Computational Study of Hydrogen Uptake in Metal Organic Framework $Ni_3(pzdc)_2(ade)_2$ and Collecting Geometric Properties of Metal Organic Frameworks to use as Descriptors in a Computational Database

Scott Simrod

A thesis submitted in partial fulfillment of the requirements for the
Master's degree in Chemistry

Department of Chemistry
Faculty of Science
University of Ottawa

Table of Contents

Abstract

Metal organic frameworks (MOFs) are a class of nanoporous materials made of organic and inorganic structural building units (SBUs). MOFs have gathered interest for many applications due to their high tunability leading to materials with very diverse pore chemistries and geometries. The focus of this thesis is on studying the gas separation and storage capabilities of MOFs and analysing their geometric properties.

The first part of this thesis work is a computational study of H_2 uptake in the MOF $Ni_3(pzdc)_2(ade)_2$ to understand its strong uptake. Synthesized in the lab of Professor Kyriakos Stylianou at the University of Oregon, the unactivated crystal structure of the MOF was experimentally determined and following the activation the MOF was found to have high uptake for H_2 at low pressures. To understand the adsorption of H_2 in the MOF, density functional theory (DFT) calculations were performed to explore the activated structure and understand the H_2 binding in the material. The H_2 uptake capacity was then explored computationally with grand canonical Monte Carlo (GCMC) simulations whereby the metal was changed to analyze the effects of the metal centre on hydrogen uptake.

The second part of this thesis involved the calculation of geometric properties for the ARC-MOF database curated with the Woo lab that contains around 280K MOFs – both experimentally characterized and computer-generated structures. The database was created to provide the community with a diverse database of MOFs for which DFT derived partial atomic charges are available to conveniently perform high-throughput screening and data-driven studies. For this thesis, the relevant geometric properties for all MOFs in the database were calculated such as the accessible surface area, accessible volume, pore limiting diameter, and largest cavity diameter. The distribution of the descriptors was plotted and compared to existing databases to demonstrate the diversity of the database.

List of Figures

List of Tables

Statement of Contribution

The experimental work performed in chapter 3 was done by Nan Chieh Chiu, a student from the Materials Discovery Laboratory at Oregon State University supervised by Dr. Kyriakos C. Stylianou. This includes chemical synthesis, X-ray diffraction, Fourier transform infrared spectroscopy, and gas adsorption-desorption isotherms. The diversity analysis in chapter 4 was done by Jake Burner, a student of the Woo lab group from the University of Ottawa.

List of Acronyms

AP-RDF	Atomic Property-weighted Radial Distribution Function
ASR	All Solvent Removed
CCS	Carbon Capture and Storage
CoRE	Computational Ready Experimental
CSD	Cambridge Structural Database
DFT	Density Functional Theory
ESP	Electrostatic Potential
FSR	Free Solvent Removed
GCMC	Grand Canonical Monte Carlo
GHG	Greenhouse Gases
MAE	Monoethanolamine
MC	Monte Carlo
ML	Machine Learning
MOFs	Metal Organic Frameworks
PAW	Projector Augmented Wave
PBC	Periodic Boundary Conditions
PSA	Pressure Swing Adsorption
PXRD	Powder X-Ray Diffraction
QMOF	Quantum Metal Organic Frameworks
QSPR	Quantitative Structure-Property Relationship
RAC	Revised Auto-Correlation
REPEAT	Repeating Electrostatic Potential Extracted Atomic charge
SBU	Structural Binding Unit
SCXRD	Single Crystal X-Ray Diffraction
ToBasCCo	Topology Based Crystal Constructor
TPSA	Temperature-Pressure Swing Adsorption
TSA	Temperature Swing Adsorption
VASP	Vienna *Ab Initio* Simulation Package

1 Introduction

1.1 Importance of Clean Fuel Sources

Over the past decade, it has become apparent that changes to the environment are occurring due to climate change caused by human pollution. This is due to the greenhouse effect causing temperature increase in the planet's atmosphere and has begun to show effects on plant and animal life across the world.[1,2] The increase in temperature and subsequent increase in sea level[3] will drastically affect the way humans are able to live on the planet. The need to reduce greenhouse gas emissions in an economical way has remained a problem and has not improved as greenhouse gas emissions continue to increase yearly.[4,5] When these emissions are divided by sector, energy production is responsible for the majority of CO_2 and agriculture produces the most CH_4 and N_2O.[4] Many places around the world are beginning to adopt energy generation from sources that do not produce greenhouse gases such as nuclear plants or renewable power (solar, wind, etc.), however these sites cannot be built in certain regions leading to the continued reliance on green house gas emitting fuels for stationary energy production.

Carbon capture and storage (CCS) represents one of the many ways to reduce greenhouse gas emissions while we transition to zero emission energy generation. CCS refers to methods that aims to reduce greenhouse gas emissions by collecting it before release into the atmosphere. This collected material can be permanently stored in underground reservoirs in order to prevent release into the atmosphere.[6] Though storage is considered a current practical option, the collected gases could also be utilized or transformed into other chemicals. However, no large-scale utilization practices are currently available that make significant reductions in our greenhouse gas emissions. CO_2 capture from large stationary sources such as a powerplant is currently viewed as the most practical form of large-scale CCS. Depending on what fuel the power plant uses and how it is run can result in different gases and concentrations of gases being produced thus they need different methods to be collected. Greenhouse gas sources can also appear in the productions of certain products such as steel[7] and cement[8] which produces CO_2 that need to be collected as well.

There are different methods for energy generation when burning fossil fuels with their own carbon capture methods, the three worth pursuing are pre-combustion, post-combustion and oxy-fuel combustion carbon capture.[9,10] Pre-combustion involves placing the coal in an environment of steam and oxygen at high temperature and pressure to create a gas of carbon monoxide and hydrogen known as syngas. The high temperatures and pressures causes a drop in powerplant efficiency of around 20% compared to traditional energy production methods.[9] The carbon monoxide is converted to carbon dioxide by reacting with water and is separated from the hydrogen, where the hydrogen is then burned for energy.[11] Oxy-fuel burns coal in oxygen resulting in an output of only CO_2 and H_2O which can be separated by condensation. These methods both require fundamental process changes and cannot be used in existing powerplants.

The last method, post-combustion carbon capture, is considered the most general due to the ability to easily retrofit onto an existing plant. Post-combustion involves separating CO_2 from the exhaust gas of a power plant. Exhaust gas composition depends on the energy production method. The current large scale method for post-combustion CO_2 capture uses an aqueous amine based process which can collect CO_2 at low concentrations and be retrofit onto existing power plants. After CO_2 is collected by the solution, heat needs to be applied to recover the CO_2 for permanent storage. This utilizes more than 30% of energy generated from the power plant.[12] Moreover, the solutions are highly corrosive and the amines chemically degrade which requires continual replacement of the sorbent.[13,14]

The Boundary dam power plant in Saskatchewan is the first large scale coal burning plant to use CCS technologies. One of the four boilers was retrofitted with a liquid amine capture system that started in June 2014.[15] Due to the aforementioned problems with the liquid amine capture systems, the CO_2 capture has not been expanded and the system is being decomissioned.[16] In Texas, the Petra Nova plant started liquid amine based carbon capture in 2017 with the goal of capturing 4.6 million tons of carbon dioxide. The plant was shut down in 2020 with too many outages over the time period it was open. Additionally, it failed to meet its carbon capture goals by 17 %.[17]

Alternatives to liquid amine based CO_2 capture are technologies that use porous solids. Porous solids based technologies could have a lower energy cost of CO_2 separation due to the lower binding energy to CO_2 than amines, which chemically react, and the lower heat capacity of the solid sorbent compared to the aqueous amine. Some of these technologies that use porous solids are also quite mature. For example, large scale pressure swing adsorption (PSA) systems have been used for many decades to remove CO_2 from natural gas.[18] With PSA, the CO_2 is adsorbed into the porous solid and then the near pure CO_2 is recovered by applying a vacuum (swinging the pressure). Other methods to recover pure CO_2 that are more energy efficient can also be developed, such as steam swing, temperature swing adsorption or combinations thereof. However, such technologies are not as mature and would require further development.

