara and T. D. Bennett, "Crystallography of Metal-Organic Frameworks," *IUCrJ*, vol. 1, no. 6, pp. 563–570, 2014.

[102] T. Chen and T. A. Manz, "Identifying Misbonded Atoms in the 2019 CoRE Metal-Organic Framework Database," *RSC Adv.*, vol. 10, no. 45, pp. 26944–26951, 2020.

[103] X. Zhang, Z. Chen, X. Liu, S. L. Hanna, X. Wang, R. Taheri-Ledari, A. Maleki, P. Li, and O. K. Farha, "A Historical Overview of the Activation and Porosity of Metal-Organic Frameworks," *Chem. Soc. Rev.*, vol. 49, no. 20, pp. 7406–7427, 2020.

[104] P. A. Konik, E. A. Berdonosova, I. M. Savvotin, and S. N. Klyamkin, "TheInfluence of Amide Solvents on Gas Sorption Properties of Metal-Organic Frameworks MIL-101 and ZIF-8," *Micropor. Mesopor. Mater.*, vol. 277, pp. 132–135, mar 2019.

[105] S. Velioglu and S. Keskin, "Revealing the Effect of Structure Curations on the Simulated CO_2 Separation Performances of MOFs," *Mater. Adv*, vol. 1, no. 3, pp. 341–353, 2020.

[106] O. Erhart, P. A. Georgiev, and H. Krautscheid, "Desolvation Process in the Flexible Metal-Organic Framework [Cu(Me-4py-trz-ia)], Adsorption of Dihydrogen and Related Structure Responses," *CrystEngComm*, vol. 21, no. 43, pp. 6523–6535, 2019.

[107] J. Ethiraj, F. Bonino, J. G. Vitillo, K. A. Lomachenko, C. Lamberti, H. Reinsch, K. P. Lillerud, and S. Bordiga, "Solvent-Driven Gate Opening in MOF-76-Ce: Effect on CO_2 Adsorption," *ChemSusChem*, vol. 9, pp. 713–719, apr 2016.

[108] A. Nandy, G. Terrones, N. Arunachalam, C. Duan, D. W. Kastner, and H. J. Kulik, "MOFSimplify, Machine Learning Models with Extracted Stability Data of Three Thousand Metal-Organic Frameworks," *Scientific Data*, vol. 9, p. 74, 2022.

[109] J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden, and T. K. Woo, "ARC-MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning," *Chem. Mater*, vol. 35, no. 3, pp. 900–916, 2023.

[110] J. Park, Y. Lim, S. Lee, and J. Kim, "Computational Design of Metal-Organic Frameworks with Unprecedented High Hydrogen Working Capacity and High Synthesizability," *Chem. Mater.*, vol. 35, no. 1, pp. 9–16, 2023.

[111] Y. Luo, S. Bag, O. Zaremba, A. Cierpka, J. Andreo, S. Wuttke, P. Friederich, and M. Tsotsalas, "MOF Synthesis Prediction Enabled by Automatic Data Mining and Machine Learning**," *Ang. Chem. Int. Ed.*, vol. 61, no. 19, p. e202200242, 2022.

[112] A. J. Fletcher, K. M. Thomas, and M. J. Rosseinsky, "Flexibility in Metal-Organic Framework Materials: Impact on Sorption Properties," *J. Solid State Chem.*, vol. 178, pp. 2491–2510, 2005.

[113] H. Jin and Y. Li, "Flexibility of Metal-Organic Frameworks for Separations: Utilization, Suppression and Regulation," *Curr. Opin. Chem. Eng.*, vol. 20, pp. 107–113, 2018.

[114] S. Vandenhaute, M. Cools-Ceuppens, S. DeKeyser, T. Verstraelen, and V. Van Speybroeck, "Machine Learning Potentials for Metal-Organic Frameworks using an Incremental Learning Approach," *Npj Comput. Mater.*, vol. 9, p. 19, feb 2023.

[115] F. Salles, A. Ghoufi, G. Maurin, R. G. Bell, C. Mellot-Draznieks, and G. Féry, "Molecular Dynamics Simulations of Breathing MOFs: Structural Transformations of MIL-53 upon Thermal Activation and CO_2 Adsorption," *Ang. Chem.*, vol. 47, no. 44, pp. 8321–8533, 2008.

[116] S. M. Rogge, R. Goeminne, R. Demuynck, J. J. Guitiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Modeling Gas Adsorption in Flexible Metal-Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes," *Adv. Theor. Simul.*, vol. 2, no. 4, p. 1800177, 2019.

[117] A. I. Kirkland and N. Young, "Advances in High-Resolution Transmission Electron Microscopy for Materials Science," *Microscopy and Analysis*, vol. 26, no. 6, pp. 19–24, 2012.

[118] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, "Electron Ptychography of 2D Materials to Deep sub-Ångström Resolution," *Nature*, vol. 559, no. 7714, pp. 343–349, 2018.

[119] Z. Chen, Y. Jiang, Y.-T. Shao, M. E. Holtz, M. Odstrcil, M. Guizar-Sicairos, I. Hanke, S. Ganschow, D. G. Schlom, and D. A. Muller, "Limits Set By Lattice Vibrations," *Science*, vol. 372, no. May, pp. 826–831, 2021.

[120] D. S. Su, B. Zhang, and R. Schlo, "Electron Microscopy of Solid Catalysts - Transforming from a Challenge to a Toolbox," *Chem. Rev.*, vol. 115, pp. 2818–2882, 2015.

[121] R. W. Lodge, G. A. Rance, M. W. Fay, and A. N. Khlobystov, "Movement of Palladium Nanoparticles in Hollow Graphitised Nanofibres: The Role of Migration and Coalescence in Nanocatalyst Sintering During the Suzuki-Miyaura Reaction," *Nanoscale*, vol. 10, no. 40, pp. 19046–19051, 2018.

[122] J. H. Park, N. M. Schneider, J. M. Grogan, M. C. Reuter, H. H. Bau, S. Kodambaka, and F. M. Ross, "Control of Electron Beam-Induced Au Nanocrystal Growth Kinetics through Solution Chemistry," *Nano Lett.*, vol. 15, no. 8, pp. 5314–5320, 2015.

[123] H. Zheng, U. M. Mirsaidov, L. W. Wang, and P. Matsudaira, "Electron Beam Manipulation of Nanoparticles," *Nano Lett.*, vol. 12, no. 11, pp. 5644–5648, 2012.

[124] J. Lyu, X. Gong, S. J. Lee, K. Gnanasekaran, X. Zhang, M. C. Wasson, X. Wang, P. Bai, X. Guo, N. C. Gianneschi, and O. K. Farha, "Phase Transitions in Metal-Organic Frameworks Directly Monitored through in Situ Variable Temperature Liquid-Cell Transmission Electron Microscopy and in Situ X-ray Diffraction," *Journal of the American Chemical Society*, vol. 142, no. 10, pp. 4609–4615, 2020.

[125] S. T. Skowron, T. W. Chamberlain, J. Biskupek, U. Kaiser, E. Besley, and A. N. Khlobystov, "Chemical Reactions of Molecules Promoted and Simultaneously Imaged by the Electron Beam in Transmission Electron Microscopy," *Acc. Chem. Res.*, vol. 50, pp. 1797–1807, aug 2017.

[126] Y. Miao, D. T. Lee, M. Dorneles de Mello, M. K. Abdel-Rahman, P. Corkery, J. A. Boscoboinik, D. H. Fairbrother, and M. Tsapatsis, "Electron beam induced modification of ZIF-8 membrane permeation properties," *Chem. Commun.*, vol. 57, pp. 5250–5253, may 2021.

[127] C. Wiktor, M. Meledina, S. Turner, O. I. Lebedev, and R. A. Fischer, "Transmission Electron Microscopy on Metal-Organic Frameworks-a Review," *J. Mater. Chem. A*, vol. 5, p. 14969, 2017.

[128] W. Kühlbrandt, "The Resolution Revolution," *Science*, vol. 343, no. 6178, pp. 1443–1444, 2014.

[129] D. Zhang, Y. Zhu, X. Ying, C.-e. Hsiung, R. Sougrat, K. Li, and Y. Han, "Atomic-Resolution Transmission Electron Microscopy of Electron Beam – Sensitive Crystalline Materials," *Mater. Sci.*, vol. 359, pp. 675–679, 2018.

[130] Y. Zhu, J. Ciston, B. Zheng, X. Miao, C. Czarnik, Y. Pan, and R. Sougrat, "Unravelling Surface and Interfacial Structures of a Metal-Organic Framework by Transmission Electron Microscopy," *Nat. Mater. Lett.*, vol. 16, no. February, pp. 4–9, 2017.

[131] T. W. Chamberlain, J. Biskupek, S. T. Skowron, P. A. Bayliss, E. Bichoutskaia, U. Kaiser, and A. N. Khlobystov, "Isotope Substitution Extends the Lifetime of Organic Molecules in Transmission Electron Microscopy," *Small*, vol. 11, pp. 622–629, feb 2015.

[132] S. T. Skowron, I. V. Lebedeva, A. M. Popov, and E. Bichoutskaia, "Approaches to Modelling Irradiation-Induced Processes in Transmission Electron Microscopy," *Nanoscale*, vol. 5, pp. 6677–6692, jul 2013.

[133] S. Kretschmer, T. Lehnert, U. Kaiser, and A. V. Krasheninnikov, "Formation of Defects in Two-Dimensional MoS_2 in the Transmission Electron Microscope at Electron Energies below the Knock-on Threshold: The Role of Electronic Excitations," *Nano Lett.*, vol. 20, pp. 2865–2870, 2020.

[134] K. Sing, "The Use of Nitrogen Adsorption for the Characterisation of Porous Materials," *Colloids Surf., A*, vol. 187-188, pp. 3–9, 2001.

[135] E. W. Dijkstra, "A Note on Two Problems in Connexion with Graphs," *Numer. Math.*, vol. 1, pp. 269–271, 1959.

[136] Dü)ren, Tina and Millange, Franck and Fé)rey, Gérard and Walton, Krista S. and Snurr, Randall Q., "Calculating Geometric Surface Areas as a Characterization Tool for Metal - Organic Frameworks," *J. Phys. Chem. C.*, vol. 111, pp. 15350–15356, oct 2007.

[137] D. Ongari, P. G. Boyd, S. Barthel, M. Witman, M. Haranczyk, and B. Smit, "Accurate Characterization of the Pore Volume in Microporous Crystalline Materials," *Langmuir*, vol. 33, pp. 14529–14538, dec 2017.

[138] S. Brunauer, P. H. Emmett, and E. Teller, "Adsorption of Gases in Multimolecular Layers," *J. Am. Chem. Soc.*, vol. 60, no. 2, pp. 309–319, 1938.

[139] J. W. Osterrieth, J. Rampersad, D. Madden, N. Rampal, L. Skoric, B. Connolly, M. D. Allendorf, V. Stavila, J. L. Snider, R. Ameloot, J. Marreiros, C. Ania, D. Azevedo, E. Vilarrasa-Garcia, B. F. Santos, X. H. Bu, Z. Chang, H. Bunzen, N. R. Champness, S. L. Griffin, B. Chen, R. B. Lin, B. Coasne, S. Cohen, J. C. Moreton, Y. J. Colón, L. Chen, R. Clowes, F. X. Coudert, Y. Cui, B. Hou, D. M. D'Alessandro, P. W. Doheny, M. Dincă, C. Sun, C. Doonan, M. T. Huxley, J. D. Evans, P. Falcaro, R. Ricco, O. Farha, K. B. Idrees, T. Islamoglu, P. Feng, H. Yang, R. S. Forgan, D. Bara, S. Furukawa, E. Sanchez, J. Gascon, S. Telalović, S. K. Ghosh, S. Mukherjee, M. R. Hill, M. M. Sadiq, P. Horcajada, P. Salcedo-Abraira, K. Kaneko, R. Kukobat, J. Kenvin, S. Keskin, S. Kitagawa, K. ichi Otake, R. P. Lively, S. J. DeWitt, P. Llewellyn,

B. V. Lotsch, S. T. Emmerling, A. M. Pütz, C. Martí-Gastaldo, N. M. Padial, J. García-Martínez, N. Linares, D. Maspoch, J. A. Suárez del Pino, P. Moghadam, R. Oktavian, R. E. Morris, P. S. Wheatley, J. Navarro, C. Petit, D. Danaci, M. J. Rosseinsky, A. P. Katsoulidis, M. Schröder, X. Han, S. Yang, C. Serre, G. Mouchaham, D. S. Sholl, R. Thyagarajan, D. Siderius, R. Q. Snurr, R. B. Goncalves, S. Telfer, S. J. Lee, V. P. Ting, J. L. Rowlandson, T. Uemura, T. Iiyuka, M. A. van der Veen, D. Rega, V. Van Speybroeck, S. M. Rogge, A. Lamaire, K. S. Walton, L. W. Bingel, S. Wuttke, J. Andreo, O. Yaghi, B. Zhang, C. T. Yavuz, T. S. Nguyen, F. Zamora, C. Montoro, H. Zhou, A. Kirchon, and D. Fairen-Jimenez, "How Reproducible are Surface Areas Calculated from the BET Equation?," *Adv. Mater.*, vol. 34, p. 2201502, 2022.

[140] H. Cox, R. L. Johnston, and J. N. Murrell, "Empirical Potentials for Modeling Solids, Surfaces, and Clusters," *Journal of Solid State Chemistry*, vol. 145, no. 2, pp. 517–540, 1999.

[141] M. S. Daw and M. I. Baskes, "Embedded-Atom Method: Derivation and Application to Impurities, Surfaces and Other Defects in Metals," *Phys. Rev. B*, vol. 29, no. 12, pp. 6443–6453, 1983.

[142] J. Cai and Y. Ye, "Simple Analytical Embedded-Atom-Potential Model Including a Long-Range Force for fcc Metals and Their Alloys," *Phys. Rev. B - Condens. Matter Mater Phys.*, vol. 54, pp. 8398–8410, sep 1996.