1.1.1 Hydrogen as a Fuel Source

Though CCS is a promising method for GHG reduction, clean energy sources are still explored as alternatives. There has been significant research into alternative energy sources that reduce greenhouse gas emissions but many lack the practicality that fossil fuels have.[19] Though there have been large improvements in stationary energy generation to move away from sources that produce greenhouse gases, providing clean energy to the transportation sector is still an ongoing problem.[20] Fossil fuels remain the most practical method to generate energy for transportation as high volumetric energy density as shown in Table 1.1. Though the fuel alternatives feature much higher gravimetric energy densities, at standard temperature and pressure they exist as gases with low volumetric energy density.

Table 1.1: Volumetric and gravimetric energy densities of common fuels in different states[19]

Material	Gravimetric energy density (MJ/kg)	Volumetric energy density (MJ/L) (liquid)	Volumetric energy density (MJ/L) (gas)	Volumetric energy density (MJ/L) (compressed)
Hydrogen	143	10.1	0.0107	5.6 (700 bar)
Methane	55.6	-	0.0378	-
Natural gas	53.6	22.2	0.0364	9 (250 bar)
Gasoline	46.4	34.2	-	-
Biodiesel oil	42.2	33	-	-

One alternative fuel that is known to produce no greenhouse gases is molecular hydrogen (H_2). Hydrogen is a combustible gas that reacts with oxygen to produce only water and has a gravimetric energy density of 143 MJ/kg which is significantly higher than most gases that are used as fuel.[19] There are still issues with hydrogen as a fuel that prevents it's widespread adoption such as low volumetric density[21] and difficulty to produce it without greenhouse gas production.[22]

Low density of hydrogen as a gas requires compression to meet the same standard as other fuels that produce greenhouse gases but can be stored in smaller volumes. Despite hydrogen storing 2.6 times the amount of energy of gasoline per kilogram, it takes up four times the volume of gasoline[22] as shown in Table 1.1. Similar to other gases, hydrogen can be stored in pressurized containers which is the current method of storage in current hydrogen vehicles.[23,24] Most compressed storage vessels need heavy materials to ensure safety therefore research is being done to create light weight materials that can densely store gas for fuel efficiency.[23,25] There are alternatives which can bypass the problems above such as producing the hydrogen before use or collecting the hydrogen in materials capable of densifying the gas. A comparison of large scale hydrogen storage[26] shows that metal-hydride, liquid, and cryo-adsorbers storage take up much less volume than pressurized containers of hydrogen but have other costs preventing them from being widely deployed. Metal hydrides are the heaviest[27] storage solution causing excess energy use during transportation and requires heat for hydrogen release.[24] Hydrogen needs to remain below 33.2 K to stay as a liquid which has a large energetic cost to cool below this temperature and it needs to be stored in an open system to prevent over pressure.[28,29] There are some methods that meet the Department of Energy's (DOE) specific targets[30] seen in Table 1.2 for hydrogen storage but many come with issues mentioned above preventing the widespread adoption of hydrogen as a transportation fuel.

Table 1.2: US DOE onboard hydrogen storage targets in light-duty vehicles[28]

Storage	2020	2025	Target
Gravimetric Capacity (wt %)	4.5	5.5	6.5
Volumetric Capacity (g/L)	30	40	50
Cost (\$/kWh) (\$/kg)			10

1.2 MOFs

The use of porous solids for gas separation and storage has been explored with many classes of materials but, metal organic frameworks (MOFs) provide a unique platform to develop materials.

A MOF is made of a repeating series of organic and inorganic structural building units (SBUs), which form covalent or dative bonds with one another.[31] MOFs are considered highly tunable as different SBUs can be combined for intentional design of pores to interact with a specific gas. MOF UiO-66 is presented in Figure 1.1 (c) as a popular example with an easy to view structure, composed of the organic linker benzene dicarboxylate (b) and the inorganic linker $Zr_6O_4(OH)_4$ (a). MOFs have been presented for many different applications such as gas selectivity and storage,[32–34] batteries,[35,36] drug delivery,[37,38] and water harvesting.[39,40]

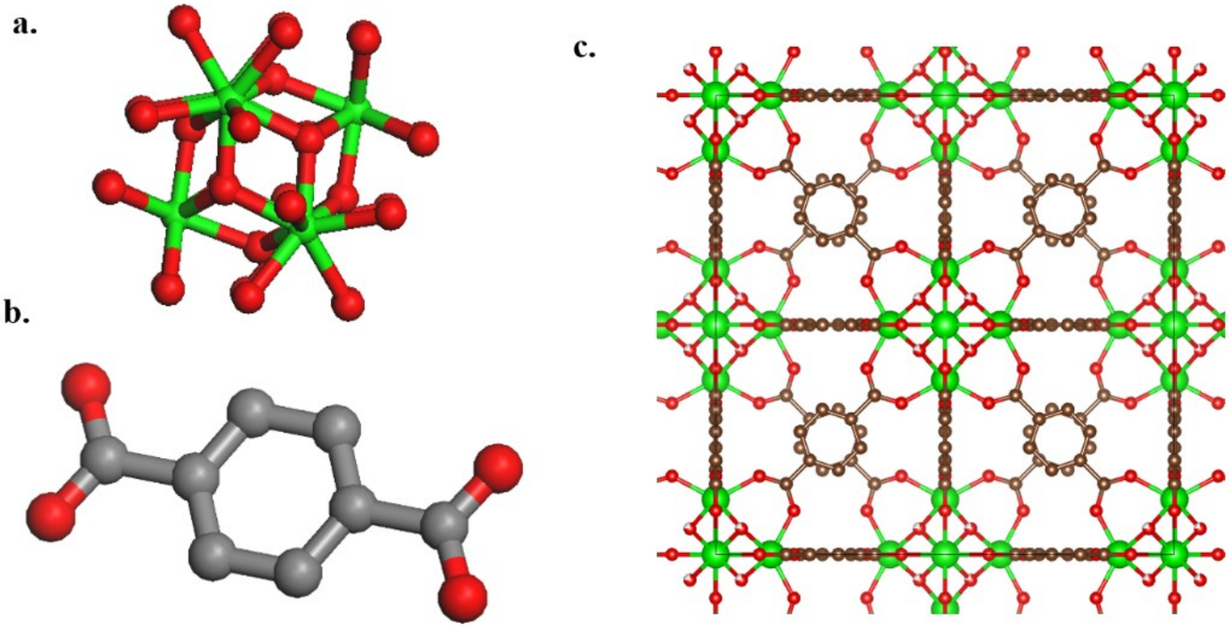

Figure 1.1: UiO-66 with (a) $Zr_6O_4(OH)_4$ as the inorganic cluster, (b) benzene dicarboxylate (bdc^{2-}) is the organic linker and (c) as the completed structure (unprotonated) where black lines represent the repeating unit cell. Red are oxygen atoms, grey are carbon atoms and green are zirconium atoms.

1.2.1 MOF Geometric Properties

MOFs can have small pore sizes and large pore volumes and surface areas which allows for gases to enter the material and interact with its internal structure. The largest reported surface area of a MOF is >7000 m^2/g and the largest theoretical surface area is ~14600 m^2/g.[41] It is important to be able to characterize the pores of a MOF in ways that are useful to a researcher. The relevant geometric properties that are usually calculated for a MOF are surface area, volume, largest cavity diameter, pore limiting diameter, void fraction and density. To determine the structure of the pores, a probe, i.e. He, N_2, H_2, travels across the surface of the MOFs structure drawing a line from the centre of the probe or the edge of the particle.[42] Everything that exists between the lines drawn are the pore cavities seen in blue on Figure 1.2. In a MOF, the surface area refers to the accessible surface that a given molecule can interact with inside of a pore generated by the probe and can either be reported gravimetrically (m^2/g) or volumetrically (m^2/cm^3). The MOF's accessible volume is then the volume that exists in these pores created by the probe and can be reported gravimetrically (cm^3/g) or as the void fraction, a ratio of the pore volume to the volume of the unit cell. There are two relevant pores to look at which are the pore limiting diameter and the largest cavity diameter. The pore limiting diameter determines what guests can enter a MOF and which can not and the largest cavity diameter is important for understanding diffusion rates of gas in a MOF.[43] These values are calculated because geometric properties can be used to predict uptake.[44–46]

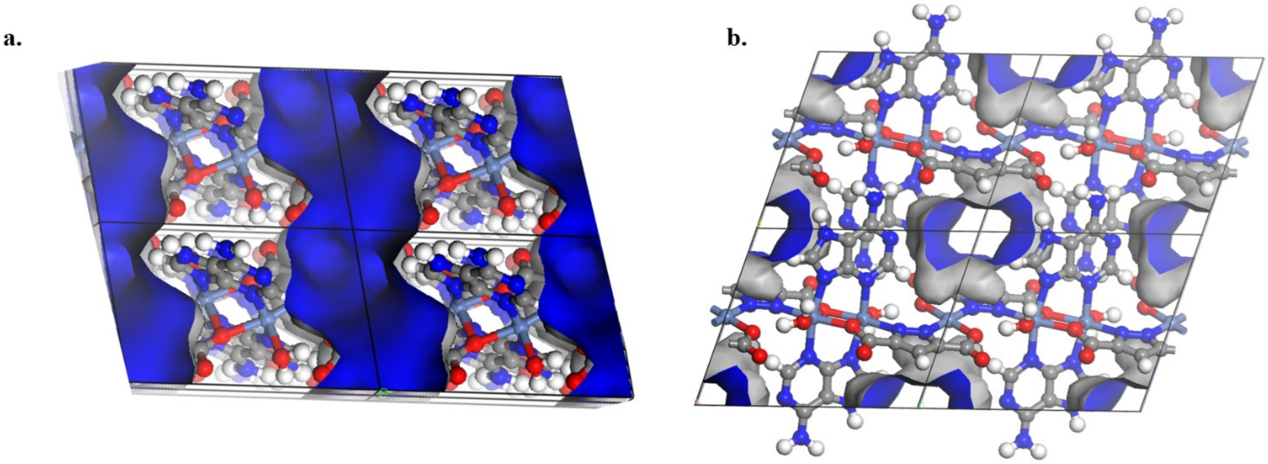

Figure 1.2: The internal surface of a MOFs pores where the blue walls are inside the pore and gray walls are outside with a view from (a) the y axis and (b) the x axis.

1.2.2 H_2 Storage in MOFs

Before MOFs came into the picture, zeolites and porous carbons were studied for hydrogen storage. This led to the discovery that there is a relation between accessible surface area and storage capacity[47] in porous materials. Due to this relation MOFs have become promising candidates for hydrogen storage due to their large and tunable pore size and surface areas. The first MOF for hydrogen storage was demonstrated on MOF-5 which possesses a large internal surface area, 2500 – 3000 m^2/g, and it was found that the maximum uptake was 4.5 wt % at 1 bar (weight percent, mass of hydrogen vs mass of MOF material) at 77 K.[48] This was later corrected to 1.3 wt % at 1 bar[49,50] More porous solids of note with experimental results are MOF-74 (Ni)[51] with a storage capacity of 1.8 wt % and HKUST-1[52] with an adsorption capacity of 2.6 wt %. Further research into hydrogen uptake in other MOFs lead to these conclusions; Hydrogen uptake has a linear correlation to accessible surface area (Chahine's rule: ~ 1 wt % per 500 m^2/g),[53,54] low heats of adsorption prevent the need for additional temperature maintenance, and fast adsorption and desorption is important for transfer to the power generation step.[55,56]

For uptake, there are two forms that are commonly measured, excess and absolute.[50] Excess uptake (or capacity) is the amount of gas adsorbed over a gas that does not interact with the framework under the same pressure and temperature. Absolute uptake is the difference between the density of the free gas with the gas in the MOF pores. It is recommended that both of these values are reported such that researchers can more easily compare materials.

There have been more studies on hydrogen uptake in porous materials to attempt to determine a structure property relation.[57] In 2019, a computational screening of ~ 500000 MOFs, both real and hypothetical, was done to test for hydrogen storage at 77 K with a pressure swing between 100 and 5 bar.[58] The importance of this research was demonstrating the relationship between geometric properties and usable capacity. As a MOF's pore volume, gravimetric surface area and void fraction increases, so does usable capacity, the reversibly stored gas between delivery and maximum storage pressure. The inverse of this correlation holds true for a small range of density and volumetric surface area.

1.3 Computational simulations of MOFs

The increase in computing power over the last few decades has allowed for computational simulations to be an integral part of chemistry including materials development. The main goals of these tools are to provide insight into the chemical interactions of a given system. The advantage of computational simulations is that they can be run en masse to search for high performers without needing to go through the process of making the materials. This process is known as computational high-throughput screening.

There are many different methods to perform computational simulations that are used for specific purposes that prioritise accuracy and time of calculation. Grand Canonical Monte Carlo (GCMC) simulations can be used to simulate gas adsorption isotherms, a key experimental measurement of the adsorption capacity of a material. Density Functional Theory (DFT) calculations are used to calculate the electronic structure of an atomic system which can be used to optimize the atom positions, calculate binding energies, and model reactive interactions between atoms. Molecular Dynamics simulations are used to study the time evolution of the atoms in a system and can provide estimates of the diffusion coefficients of a guest in a material. The details of these methods are expanded on in Chapter 2 of this thesis.

The Woo lab previously performed a computational analysis on Calgary Framework 20 (CALF-20),[59] a water stable MOF with CCS capabilities. The computational methods were able to simulate adsorption isotherms which were in close agreement with experimental results. The computationally calculated binding sites were also studied to understand the guest interactions. Savante is currently working to scale up the MOF manufacturing process for industrial use.[60]

1.4 MOF Criteria for Gas Separation and Storage

While MOFs are being studied for a variety of use cases, the Woo lab group primarily investigates them for gas separation and storage applications. In a theoretical study, a researcher must determine how to identify MOFs that would be good at these tasks. The main properties looked at for CCS abilities are working capacity and selectivity.