[143] G. U. Jeong and B. J. Lee, "Interatomic Potentials for Pt-C and Pd-C Systems and a Study of Structure-Adsorption Relationship in Large Pt/Graphene System," *Comp. Mater. Sci*, vol. 185, p. 109946, dec 2020.

[144] J. N. Murrell and R. E. Mottram, "Potential Energy Functions for Atomic Solids," *Mol. Phys.*, vol. 69, pp. 571–585, feb 1990.

[145] D. Dubbeldam, A. Torres-Knoop, and K. S. Walton, "On the Inner Workings of Monte Carlo Codes," *Mol. Simul.*, vol. 39, pp. 1253–1292, 2013.

[146] B. C. Yeo, D. Kim, H. Kim, and S. S. Han, "High-Throughput Screening to Investigate the Relationship between the Selectivity and Working Capacity of Porous Materials for Propylene/Propane Adsorptive Separation," *J. Phys. Chem. C*, vol. 120, pp. 24224–24230, oct 2016.

[147] T. M. Becker, L. C. Lin, D. Dubbeldam, and T. J. Vlugt, "Polarizable Force Field for CO2 in M-MOF-74 Derived from Quantum Mechanics," *J. Phys. Chem. C*, vol. 122, no. 42, pp. 24488–24498, 2018.

[148] T. M. Becker, J. Heinen, D. Dubbeldam, L. C. Lin, and T. J. Vlugt, "Polarizable Force Fields for CO2 and CH4 Adsorption in M-MOF-74," *J. Phys. Chem. C*, vol. 121, no. 8, pp. 4659–4673, 2017.

[149] D. Nocito and G. J. Beran, "Reduced Computational Cost of Polarizable Force Fields by a Modification of the Always Stable Predictor-Corrector," *J. Chem. Phys.*, vol. 150, p. 151103, 2019.

[150] M. J. Lennox, M. Bound, A. Henley, and E. Besley, "The Right Isotherms for the Right Reasons? Validation of Generic Force Fields for Prediction of Methane Adsorption in Metal-Organic Frameworks," *Mol. Simul.*, vol. 43, pp. 828–837, jul 2017.

[151] S. L. Mayo, B. D. Olafson, and W. A. G. Iii, "DREIDING: A Generic Force Field for Molecular Simulations," *J. Phys. Chem*, vol. 94, p. 91101, 1990.

[152] M. G. Martin and J. I. Siepmann, "Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes," *J. Phys. Chem. B*, vol. 102, no. 14, pp. 2569–2577, 1998.

[153] J. J. Potoff and J. I. Siepmann, "Vapor-Liquid Equilibria of Mixtures Containing Alkanes, Carbon Dioxide, and Nitrogen," *AIChE Journal*, vol. 47, pp. 1676–1682, jul 2001.

[154] R. S. Mulliken, "Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I," *J. Chem. Phys.*, vol. 23, no. 10, pp. 1833–1840, 1955.

[155] R. F. W. Bader, "Atoms in Molecules," *Acc. Chem. Res.*, vol. 18, pp. 9–15, 1985.

[156] F. L. Hirshfeld, "Bonded-Atom Fragments for Describing Molecular Charge Densities," *Theoretica Chimica Acta*, vol. 44, no. 2, pp. 129–138, 1977.

[157] T. C. Lillestolen and R. J. Wheatley, "Redefining the Atom: Atomic Charge Densities Produced by an Iterative Stockholder Approach," *Chem. Commun.*, vol. 7345, no. 45, pp. 5909–5911, 2008.

[158] T. A. Manz and D. S. Sholl, "Chemically Meaningful Atomic Charges That Reproduce the Electrostatic Potential in Periodic and Nonperiodic Materials," *J. Chem. Theory Comput.*, vol. 6, no. 8, pp. 2455–2468, 2010.

[159] C. Campa
simná, B. Mussard, and T. K. Woo, "Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional," *J. Chem. Theory Comput.*, vol. 5, pp. 2866–2878, 2009.

[160] A. K. Rappe and W. A. G. Iii, "Charge Equilibration for Molecular Dynamics Simulations," *J. Phys. Chem*, vol. 95, pp. 3358–3363, 1991.

[161] D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin, and B. Smit, "Evaluating Charge Equilibration Methods To Generate Electrostatic Fields in Nanoporous Materials," *J. Chem. Theory Comput.*, vol. 15, no. 1, p. 18, 2019.

[162] C. E. Wilmer, K. C. Kim, and R. Q. Snurr, "An Extended Charge Equilibration Method," *J. Phys. Chem. Lett.*, vol. 3, no. 17, pp. 2506–2511, 2012.

[163] J. Walker, D. Halliday, and R. Resnick, *Fundamentals of Physics*. Wiley, 2007.

[164] D. J. Adams, "On the use of the Ewald summation in computer simulation," *J. Chem. Phys*, vol. 78, no. 5, pp. 2585–2590, 1983.

[165] E. Schrödinger, "An Undulatory Theory of the Mechanics of Atoms and Molecules," *Phys. Rev.*, vol. 28, no. 6, pp. 1049–1070, 1926.

[166] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927.

[167] P. A. M. Dirac, "Quantum Mechanics of Many-Electron Systems," *Proc. R. soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character*, vol. 123, no. 792, pp. 714–733, 1929.

[168] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods," *Math. Proc. Cambridge Philos. Soc.*, vol. 24, no. 1, pp. 89–110, 1927.

[169] N. E. Schultz, Y. Zhao, and D. G. Truhlar, "Databases for Transition Element Bonding: Metal-Metal Bond Energies and Bond Lengths and Their Use To Test Hybrid, Hybrid Meta, and Meta Density Functionals and Generalized Gradient Approximations," *J. Phys. Chem. A.*, vol. 109, no. 19, pp. 4388–4403, 2005.

[170] N. E. Schultz, Y. Zhao, and D. G. Truhlar, "Density Functionals for Inorganometallic and Organometallic Chemistry," *J. Phys. Chem.*, vol. 109, no. 49, pp. 11127–11143, 2005.

[171] J. Sun, A. Ruzsinszky, and J. P. Perdew, "Strongly Constrained and Appropriately Normed Semilocal Density Functional," *Phys. Rev. Lett.*, vol. 115, p. 036402, 2015.

[172] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, no. 3, pp. 3865–3868, 1996.

[173] J. Tao and J. P. Perdew, "Climbing the Density Functional Ladder : Nonempirical Meta – Generalized Gradient Approximation Designed for Molecules and Solids," *Phys. Rev. Lett.*, vol. 91, no. 14, p. 146401, 2003.

[174] A. P. Bartók and J. R. Yates, "Regularized SCAN functional," *J. Chem Phys.*, vol. 150, p. 161101, apr 2019.

[175] N. Metropolis, A. W. Rosenblush, R. M. Rosenblush, and A. H. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys*, vol. 21, no. 6, pp. 1087–1092, 1953.

[176] D. A. Chesnut and Z. W. Salsburg, "Monte Carlo Procedure for Statistical Mechanical Calculations in a Grand Canonical Ensemble of Lattice Systems," *J. Chem. Phys*, vol. 38, p. 2861, 1963.

[177] F. Wang and D. P. Landau, "Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States," *Phys. Rev. Lett.*, vol. 86, p. 2050, 2001.

[178] S. Brandani, E. Mangano, and L. Sarkisov, "Net, Excess and Absolute Adsorption and Adsorption of Helium," *Adsorption*, vol. 22, no. 2, pp. 261–276, 2016.

[179] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal*, vol. 3, no. 3, pp. 535–554, 1959.

[180] M. Fernandez, T. K. Woo, C. E. Wilmer, and R. Q. Snurr, "Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks," *J. Phys. Chem. C*, vol. 117, pp. 7681–7689, apr 2013.

[181] F. Strieth-Kalthoff, F. Sandfort, M. H. Segler, and F. Glorius, "Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry," *Chem. Soc. Rev.*, vol. 49, no. 17, pp. 6154–6168, 2020.

[182] N. Artrith, K. T. Butler, F. X. Coudert, S. Han, O. Isayev, A. Jain, and A. Walsh, "Best Practices in Machine Learning for Chemistry," *Nat. Chem.*, vol. 13, no. 6, pp. 505–508, 2021.

[183] S. Barthel, E. V. Alexandrov, D. M. Proserpio, and B. Smit, "Distinguishing Metal - Organic Frameworks," *Cryst. Growth. Des.*, vol. 18, pp. 1738–1747, 2018.

[184] S. Majumdar, S. M. Moosavi, K. M. Jablonka, D. Ongari, and B. Smit, "Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening," *ACS Appl. Mater. Interfaces*, vol. 13, pp. 61004–61014, 2021.

[185] R. Gurnani, Z. Yu, C. Kim, D. S. Sholl, and R. Ramprasad, "Interpretable Machine Learning-Based Predictions of Methane Uptake Isotherms in Metal - Organic Frameworks," *Chem. Mater.*, vol. 33, pp. 3543–3552, 2021.

[186] M. Hirohara, Y. Saito, Y. Koda, K. Sato, and Y. Sakakibara, "Convolutional Neural Network Based on SMILES Representation of Compounds for Detecting Chemical Motif," *BMC Bioinform.*, vol. 19, p. 526, 2018.

[187] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit, and M. Haranczyk, "What Are the Best Materials to Separate a Xenon/Krypton Mixture?," *Chem. Mater.*, vol. 27, pp. 4459–4475, jun 2015.

[188] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[189] C. Altintas, O. F. Altundal, S. Keskin, and R. Yildirim, "Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation," *J. Chem. Inf. Model.*, vol. 61, pp. 2131–2146, 2021.

[190] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York: Springer, 2013.

[191] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. San Diego: Elsevier, 4 ed., 2009.

[192] F. G. Kerry, *Industrial Gas Handbook: Gas Separation and Purification. 1st ed.* Boca Raton, FL: CRC Press, 2007.

[193] D. Ianovski, K. Munakata, S. Kanjo, Y. Yokoyama, A. Koga, S. Yamatsuki, K. Tanaka, T. Fukumatsu, M. Nishikawa, and Y. Igarashi, "Adsorption of Noble Gases on H-Mordenite," *J. Nucl. Sci. Technol.*, vol. 39, no. 11, pp. 1213–1218, 2002.

[194] X. Feng, Z. Zong, S. K. Elsaidi, J. B. Jasinski, R. Krishna, P. K. Thallapally, and M. A. Carreon, "Kr/Xe Separation over a Chabazite Zeolite Membrane," *J. Am. Chem. Soc.*, vol. 138, pp. 9791–9794, aug 2016.

[195] P. K. Thallapally, J. W. Grate, and R. K. Motkuri, "Facile Xenon Capture and Release at Room Temperature using a Metal–Organic Framework: a Comparison with Activated Charcoal," *Chem. Commun.*, vol. 48, pp. 347–349, dec 2012.

[196] D. Banerjee, A. J. Cairns, J. Liu, R. K. Motkuri, S. K. Nune, C. A. Fernandez, R. Krishna, D. M. Strachan, and P. K. Thallapally, "Potential of Metal–Organic Frameworks for Separation of Xenon and Krypton," *Acc. Chem. Res.*, vol. 48, pp. 211–219, feb 2015.

[197] C. A. Fernandez, J. Liu, P. K. Thallapally, and D. M. Strachan, "Switching Kr/Xe Selectivity with Temperature in a Metal-Organic Framework," *J. Am. Chem. Soc.*, vol. 134, pp. 9046–9049, jun 2012.

[198] T. Wu, X. Feng, S. K. Elsaidi, P. K. Thallapally, and M. A. Carreon, "Zeolitic Imidazolate Framework-8 (ZIF-8) Membranes for Kr/Xe Separation," *Ind. Eng. Chem. Res.*, vol. 56, pp. 1682–1686, feb 2017.

[199] E. Ren and F.-X. Coudert, "Thermodynamic Exploration of Xenon/Krypton Separation Based on a High-Throughput Screening," *Faraday Discuss.*, vol. 231, pp. 201–223, 2021.

[200] S.-J. Lee, T.-U. Yoon, A.-R. Kim, S.-Y. Kim, K.-H. Cho, Y. K. Hwang, J.-W. Yeon, and Y.-S. Bae, "Adsorptive Separation of xenon/krypton Mixtures using a Zirconium-Based Metal-Organic Framework with High Hydrothermal and Radioactive Stabilities," *J. Hazard. Mater.*, vol. 320, pp. 513–520, dec 2016.

[201] K. S. Subrahmanyam, I. Spanopoulos, J. Chun, B. J. Riley, P. K. Thallapally, P. N. Trikalitis, and M. G. Kanatzidis, "Chalcogenide Aerogels as Sorbents for Noble Gases (Xe, Kr)," *ACS Appl. Mater. Interfaces*, vol. 9, pp. 33389–33394, oct 2017.

[202] F. F. Castellani, G. G. Curzio, and A. F. Gentil, "Krypton Diffusion in Granular Charcoal," *Anal. Chem.*, vol. 48, pp. 599–600, mar 1976.

[203] C. J. Jameson, A. K. Jameson, and H.-M. Lim, "Competitive Adsorption of Xenon and Krypton in Zeolite NaA: 129Xe Nuclear Magnetic Resonance Studies and Grand Canonical Monte Carlo Simulations," *J. Chem. Phys.*, vol. 104, p. 1709, jun 1996.

[204] S. M. Kuznicki, A. Ansón, A. Koenig, T. M. Kuznicki, T. Haastrup, E. M. Eyring, and D. Hunter, "Xenon Adsorption on Modified ETS-10," *J. Phys. Chem. C.*, vol. 111, pp. 1560–1562, feb 2007.

[205] G. Cruciani, "Zeolites upon Heating: Factors Governing their Thermal Stability and Structural Changes," *J. Phys. Chem. Solids*, vol. 67, pp. 1973–1994, sep 2006.