1.4.1 Working Capacity

The working capacity is calculated as the difference between the gas adsorbed at adsorption conditions (typically low temperature and high pressure) and desorption conditions (high temperature and low pressure), Δq.[61] To measure working capacity, an isotherm must be calculated either computationally or from experiment. An isotherm, shown in Figure 1.3, is a plot where the uptake of a gas by a material is plotted against the partial pressure of the gas at a fixed temperature. The larger the working capacity is, the more gas can be adsorbed per unit of material per adsorption/desorption cycle.

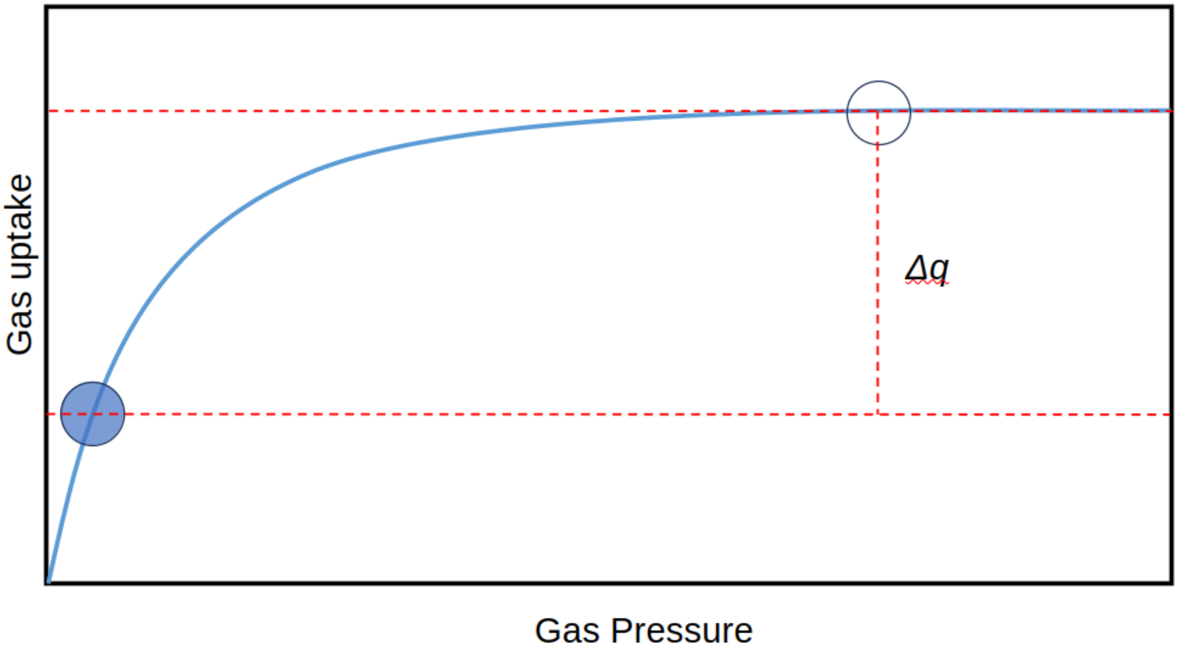

Figure 1.3: Sample isotherm of gas uptake in a MOF. Closed circle represents desorption conditions and open circle represents adsorption conditions. Δq is the working capacity of the system.

1.4.2 Selectivity

Selectivity refers to the metric of measuring the ability of a MOF to adsorb one molecule over another. Equation 1.1 shows how this is calculated

$$\boldsymbol{S = \frac{q_i/p_i}{q_j/p_j}} \tag{1.1}$$

where *S* is the selectivity, *q* refers to the uptake of specified gas (i/j), and *p* refers to the partial pressure of the specified gas (i/j).[62] These values can also be calculated from experiment or from computational simulations.

Selectivity is an important adsorption property for gas separation applications. For example, flue gas from post-combustion carbon capture contains several gases, mostly N_2, other than CO_2. Therefore, if the MOF does not selectively adsorb CO_2 more energy input will be required to compress and sequester the nitrogen.

1.5 MOF Databases

The components of a MOF are very general, leading to large numbers of possibilities for what a MOF can consist of. The huge number of MOFs as well as their diverse use cases mean that having a resource that allows an interested party to easily parse information that has already been researched is important. This had led to the creation of a large number of MOF databases that are curated for specific interests.

1.5.1 Experimental MOF Databases

The Cambridge Crystallographic Data Centre (CCDC) is an organization established by Cambridge University with the goal of collecting discovered crystal structures in 1965 by cataloguing them in the Cambridge Structural Database (CSD).[63] A deposited material's structure is determined by X-ray or neutron diffraction and are uploaded to the database as a *.cif* (crystal structure file) file. Structures in these databases can contain errors such as missing hydrogen atoms, missing counter ions, and crystallographic disorder, seen as hyper-coordinated or overlapping atoms. The submitted MOFs may also be in an inactivated state, where solvent is still present in the molecule, so it is up to a researcher to determine the MOF's activated structure. Solvent can affect pore structure by obstructing pores. Watanabe performed a study on CSD MOFs in 2012 needing to remove MOFs based on disorder, lack of hydrogen and presence of solvent leaving 1163 of ~30,000 starting materials.[64] Goldsmith conducted another study in 2013 using the CSD database but used a set of automated scripts to determine which structures were MOFs and remove the solvent from the structures which reduced a starting ~38,800 materials to ~22,700.[65] To make searching for MOFs in the CSD easier, an integrated database was created from the CSD using automated tools known as the CSD MOF subset.[66] To prevent future groups from having to filter such a large number of materials from the CSD, groups have tried to create their own libraries for ease of use.

In 2014, Chung et al created the Computational-Ready Experimental (CoRE) Database for use in high-throughput screening.[67] CoRE's goal was to have a database of experimental MOFs that are ready to use for computational simulations. CoRE was created by taking materials from the CSD and running them through a series of automated scripts and manual checks to collect materials that fit the definition of MOF and are not considered disordered. The CoRE database includes ~14,000 MOFs.

These collected materials were then run through a set of scripts to remove the solvent. CoRE's last update was in 2019[68] and it separated the database into all solvent removed (ASR) and free solvent removed (FSR). Figure 1.4 shows the same MOF with different parts of the solvent removed where 1.4 (a) has no solvent removed, (b) has the free solvent removed, and (c) has all of the solvent removed. Solvent is considered bound when the distance between the metal and solvent oxygen atoms are less than 0.4 Å, it is considered free otherwise. The process of removing a solvent from a MOF is referred to as activation and can result in a change in the structure as it can create open metal sites or change the pore geometries. The CoRE database only has the solvent removed from the structure, without any optimization of the MOF's atom locations afterwards, resulting in potentially inaccurate structures in the database. The changes that a MOF undergoes after activation will affect uptake[69] so it is important to ensure that all structures accurately represent experimental materials.

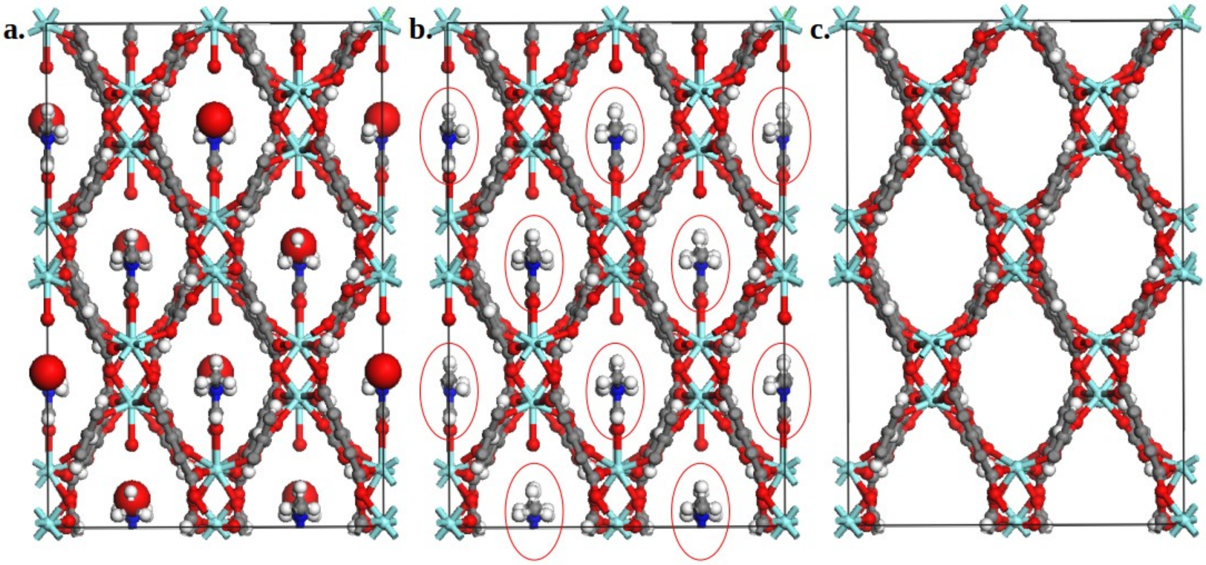

Figure 1.4: A sample MOF containing solvent, refcode NADZEZ, containing (a) all solvent with the free solvent represented by the larger red spheres, (b) with the bound solvent circled in red and (c) with all solvent removed

The Quantum MOF database (QMOF) was another database created from corrected CSD structures.[70] The goal of this database is to calculate quantum chemical properties of the MOFs such that machine learning models can be trained on the database. The database contained ~15000 structures after the filtering and contained various properties computed by DFT such as band gaps, charge densities and absolute energies.

1.5.2 Hypothetical MOF Databases

Due to the knowledge collected from experimental data, researchers can isolate SBUs and attempt to create new structures with them. Computationally, a researcher would not have to go through the effort of attempting to synthesise compounds in a lab and instead could create structures based on previous knowledge on possible SBUs. Groups can then use these hypothetical MOFs (hMOFs) for high-throughput screening to identify high performing candidates.

The first example of one of these hypothetical databases was created in 2012 by Wilmer et al.[71] The SBUs for this database were collected from existing crystallographic data then grouped based on reagents used in synthesis. These building blocks can bind based on chemical composition and geometry alone. In total they collected five inorganic SBUs, 41 organic SBUs (these could have different terminal groups) and 12 functional groups generated from experimental data to use in creating over 137000 hMOFs. One study by Gurnani et al. used machine learning to generate models predicting methane adsorption in a MOF.[72] However, the low number of SBUs and 98% of the MOFs having the same topology has led to the database lacking diversity and SBUs containing errors that will be mentioned in section 1.5.3. Additionally, it was found that many of the structures possessed near overlapping atoms.

Woo lab PhD. student, Peter Boyd, created ToBasCCo (Topology Based Crystal Constructor), an algorithm to create hMOFs based on graph theory.[73] ToBasCCo was used to create what is now called the Boyd-Woo database used in two studies. The Boyd-Woo database was used to determine a structure-property relations using machine learning to identify MOFs with strong CO_2 uptake and CO_2/CH_4 selectivity.[74,75] This database was used to identify high performing binding sites for CO_2 capture under wet flue gas conditions. Two of these hypothetical MOFs were then synthesized[76] and shown to capture CO_2 under humid conditions as predicted.

The last hypothetical database of note is by Mujumdar et al.[77] where the focus was on the chemical diversity of the structures. The group collected 14 metal nodes that are not commonly found in hypothetical databases with a series of metal centres and connection points and the organic linkers were taken from different databases. This led to a database of ~20000 hypothetical MOFs.

1.5.3 Errors in Existing Databases

Further examination into the CoRE database by PhD. students, Andrew White and Marco Gibaldi of the Woo lab, determined that the popular CoRE database contains a large fraction of structures (~50%) that have serious structural errors.[78] Moreover, since most hypothetical MOF databases derived their structural binding units from the CoRE database, most hMOF databases have structural error rates of more than 40%. This then puts into question the many studies that use these databases for screening or for building machine learning models.

1.6 Machine Learning and Descriptors

The goal of screening existing databases is to search for high performing candidates however; the sheer number of MOFs can make these screenings prohibitively expensive if the computation of the property of interest is time consuming. An alternative approach can be to use machine learning algorithms aid in development and identification of high performing MOFs. An explanation of how machine learning works is beyond the scope of this thesis, but in general terms, it is a method of computationally identifying patterns in sets of data. To perform this the ML model needs be able to interpret the data that it is given, as such the properties of a material such as pore geometries, atom distances, atom charges need to be quantified as descriptors. Descriptors are quantitative values used as input of a machine learning model. The descriptors used in this thesis will be explained in the methods section.

1.7 Thesis Outline

The organization of this thesis is as follows. Chapter 2 is a review of the various tools and methods used to analyze gas adsorption in MOFs. Programs such as GALA and VASP will be described as well as techniques such as periodic boundary conditions and GCMC simulations.

Chapter 3 is an overview of the computational simulations that were performed on a MOF created by the lab of Professor Kyriakos Stylianou in order to determine its properties. The goal of this work was to determine the cause of the unusually high hydrogen uptake of the MOF through analysis of the experimental data and a computational study. The MOF’s activated structure needed to be determined first through DFT structural optimizations as only the unactivated structure was determined through powder X-ray diffraction. The binding mode of hydrogen was then determined through computational simulations. Studying the cause of the strong uptake of the MOF can lead to the rational design of superior adsorbents.

Chapter 4 contains contributions to the ARC-MOF database. The database is a collection of hypothetical MOF structures containing DFT-derived partial atomic charges, various descriptors, and GCMC simulated adsorption data such that the database is computationally ready for high throughput screening and machine learning. This author’s work focused on collecting the geometric properties of the MOFs to understand the distribution of properties in the database. The goal is to determine the database MOFs were diverse with respect to the descriptors for the use of machine learning and high throughput screening.

Finally, the thesis ends with Chapter 5, which includes a summary of the conclusions of the thesis.

1.8 References

(1) *Global Energy Review 2025*. (2025). www.iea.org

(2) Mitchell, J. F. B. The "Greenhouse" Effect and Climate Change. **1989**, No. 89, 115–139.