[206] R. S. Patil, D. Banerjee, C. M. Simon, J. L. Atwood, and P. K. Thallapally, "Noria: A Highly Xe-Selective Nanoporous Organic Solid," *Chem. Eur. J.*, vol. 22, no. 36, pp. 12618–12623, 2016.

[207] K. V. Lawler, Z. Hulvey, and P. M. Forster, "Nanoporous Metal Formates for Krypton/Xenon Separation," *Chem. Commun.*, vol. 49, pp. 10959–10961, oct 2013.

[208] O. Benson, I. da Silva, S. P. Argent, R. Cabot, M. Savage, H. G. Godfrey, Y. Yan, S. F. Parker, P. Manuel, M. J. Lennox, T. Mitra, T. L. Easun, W. Lewis, A. J. Blake, E. Besley, S. Yang, and M. Schröder, "Amides Do Not Always Work: Observation of Guest Binding in an Amide-Functionalized Porous Metal–Organic Framework," *J. Am. Chem. Soc*, vol. 138, pp. 14828–14831, nov 2016.

[209] Y. Yampolskii, I. Pinnau, and B. D. Freeman, *Materials Science of Membranes for Gas and Vapor Sepatation.* Chichester, UK: John Wiley and Sons, Ltd, 2006.

[210] H. A. Lorentz, "Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase," *Annalen der Physik*, vol. 248, pp. 127–136, jan 1881.

[211] M. K. Rana, H. S. Koh, H. Zuberi, and D. J. Siegel, "Methane Storage in Metal-Substituted Metal-Organic Frameworks: Thermodynamics, Usable Capacity, and the Impact of Enhanced Binding Sites," *J. Phys. Chem*, vol. 118, pp. 2929–2942, feb 2014.

[212] S. Brunauer, L. S. Deming, W. E. Deming, and E. Teller, "On a Theory of the van der Waals Adsorption of Gases," *J. Am. Chem. Soc*, vol. 62, no. 7, pp. 1723–1732, 1940.

[213] H. Demir and S. Keskin, "Computational Investigation of Multifunctional MOFs for Adsorption and Membrane-Based Separation of CF4/CH4, CH4/H2, CH4/N2, and N2/H2 Mixtures," *Mol. Sys. Des. Eng.*, vol. 7, no. 12, pp. 1707–1721, 2022.

[214] M. S. Shah, M. Tsapatsis, and J. I. Siepmann, "Identifying Optimal Zeolitic Sorbents for Sweetening of Highly Sour Natural Gas," *Angew. Chem. Int. Ed.*, vol. 55, pp. 5938–5942, may 2016.

[215] Z. Qiao, Q. Xu, and J. Jiang, "Computational Screening of Hydrophobic Metal–Organic Frameworks for the Separation of H_2S and CO_2 From Natural Gas," *J. Mater. Chem. A*, vol. 6, pp. 18898–18905, oct 2018.

[216] M. Rahimi, J. K. Singh, and F. Müller-Plathe, "Adsorption and Separation of Binary and Ternary Mixtures of SO_2, CO_2 and N_2 by Ordered Carbon Nanotube Arrays: Grand-Canonical Monte Carlo Simulations," *Phys. Chem. Chem. Phys.*, vol. 18, pp. 4112–4120, jan 2016.

[217] V. A. Solanki and B. Borah, "Exploring the Potentials of Metal-Organic Frameworks as Adsorbents and Membranes for Separation of Hexane Isomers," *J. Phys. Chem. C*, vol. 123, no. 29, pp. 17808–17822, 2019.

[218] J. Shang, A. Hanif, G. Li, G. Xiao, J. Zhe Lui, P. Xiao, and P. A. Webley, "Separation of CO_2 and CH_4 by Pressure Swing Adsorption Using a Molecular Trapdoor Chabazite Adsorbent for Natural Gas Purification," *Ing. Eng. Chem. Res*, vol. 59, no. 16, pp. 7857–7865, 2020.

[219] PG&E, "Pressure Swing Adsorption Technical Analysis [White paper]," 2018.

[220] D. Ko, R. Siriwardane, and L. T. Biegler, "Optimization of a Pressure-Swing Adsorption Process using Zeolite 13X for CO_2 Sequestration," *Ind. Eng. Chem. Res.*, vol. 42, no. 2, pp. 339–348, 2003.

[221] M. H. Mohamed, S. K. Elsaidi, T. Pham, K. A. Forrest, H. T. Schaef, A. Hogan, L. Wojtas, W. Xu, B. Space, M. J. Zaworotko, and P. K. Thallapally, "Hybrid Ultra-Microporous Materials for Selective Xenon Adsorption and Separation," *Angew. Chem. Int. Ed.*, vol. 55, pp. 8285–8289, jul 2016.

[222] J. Wieme, S. Vandenbrande, A. Lamaire, V. Kapil, L. Vanduyfhuys, and V. V. Speybroeck, "Thermal Engineering of Metal-Organic Frameworks for Adsorption Applications: A Molecular Simulation Perspective," *ACS Appl. Mater. Interfaces*, vol. 11, no. 42, pp. 38697–38707, 2019.

[223] L. Riboldi and O. Bolland, "Overview on Pressure Swing Adsorption (PSA) as CO_2 Capture Technology: State-of-the-Art, Limits and Potentials," *Energy Procedia*, vol. 114, pp. 2390–2400, jul 2017.

[224] J. Sánchez-Laínez, L. Paseta, M. Navarro, B. Zornoza, C. Téllez, and J. Coronas, "Ultrapermeable Thin Film ZIF-8/Polyamide Membrane for H2/CO2 Separation at High Temperature without Using Sweep Gas," *Adv. Mater. Interfaces*, vol. 5, p. 1800647, oct 2018.

[225] E. R. Van Selow, P. D. Cobden, P. A. Verbraeken, J. R. Hufton, and R. W. Van Den Brink, "Carbon Capture by Sorption-Enhanced Water-Gas Shift Reaction Process using Hydrotalcite-Based Material," *Ind. Eng. Chem. Res.*, vol. 48, pp. 4184–4193, may 2009.

[226] Y. Ban, N. Cao, and W. Yang, "Metal-Organic Framework Membranes and Membrane Reactors: Versatile Separations and Intensified Processes," *Research (Wash DC)*, vol. 2020, p. 1583451, may 2020.

[227] L. T. Glasby and P. Z. Moghadam, "Hydrogen Storage in MOFs: Machine Learning for Finding a Needle in a Haystack," *Patterns*, vol. 2, no. 7, p. 100305, 2021.

[228] M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji, and T. K. Woo, "Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO_2 Capture," *J. Phys. Chem. Lett.*, vol. 5, pp. 3056–3060, sep 2014.

[229] B. Bluemling, A. P. Mol, and Q. Tu, "The Social Organization of Agricultural Biogas Production and Use," *Energy Policy*, vol. 63, pp. 10–17, dec 2013.

[230] M. Struk, I. Kushkevych, and M. Vítězová, "Biogas Upgrading Methods: Recent Advancements and Emerging Technologies," *Rev Environ Sci Biotechnol*, vol. 19, pp. 651–671, 2020.