(3) Roy, P.; Pal, S. C.; Chakrabortty, R.; Chowdhuri, I.; Saha, A.; Shit, M. Effects of Climate Change and Sea-Level Rise on Coastal Habitat: Vulnerability Assessment, Adaptation Strategies and Policy Recommendations. *J. Environ. Manage.* **2023**, *330*, 117187. https://doi.org/10.1016/j.jenvman.2022.117187.

(4) International Energy Agency. CO_2 Emission from Fuel and Combustion: Overview. *Int. Energy Agency* **2019**, 1–514.

(5) Ahmed, M.; Shuai, C.; Ahmed, M. Analysis of Energy Consumption and Greenhouse Gas Emissions Trend in China, India, the USA, and Russia. *Int. J. Environ. Sci. Technol.* **2023**, *20* (3), 2683–2698. https://doi.org/10.1007/s13762-022-04159-y.

(6) Rackley, S. Carbon Capture and Storage. *Carbon Capture and Storage* **2009**, *325* (September), 1–392. https://doi.org/10.1016/C2009-0-19306-6.

(7) Lei, T.; Wang, D.; Yu, X.; Ma, S.; Zhao, W.; Cui, C.; Meng, J.; Tao, S.; Guan, D. Global Iron and Steel Plant CO2 Emissions and Carbon-Neutrality Pathways. *Nature* **2023**, *622* (7983), 514–520. https://doi.org/10.1038/s41586-023-06486-7.

(8) York, I. N.; Europe, I. Concrete Needs to Lose Its Colossal Carbon Footprint. *Nature* **2021**, *597* (7878), 593–594. https://doi.org/10.1038/d41586-021-02612-5.

(9) Kanniche, M.; Gros-Bonnivard, R.; Jaud, P.; Valle-Marcos, J.; Amann, J. M.; Bouallou, C. Pre-Combustion, Post-Combustion and Oxy-Combustion in Thermal Power Plant for CO2 Capture. *Appl. Therm. Eng.* **2010**, *30* (1), 53–62. https://doi.org/10.1016/j.applthermaleng.2009.05.005.

(10) Underschultz, J.; Dodds, K.; Michael, K.; Sharma, S.; Wall, T.; Whittaker, S. Carbon Capture and Storage. *Sustain. Miner. Energy Sect.* **2016**, 437–452. https://doi.org/10.1201/9781315369853-23.

(11) Folger, P. Carbon Capture: A Technology Assessment. *Technol. Issues Carbon Capture* **2011**, 37–153.

(12) Ho, M. T.; Allinson, G. W.; Wiley, D. E. Reducing the Cost of CO2 Capture from Flue Gases Using Pressure Swing Adsorption. *Ind. Eng. Chem. Res.* **2008**, *47* (14), 4883–4890. https://doi.org/10.1021/ie070831e.

(13) Zhang, L.; Xie, J. L.; Fu, D. Corrosion Behavior of Carbon Steel in the CO 2 Absorption Process Using MEA Activated K 2 CO 3 Aqueous Solutions. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *199* (4), 3917–3924. https://doi.org/10.1088/1755-1315/199/4/042041.

(14) Lepaumier, H.; Picq, D.; Carrette, P. L. New Amines for CO2 Capture. I. Mechanisms of Amine Degradation in the Presence of CO2. *Ind. Eng. Chem. Res.* **2009**, *48* (20), 9061–9067. https://doi.org/10.1021/ie900472x.

(15) Boundary Dam Integrated Carbon Capture and Sequestration Demonstration Project https://web.archive.org/web/20140810095512/http://www.globalccsinstitute.com/project/boundary-dam-integrated-carbon-capture-and-sequestration-demonstration-project.

(16) Leo, G. SNC-Lavalin-built carbon capture facility has "serious design issues": SaskPower https://www.cbc.ca/news/canada/saskatchewan/snc-lavalin-carbon-capture-project-

saskpower-1.3291554.
(17) Groom, N. Problems plagued U.S. CO2 capture project before shutdown: document https://www.reuters.com/article/us-usa-energy-carbon-capture-idUSKCN2523K8%0Ahttps://www.reuters.com/article/us-usa-energy-carbon-capture-idUSKCN2523K8 (accessed Aug 28, 2024).
(18) Kim, S.; Ko, D.; Moon, I. Dynamic Optimization of a Dual Pressure Swing Adsorption Process for Natural Gas Purification and Carbon Capture. *Ind. Eng. Chem. Res.* **2016**, *55* (48), 12444–12451. https://doi.org/10.1021/acs.iecr.5b04157.
(19) Mazloomi, K.; Gomes, C. Hydrogen as an Energy Carrier: Prospects and Challenges. *Renew. Sustain. Energy Rev.* **2012**, *16* (5), 3024–3033. https://doi.org/10.1016/j.rser.2012.02.028.
(20) Anenberg, S. C.; Miller, J.; Henze, D.; Minjares, R. A. Y. Pollution-Related Health Impacts of Transportation Sector Emission in 2010 and 2015. *ICCT, Clim. Clean Air Coal.* **2019**, 1–55.
(21) Verhelst, S.; Wallner, T. Hydrogen-Fueled Internal Combustion Engines. *Prog. Energy Combust. Sci.* **2009**, *35* (6), 490–527. https://doi.org/10.1016/j.pecs.2009.08.001.
(22) Balat, M. Potential Importance of Hydrogen as a Future Solution to Environmental and Transportation Problems. **2008**, *33*, 4013–4029. https://doi.org/10.1016/j.ijhydene.2008.05.047.
(23) Kumar, N., Lee, S. Y., & Park, S. J. (2024). Advancements in hydrogen storage technologies: A comprehensive review of materials, methods, and economic policy. In *Nano Today* (Vol. 56). Elsevier B.V. https://doi.org/10.1016/j.nantod.2024.102302
(24) Mehr, A. S., Phillips, A. D., Brandon, M. P., Pryce, M. T., & Carton, J. G. (2024). Recent challenges and development of technical and technoeconomic aspects for hydrogen storage, insights at different scales; A state of art review. In *International Journal of Hydrogen Energy* (Vol. 70, pp. 786–815). Elsevier Ltd. https://doi.org/10.1016/j.ijhydene.2024.05.182
(25) Carpetis, C. Estimation of Storage Costs for Large Hydrogen Storage Facilities. *Int. J. Hydrogen Energy* **1982**, *7* (2), 191–203. https://doi.org/10.1016/0360-3199(82)90146-X.
(26) Carpetis, C. ESTIMATION OF STORAGE COSTS FOR LARGE HYDROGEN STORAGE FACILITIES. *Int. J. Hydrog. Energy,* **1982**, *7* (2), 191–203.
(27) Das, L. M.; Gulati, R.; Gupta, P. K. Performance Evaluation of a Hydrogen-Fuelled Spark Ignition Engine Using Electronically Controlled Solenoid-Actuated Injection System. *Int. J. Hydrogen Energy* **2000**, *25* (6), 569–579. https://doi.org/10.1016/S0360-3199(99)00059-2.
(28) Schlapbach, L.; Züttel, A. For Mobile Applications. *Nature* **2001**, *414* (November), 353–358. https://doi.org/10.1038/35104634.
(29) Zhou, L. Progress and Problems in Hydrogen Storage Methods. *Renew. Sustain. Energy Rev.* **2005**, *9* (4), 395–408. https://doi.org/10.1016/j.rser.2004.05.005.
(30) Hydrogen Storage https://www.energy.gov/eere/fuelcells/hydrogen-storage (accessed Jul 24, 2024).
(31) Batten, S. R.; Champness, N. R.; Chen, X.-M.; Garcia-Martinez, J.; Kitagawa, S.; Öhrström, L.; O’Keeffe, M.; Paik Suh, M.; Reedijk, J. Terminology of Metal–Organic Frameworks and Coordination Polymers (IUPAC Recommendations 2013). *Pure Appl. Chem.* **2013**, *85* (8), 1715–1724. https://doi.org/10.1351/PAC-REC-12-11-20.
(32) Jia, T.; Gu, Y.; Li, F. Progress and Potential of Metal-Organic Frameworks (MOFs) for

Gas Storage and Separation: A Review. *J. Environ. Chem. Eng.* **2022**, *10* (5), 108300. https://doi.org/10.1016/j.jece.2022.108300.

(33) Marco-Lozar, J. P.; Juan-Juan, J.; Suárez-García, F.; Cazorla-Amorós, D.; Linares-Solano, A. MOF-5 and Activated Carbons as Adsorbents for Gas Storage. *Int. J. Hydrogen Energy* **2012**, *37* (3), 2370–2381. https://doi.org/10.1016/j.ijhydene.2011.11.023.

(34) Alezi, D.; Belmabkhout, Y.; Suyetin, M.; Bhatt, P. M.; Weseli??ski, L. J.; Solovyeva, V.; Adil, K.; Spanopoulos, I.; Trikalitis, P. N.; Emwas, A. H.; Eddaoudi, M. MOF Crystal Chemistry Paving the Way to Gas Storage Needs: Aluminum-Based Soc -MOF for CH4, O2, and CO2 Storage. *J. Am. Chem. Soc.* **2015**, *137* (41), 13308–13318. https://doi.org/10.1021/jacs.5b07053.

(35) Zhang, X.; Chen, A.; Zhong, M.; Zhang, Z.; Zhang, X.; Zhou, Z.; Bu, X. H. *Metal–Organic Frameworks (MOFs) and MOF-Derived Materials for Energy Storage and Conversion*; Springer Singapore, 2019; Vol. 2. https://doi.org/10.1007/s41918-018-0024-x.

(36) Ye, Z.; Jiang, Y.; Li, L.; Wu, F.; Chen, R. *Rational Design of MOF-Based Materials for Next-Generation Rechargeable Batteries*; 2021; Vol. 13. https://doi.org/10.1007/s40820-021-00726-z.

(37) Hasan, M. N.; Bera, A.; Maji, T. K.; Pal, S. K. Sensitization of Nontoxic MOF for Their Potential Drug Delivery Application against Microbial Infection. *Inorganica Chim. Acta* **2021**, *523* (March), 120381. https://doi.org/10.1016/j.ica.2021.120381.

(38) Mallakpour, S.; Nikkhoo, E.; Hussain, C. M. Application of MOF Materials as Drug Delivery Systems for Cancer Therapy and Dermal Treatment. *Coord. Chem. Rev.* **2022**, *451*, 214262. https://doi.org/10.1016/j.ccr.2021.214262.

(39) Kim, H.; Rao, S. R.; Kapustin, E. A.; Zhao, L.; Yang, S.; Yaghi, O. M.; Wang, E. N. Adsorption-Based Atmospheric Water Harvesting Device for Arid Climates. *Nat. Commun.* **2018**, *9* (1), 1191. https://doi.org/10.1038/s41467-018-03162-7.

(40) Xu, W.; Yaghi, O. M. Metal–Organic Frameworks for Water Harvesting from Air, Anywhere, Anytime. *ACS Cent. Sci.* **2020**, *6* (8), 1348–1354. https://doi.org/10.1021/acscentsci.0c00678.

(41) Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A.; Snurr, R. Q.; Nguyen, S. T.; Yazaydın, Ö.; Hupp, J. T. Metal-Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit? *J. Am. Chem. Soc.* **2012**, *134* (36), 15016–15021. https://doi.org/10.1021/ja3055639.

(42) Sarkisov, L.; Bueno-Perez, R.; Sutharson, M.; Fairen-Jimenez, D. Materials Informatics with PoreBlazer v4.0 and the CSD MOF Database. *Chem. Mater.* **2020**, *32* (23), 9849–9867. https://doi.org/10.1021/acs.chemmater.0c03575.

(43) Haldoupis, E.; Nair, S.; Sholl, D. S. Efficient Calculation of Diffusion Limitations in Metal Organic Framework Materials: A Tool for Identifying Materials for Kinetic Separations. *J. Am. Chem. Soc.* **2010**, *132* (21), 7528–7539. https://doi.org/10.1021/ja1023699.

(44) Fernandez, M.; Woo, T. K.; Wilmer, C. E.; Snurr, R. Q. Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks. *J. Phys. Chem. C* **2013**, *117* (15), 7681–7689. https://doi.org/10.1021/jp4006422.

(45) Fernandez, M.; Barnard, A. S. Geometrical Properties Can Predict CO2 and N2 Adsorption Performance of Metal-Organic Frameworks (MOFs) at Low Pressure. *ACS*

Comb. Sci. **2016**, *18* (5), 243–252. https://doi.org/10.1021/acscombsci.5b00188.

(46) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Quantitative Structure–Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO2Working Capacity and CO2/CH4Selectivity for Methane Purification. *Eur. J. Inorg. Chem.* **2016**, *2016* (27), 4505–4511. https://doi.org/10.1002/ejic.201600365.

(47) Hirscher, M.; Yartys, V. A.; Baricco, M.; Bellosta von Colbe, J.; Blanchard, D.; Bowman, R. C.; Broom, D. P.; Buckley, C. E.; Chang, F.; Chen, P.; Cho, Y. W.; Crivello, J. C.; Cuevas, F.; David, W. I. F.; de Jongh, P. E.; Denys, R. V.; Dornheim, M.; Felderhoff, M.; Filinchuk, Y.; Froudakis, G. E.; Grant, D. M.; Gray, E. M. A.; Hauback, B. C.; He, T.; Humphries, T. D.; Jensen, T. R.; Kim, S.; Kojima, Y.; Latroche, M.; Li, H. W.; Lototskyy, M. V.; Makepeace, J. W.; Møller, K. T.; Naheed, L.; Ngene, P.; Noréus, D.; Nygård, M. M.; Orimo, S. ichi; Paskevicius, M.; Pasquini, L.; Ravnsbæk, D. B.; Veronica Sofianos, M.; Udovic, T. J.; Vegge, T.; Walker, G. S.; Webb, C. J.; Weidenthaler, C.; Zlotea, C. Materials for Hydrogen-Based Energy Storage – Past, Recent Progress and Future Outlook. *J. Alloys Compd.* **2020**, *827*. https://doi.org/10.1016/j.jallcom.2019.153548.

(48) Rosi, N. L.; Eckert, J.; Eddaoudi, M.; Vodak, D. T.; Kim, J.; O'Keeffe, M.; Yaghi, O. M. Hydrogen Storage in Microporous Metal-Organic Frameworks. *Science (80-.).* **2003**, *300* (5622), 1127–1129. https://doi.org/10.1126/science.1083440.

(49) Rowsell, J. L. C.; Millward, A. R.; Park, K. S.; Yaghi, O. M. Hydrogen Sorption in Functionalized Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2004**, *126* (18), 5666–5667. https://doi.org/10.1021/ja049408c.