[231] S. F. Ahmed, ·. M. Mofijur, K. Tarannum, A. T. Chowdhury, N. Rafa, ·. Samiha Nuzhat, ·. P. Senthil Kumar, ·. Dai-Viet, N. Vo, E. Lichtfouse, and ·. T. M. I. Mahlia, "Biogas Upgrading, Economy and Utilization: a Review," *Environ. Chem. Lett.*, vol. 19, pp. 4137–4164, 2021.

[232] R. L. Canevesi, K. A. Andreassen, E. A. Da Silva, C. E. Borba, and C. A. Grande, "Pressure Swing Adsorption for Biogas Upgrading with Carbon Molecular Sieve," *Ind. Eng. Chem. Res.*, vol. 57, pp. 8057–8067, jun 2018.

[233] L. Jiang, R. Q. Wang, A. Gonzalez-Diaz, A. Smallbone, R. O. Lamidi, and A. P. Roskilly, "Comparative Analysis on Temperature Swing Adsorption Cycle for Carbon Capture by Using Internal Heat/Mass Recovery," *Appl. Therm. Eng.*, vol. 169, p. 114973, 2020.

[234] L. Chen, S. Deng, R. Zhao, Y. Zhu, L. Zhao, and S. Li, "Temperature Swing Adsorption for CO_2 Capture: Thermal Design and Management on Adsorption Bed with Single-Tube/Three-Tube Internal Heat Exchanger," *Appl. Therm. Eng.*, vol. 199, no. June, p. 117538, 2021.

[235] M. R. Ghadiri, J. R. Granja, R. A. Milligan, D. E. Mcree, N. Khazanovich, S. B. Copp, S. Subramanian, M. J. Zaworotko, O. M. Yaghi, G. Li, K.-M. Park, T. Iwamoto, B. F. Hoskins, R. Robson, Z. Sun, D. A. Richardson, T. L. Groy, J. Kim, D. Whang, J. I. Lee, and K. Kim, "Hydrothermal Synthesis of a Metal-Organic Framework Containing Large Rectangular Channels," *Angew. Chem., Int. Ed. Engl*, vol. 117, no. 4, pp. 1400–1402, 1995.

[236] R. Aniruddha, I. Sreedhar, and B. M. Reddy, "MOFs in Carbon Capture - Past, Present and Future," *J. CO2 Util*, vol. 42, p. 101297, dec 2020.

[237] D. A. Gómez-Gualdró, Y. J. Coló, X. Zhang, T. C. Wang, Y.-S. Chen, J. T. Hupp, T. Yildirim, O. K. Farha, J. Zhang, and R. Q. Snurr, "Evaluating Topologically Diverse Metal-Organic Frameworks for Cryo-Adsorbed Hydrogen Storage," *Energy Environ. Sci*, vol. 9, pp. 3279–3289, 2016.

[238] G. Avci, S. Velioglu, and S. Keskin, "High-Throughput Screening of MOF Adsorbents and Membranes for H_2 Purification and CO_2 Capture," *ACS Appl. Mater. Interfaces*, vol. 10, no. 39, pp. 33693–33706, 2018.

[239] S. Boobier, D. R. Hose, A. J. Blacker, and B. N. Nguyen, "Machine Learning with Physicochemical Relationships: Solubility Prediction in Organic Solvents and Water," *Nat. Commun.*, vol. 11, p. 5753, nov 2020.

[240] M. Suyetin, "The Application of Machine Learning for Predicting the Methane Uptake and Working Capacity of MOFs," *Faraday Discuss.*, vol. 231, pp. 224–234, 2021.

[241] Z. Shi, W. Yang, X. Deng, C. Cai, Y. Yan, H. Liang, Z. Liu, and Z. Qiao, "Machine-Learning-Assisted High-Throughput Computational Screening of High Performance Metal–Organic Frameworks," *Mol. Sys. Des. Eng.*, vol. 5, pp. 725–742, may 2020.

[242] X. Wu, S. Xiang, J. Su, and W. Cai, "Understanding Quantitative Relationship between Methane Storage Capacities and Characteristic Properties of Metal-Organic Frameworks Based on Machine Learning," *J. Phys. Chem. C*, vol. 123, pp. 8550–8559, 2019.

[243] J. Rugolo, C. Atkinson, D. Boysen, and E. Rohlfing, "Mehthane Opportunities for Vehicular Energy," 2012.

[244] Arpa, "MOVE Program Overview," 2012.

[245] Y. Peng, V. Krungleviciute, I. Eryazici, J. T. Hupp, O. K. Farha, and T. Yildirim, "Methane Storage in Metal-Organic Frameworks: Current Records, Surprise Findings, and Challenges," *J. Am. Chem. Soc.*, vol. 135, pp. 11887–11894, 2013.

[246] Z. Chen, P. Li, R. Anderson, X. Wang, X. Zhang, L. Robison, L. R. Redfern, S. Moribe, T. Islamoglu, D. A. Gómez-Gualdrón, T. Yildirim, J. F. Stoddart, and O. K. Farha, "Balancing Volumetric and Gravimetric Uptake in Highly Porous Materials for Clean Energy," *Science*, vol. 368, pp. 297–303, apr 2020.

[247] Y. Ye, R. B. Lin, H. Cui, A. Alsalme, W. Zhou, T. Yildirim, Z. Zhang, S. Xiang, and B. Chen, "A Microporous Metal–Organic Framework with Naphthalene Diimide Groups for High Methane Storage," *Dalton Trans.*, vol. 49, pp. 3658–3661, mar 2020.

[248] B. Wang, A. P. Côté, H. Furukawa, M. O'keeffe, and O. M. Yaghi, "Colossal Cages in Zeolitic Imidazolate Frameworks as Selective Carbon Dioxide Reservoirs," *Nature Lett.*, vol. 453, pp. 207–212, 2008.

[249] L. Pan, K. M. Adams, H. E. Hernandez, X. Wang, C. Zheng, Y. Hattori, and K. Kaneko, "Porous Lanthanide-Organic Frameworks: Synthesis, Characterization, and Unprecedented Gas Adsorption Properties," *J. Am. Chem. Soc.*, vol. 125, pp. 3062–3067, mar 2003.

[250] X. Zhou, W. Huang, J. Miao, Q. Xia, Z. Zhang, H. Wang, and Z. Li, "Enhanced Separation Performance of a Novel Composite Material GrO@MIL-101 for CO_2/CH_4 Binary Mixture," *Chem. Eng. J.*, vol. 266, pp. 339–344, apr 2015.

[251] A. I. Sarker, A. Aroonwilas, and A. Veawab, "Equilibrium and Kinetic Behaviour of CO2 Adsorption onto Zeolites, Carbon Molecular Sieve and Activated Carbons," *Energy Procedia*, vol. 114, pp. 2450–2459, jul 2017.

[252] Y. He, W. Zhou, G. Qian, and B. Chen, "Methane Storage in Metal-Organic Frameworks," *Chem. Soc. Rev*, vol. 43, pp. 5657–5678, 2014.

[253] P. Iacomi, F. Formalik, J. Marreiros, J. Shang, J. Rogacka, A. Mohmeyer, P. Behrens, R. Ameloot, B. Kuchta, and P. L. Llewellyn, "Role of Structural Defects in the Adsorption and Separation of C3 Hydrocarbons in Zr-Fumarate-MOF (MOF-801)," *Chem. Mater.*, vol. 31, no. 20, pp. 8413–8423, 2019.

[254] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, "The Cambridge Structural Database," *Acta. Crystallogr., Sect. B: Struct. Sci.*, vol. 72, pp. 171–179, apr 2016.

[255] T. Chen and T. A. Manz, "A Collection of Forcefield Precursors for Metal-Organic Frameworks," *RSC Adv.*, vol. 9, no. 63, pp. 36492–36507, 2019.

[256] V. Presser, J. McDonough, S. H. Yeon, and Y. Gogotsi, "Effect of Pore Size on Carbon Dioxide Sorption by Carbide Derived Carbon," *Energy Environ. Sci*, vol. 4, no. 8, pp. 3059–3066, 2011.

[257] E. Mahmoud, "Evolution of the Design of CH_4 Adsorbents," *Surfaces*, vol. 3, pp. 433–466, 2020.

[258] M. H. Mohamed, Y. Yang, L. Li, S. Zhang, J. P. Ru, A. G. Jarvi, S. Saxena, G. Veser, J. K. Johnson, and N. L. Rosi, "Designing Open Metal Sites in Metal - Organic Frameworks for Paraffin/Olefin Separations," *J. Am. Chem. Soc*, vol. 141, pp. 13003–13007, 2019.

[259] Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kuš, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C. M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. Distasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T. C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S. P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. Oneill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small,

A. Sodt, T. Stein, D. Stück, Y. C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z. Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J. D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C. P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill, and M. Head-Gordon, "Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package," *Mol. Phys.*, vol. 113, pp. 184–215, jan 2015.

[260] R. Ditchfield, W. J. Hehre, and J. A. Pople, "Self — Consistent Molecular Orbital Methods . XII . Further Extensions of Gaussian — Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules," *J. Chem. Phys.*, vol. 54, pp. 724–728, 1971.

[261] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu," *J. Chem. Phys.*, vol. 132, p. 154104, 2010.

[262] M. Hmadeh, Z. Lu, Z. Liu, F. Gándara, H. Furukawa, S. Wan, V. Augustyn, R. Chang, L. Liao, F. Zhou, E. Perre, V. Ozolins, K. Suenaga, X. Duan, B. Dunn, Y. Yamamto, O. Terasaki, and O. M. Yaghi, "New Porous Crystals of Extended Metal-Catecholates," *Chem. Mater*, vol. 24, p. 19, 2012.

[263] L. Zhu, D. Zhang, M. Xue, H. Li, and S. Qiu, "Direct Observations of the MOF (UiO-66) Structure by Transmission Electron Microscopy," *CrystEngComm*, vol. 15, p. 9356, 2013.

[264] D. Xu, D. Zhang, H. Zou, L. Zhu, M. Xue, Q. Fang, and S. Qiu, "Guidance From an in situ Hot Stage in TEM to Synthesize Magnetic Metal Nanoparticles from a MOF," *Chem. Commun*, vol. 52, p. 10513, 2016.

[265] R. F. Egerton, P. Li, and M. Malac, "Radiation Damage in the TEM and SEM," *Micron*, vol. 35, pp. 399–409, 2004.

[266] O. Ugurlu, J. Haus, A. A. Gunawan, M. G. Thomas, S. Maheshwari, M. Tsapatsis, and K. A. Mkhoyan, "Radiolysis to Knock-On Damage Transition in Zeolites under Electron Beam Irradiation," *Phys. Rev. B*, vol. 83, p. 113408, 2011.

[267] C. Speckmann, J. Lang, J. Madsen, M. Reza, A. Monazam, G. Zagler, G. T. Leuthner, N. Mcevoy, C. Mangler, T. Susi, and J. Kotakoski, "Combined Electronic Excitation and Knock-On Damage in Monolayer MoS_2," *Phys. Rev. B*, vol. 107, p. 094112, 2023.

[268] A. Santana, A. Zobelli, J. Kotakoski, A. Chuvilin, and E. Bichoutskaia, "Inclusion of Radiation Damage Dynamics in High-Resolution Transmission Electron Microscopy Image Simulations: The Example of Graphene," *Phys. Rev. B - Condens. Matter Mater Phys.*, vol. 87, p. 094110, mar 2013.

[269] T. Susi, J. C. Meyer, and J. Kotakoski, "Quantifying Electron Irradiation Effects in Transmission Electron Microscopy," *Nat. Rev. Phys.*, vol. 1, pp. 397–405, 2019.

[270] A. Yoshimura, M. Lamparski, J. Giedt, D. Lingerfelt, J. Jakowski, P. Ganesh, T. Yu, B. G. Sumpter, and V. Meunier, "Quantum Theory of Electronic Excitation and Sputtering by Transmission Electron Microscopy," *Nanoscale*, vol. 15, pp. 1053–1067, 2022.

[271] S. T. Skowron, S. L. Roberts, A. N. Khlobystov, and E. Besley, "The Effects of Encapsulation on Damage to Molecules by Electron Radiation," *Micron*, vol. 120, pp. 96–103, may 2019.

[272] F. Saiz and L. Bernasconi, "Density-Functional Theory Models of Fe(iv)O Reactivity in Metal-Organic Frameworks: Self-Interaction Error, Spin Delocalisation and the Role of Hybrid Exchange," *Phys. Chem. Chem. Phys*, vol. 22, no. 22, pp. 12821–12830, 2020.

[273] Z. Cai, S. Chen, and L. W. Wang, "Dissociation Path Competition of Radiolysis Ionization-Induced Molecule Damage under Electron Beam Illumination," *Chem. Sci.*, vol. 10, no. 46, pp. 10706–10715, 2019.

[274] F. Neese, "The ORCA Program System," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 2, pp. 73–78, jan 2012.

[275] F. Neese, "Software Update: the ORCA Program System, Version 4.0," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 8, p. 1327, jan 2018.

[276] F. Neese, "Software Update: The ORCA Program System—Version 5.0," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 12, p. 1606, sep 2022.

[277] K. E. Riley, B. T. Op't Holt, and K. M. Merz, "Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties," *J. Chem. Theory Comput.*, vol. 3, no. 2, pp. 407–433, 2007.

[278] K. E. Riley and K. M. Merz, "Assessment of Density Functional Theory Methods for the Computation of Heats of Formation and Ionization Potentials of Systems Containing Third Row Transition Metals," *J. Phys. Chem.*, vol. 111, no. 27, pp. 6044–6053, 2007.

[279] D. Farmanzadeh and ·. T. Abdollahi, "Investigation on the Chemical Active Sites of Copper Nanoclusters as Nanocatalyst for the Adsorption of Acetylene: Calibration of DFT Method and Basis Set," *Theor Chem Acc*, vol. 135, p. 49, 2016.

[280] Y. Yang, M. N. Weaver, and K. M. Merz, "Assessment of the "6-31+G** + LANL2DZ" Mixed Basis Set Coupled with Density Functional Theory Methods and the Effective Core Potential: Prediction of Heats of Formation and Ionization Potentials for First-Row-Transition-Metal Complexes," *J. Phys. Chem. A*, vol. 113, no. 36, pp. 9843–9851, 2009.

[281] W. A. McKinley and H. Feshbach, "The Coulomb Scattering of Relativistic Electrons by Nuclei," *Physical Review*, vol. 74, p. 1759, dec 1948.

[282] J. C. Meyer, F. Eder, S. Kurasch, V. Skakalova, J. Kotakoski, H. Jin Park, S. Roth, A. Chuvilin, S. Eyhusen, G. Benner, A. V. Krasheninnikov, and U. Kaiser, "Accurate Measurement of Electron Beam Induced Displacement Cross Sections for Single-Layer Graphene," *Phys. Rev. Lett.*, vol. 108, no. 19, p. 239902, 2012.

[283] T. W. Chamberlain, J. Biskupek, S. T. Skowron, A. V. Markevich, S. Kurasch, O. Reimer, K. E. Walker, G. A. Rance, X. Feng, K. Müllen, A. Turchanin, M. A. Lebedeva, A. G. Majouga, V. G. Nenajdenko, U. Kaiser, E. Besley, and A. N. Khlobystov, "Stop-Frame Filming and Discovery of Reactions at the Single-Molecule Level by Transmission Electron Microscopy," *ACS Nano*, vol. 11, pp. 2509–2520, mar 2017.

[284] S. Malola and H. Häkkinen, "Prospects and Challenges for Computer Simulations of Monolayer-Protected Metal Clusters," *Nat. Commun.*, vol. 12, p. 2197, 2021.

[285] A. Chuvilin, U. Kaiser, E. Bichoutskaia, N. A. Besley, and A. N. Khlobystov, "Direct Transformation of Graphene to Fullerene," *Nat. Chem.*, vol. 2, pp. 450–453, may 2010.

[286] A. Markevich, S. Kurasch, O. Lehtinen, O. Reimer, X. Feng, K. Müllen, A. Turchanin, A. N. Khlobystov, U. Kaiser, and E. Besley, "Electron Beam Controlled Covalent Attachment of Small Organic Molecules to Graphene," *Nanoscale*, vol. 8, p. 2711, 2016.

[287] H. Fang, H. Demir, P. Kamakoti, and D. S. Sholl, "Recent Developments in First-Principles Force Fields for Molecules in Nanoporous Materials," *J. Mater. Chem. A*, vol. 2, pp. 274–291, 2014.

[288] S. P. Gedam, S. Chiriki, and D. A. Padmavathi, "Advanced Machine Learning Based Global Optimizations for Pt Nanoclusters," *J. Ind. Chem. Soc.*, vol. 100, p. 100978, may 2023.

[289] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The Amber Biomolecular Simulation Programs," *J. Comput. Chem.*, vol. 26, pp. 1668–1688, dec 2005.

[290] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[291] J. Heinen and D. Dubbeldam, "On Flexible Force Fields for Metal–Organic Frameworks: Recent Developments and Future Prospects," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 8, p. 1363, jul 2018.

[292] A. P. Sutton and J. Chen, "Long-Range Finnis-Sinclair Potentials," *Philos. Mag. Lett.*, vol. 61, no. 3, pp. 139–146, 1990.

[293] B. D. Todd and R. M. Lynden-Bell, "Surface and Bulk Properties of Metals Modelled with Sutton-Chen Potentials," *Surf. Sci.*, vol. 281, pp. 191–206, jan 1993.

[294] M. I. Baskes, "Modified Embdedded Atom Potentials for Cubic Materials and Impurities," *Phys. Rev. B*, vol. 46, no. 5, pp. 2727–2742, 1992.

[295] B. J. Lee, J. H. Shim, and I. Baskes, "Semiempirical Atomic Aotentials for the fcc Metals Cu, Ag, Au, Ni, Pd, Pt, Al, and Pb Based on First and Second Nearest-Neighbor Modified Embedded Atom Method," *Phys. Rev. B - Condens. Matter Mater Phys.*, vol. 68, p. 144112, oct 2003.

[296] F. Cleri and V. Rosato, "Tight-Binding Potentials for Transition Metals and Alloys," *Phys. Rev. B*, vol. 48, no. 1, pp. 22–23, 1993.

[297] L. D. Lloyd, R. L. Johnston, S. Salhi, and N. T. Wilson, "Theoretical Investigation of Isomer Stability in Platinum-Palladium Nanoalloy Clusters," *J. Mater. Chem.*, vol. 14, pp. 1691–1704, jun 2004.

[298] G. D. Förster, Y. Magnin, F. Rabiloud, and F. Calvo, "Effective Embedded-Atom Potential for Metallic Adsorbates on Crystalline Surfaces," *Model. Simul. Mat. Sci. Eng.*, vol. 22, p. 035015, 2014.

[299] M. Ruda and D. Farkas, "Embedded-Atom Interatomic Potentials for Hydrogen in Metals and Intermetallic Alloys," *Phys. Rev. B.*, vol. 54, no. 14, pp. 9765–9774, 1996.

[300] L. Vitos, A. V. Ruban, H. L. Skriver, and J. Kollár, "The Surface Energy of Metals," *Surf. Sci.*, vol. 411, pp. 186–202, aug 1998.

[301] J. L. Da Silva, C. Stampfl, and M. Scheffler, "Converged Properties of Clean Metal Surfaces by All-Electron First-Principles Calculations," *Surf. Sci.*, vol. 600, pp. 703–715, feb 2006.

[302] I. Galanakis, N. Papanikolaou, and P. H. Dederichs, "Applicability of the Broken-Bond Rule to the Surface Energy of the fcc Metals," *Surf. Sci.*, vol. 511, pp. 1–12, jun 2002.

[303] S. H. Yoo, J. H. Lee, Y. K. Jung, and A. Soon, "Exploring Stereographic Surface Energy Maps of Cubic Metals via an Effective Pair-Potential Approach," *Phys. Rev. B*, vol. 93, p. 035434, jan 2016.

[304] H. L. Skriver and N. M. Rosengaard, "Surface Energy and Work Function of Elemental Metals," *Phys. Rev. B.*, vol. 46, pp. 15–1992, 1992.

[305] W. R. Tyson and W. A. Miller, "Surface Free Energies of Solid Metals: Estimation from Liquid Surface Tension Measurements," *Surf. Sci.*, vol. 62, pp. 267–276, jan 1977.

[306] B. J. Keene, "Review of Data for the Surface Tension of Pure Metals," *Int. Mat. Rev.*, vol. 38, no. 4, pp. 157–192, 1993.

[307] J. Gonnissen, A. De Backer, A. J. den Dekker, J. Sijbers, and S. Van Aert, "Atom-Counting in High Resolution Electron Microscopy: TEM or STEM – That's the Question," *Ultramicroscopy*, vol. 174, pp. 112–120, mar 2017.