(50) Broom, D. P.; Webb, C. J.; Hurst, K. E.; Parilla, P. A.; Gennett, T.; Brown, C. M.; Zacharia, R.; Tylianakis, E.; Klontzas, E.; Froudakis, G. E.; Steriotis, T. A.; Trikalitis, P. N.; Anton, D. L.; Hardy, B.; Tamburello, D.; Corgnale, C.; van Hassel, B. A.; Cossement, D.; Chahine, R.; Hirscher, M. Outlook and Challenges for Hydrogen Storage in Nanoporous Materials. *Appl. Phys. A Mater. Sci. Process.* **2016**, *122* (3), 1–21. https://doi.org/10.1007/s00339-016-9651-4.

(51) Dietzel, P. D. C.; Panella, B.; Hirscher, M.; Blom, R.; Fjellvåg, H. Hydrogen Adsorption in a Nickel Based Coordination Polymer with Open Metal Sites in the Cylindrical Cavities of the Desolvated Framework. *Chem. Commun.* **2006**, *1* (9), 959–961. https://doi.org/10.1039/b515434k.

(52) Rojas-Garcia, E.; Castañeda-Ramírez, A. A.; Angeles-Beltrán, D.; López-Medina, R.; Maubert-Franco, A. M. Enhancing in the Hydrogen Storage by SWCNT/HKUST-1 Composites: Effect of SWCNT Amount. *Catal. Today* **2022**, *394–396* (August 2021), 357–364. https://doi.org/10.1016/j.cattod.2021.08.004.

(53) Broom, D. P.; Webb, C. J.; Fanourgakis, G. S.; Froudakis, G. E.; Trikalitis, P. N.; Hirscher, M. Concepts for Improving Hydrogen Storage in Nanoporous Materials. *Int. J. Hydrogen Energy* **2019**, *44* (15), 7768–7779. https://doi.org/10.1016/j.ijhydene.2019.01.224.

(54) Panella, B.; Hirscher, M.; Roth, S. Hydrogen Adsorption in Different Carbon Nanostructures. *Carbon N. Y.* **2005**, *43* (10), 2209–2214. https://doi.org/10.1016/j.carbon.2005.03.037.

(55) Hirscher, M.; Panella, B. Hydrogen Storage in Metal-Organic Frameworks. *Scr. Mater.* **2007**, *56* (10), 809–812. https://doi.org/10.1016/j.scriptamat.2007.01.005.

(56) Hirscher, M. Hydrogen Storage by Cryoadsorption in Ultrahigh-Porosity Metal-Organic

Frameworks. *Angew. Chemie - Int. Ed.* **2011**, *50* (3), 581–582. https://doi.org/10.1002/anie.201006913.

(57) Balderas-Xicohténcatl, R.; Schlichtenmayer, M.; Hirscher, M. Volumetric Hydrogen Storage Capacity in Metal–Organic Frameworks. *Energy Technol.* **2018**, *6* (3), 578–582. https://doi.org/10.1002/ente.201700636.

(58) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-09365-w.

(59) Lin, J.-B., Nguyen, T. T. T., Vaidhyanathan, R., Burner, J., Taylor, J. M., Durekova, H., Akhtar, F., Mah, R. K., Ghaffari-Nik, O., Marx, S., Fylstra, N., Iremonger, S. S., Dawson, K. W., Sarkar, P., Hovington, P., Rajendran, A., Woo, T. K., & Shimizu, G. K. H. (n.d.). *A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture*. https://www.science.org

(60) *MOF Sorbent on a Roll – A Scalable Solution for Gigaton Scale Carbon Capture*. (2021, December 16).

(61) Berger, A. H.; Bhown, A. S. Comparing Physisorption and Chemisorption Solid Sorbents for Use Separating CO2 from Flue Gas Using Temperature Swing Adsorption. *Energy Procedia* **2011**, *4*, 562–567. https://doi.org/10.1016/j.egypro.2011.01.089.

(62) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High Co 2 Working Capacity and Co 2 /h 2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123* (7). https://doi.org/10.1021/acs.jpcc.8b10644.

(63) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, *58* (3 PART 1), 380–388. https://doi.org/10.1107/S0108768102003890.

(64) Watanabe, T.; Sholl, D. S. Accelerating Applications of Metal–Organic Frameworks for Gas Adsorption and Separation by Computational Screening of Materials. *Langmuir* **2012**, *28* (40), 14114–14128. https://doi.org/10.1021/la301915s.

(65) Goldsmith, J.; Wong-Foy, A. G.; Cafarella, M. J.; Siegel, D. J. Theoretical Limits of Hydrogen Storage in Metal–Organic Frameworks: Opportunities and Trade-Offs. *Chem. Mater.* **2013**, *25* (16), 3373–3382. https://doi.org/10.1021/cm401978e.

(66) Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal-Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**. https://doi.org/10.1021/acs.chemmater.7b00441.

(67) Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal-Organic Frameworks: A Tool to Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26* (21), 6185–6192. https://doi.org/10.1021/cm502594j.

(68) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985–5998. https://doi.org/10.1021/acs.jced.9b00835.

(69) Nazarian, D.; Camp, J. S.; Chung, Y. G.; Snurr, R. Q.; Sholl, D. S. Large-Scale Refinement of Metal-Organic Framework Structures Using Density Functional Theory. *Chem. Mater.* **2017**, *29* (6), 2521–2528. https://doi.org/10.1021/acs.chemmater.6b04226.

(70) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine Learning the Quantum-Chemical Properties of Metal–Organic Frameworks for Accelerated Materials Discovery. *Matter* **2021**, *4* (5), 1578–1597. https://doi.org/10.1016/j.matt.2021.02.015.

(71) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal-Organic Frameworks. *Nat. Chem.* **2012**, *4*, 83–89. https://doi.org/10.1038/nchem.1192.

(72) Gurnani, R.; Yu, Z.; Kim, C.; Sholl, D. S.; Ramprasad, R. Interpretable Machine Learning-Based Predictions of Methane Uptake Isotherms in Metal-Organic Frameworks. *Chem. Mater.* **2021**, *33*, 3552. https://doi.org/10.1021/acs.chemmater.0c04729.

(73) Boyd, P. G. P. G. P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18* (21), 3777–3792. https://doi.org/10.1039/C6CE00407E.

(74) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z. M. Z.; Woo, T. K. Rapid and Accurate Machine Learning Recognition of High Performing Metal Organic Frameworks for CO2capture. *J. Phys. Chem. Lett.* **2014**, *5* (17), 3056–3060. https://doi.org/10.1021/jz501331m.

(75) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. 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.* **2016**, *2016* (27), 4505–4511. https://doi.org/10.1002/ejic.201600365.

(76) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; Reimer, J. A.; Navarro, J. A. R.; Woo, T. K.; Garcia, S.; Stylianou, K. C.; Smit, B. Data-Driven Design of Metal–Organic Frameworks for Wet Flue Gas CO2 Capture. *Nature* **2019**, *576* (7786), 253–256. https://doi.org/10.1038/s41586-019-1798-7.

(77) Majumdar, S.; Moosavi, S. M.; Jablonka, K. M.; Ongari, D.; Smit, B. Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening. *ACS Appl. Mater. Interfaces* **2021**, *13* (51), 61004–61014. https://doi.org/10.1021/acsami.1c16220.

(78) White, A. J.; Gibaldi, M.; Burner, J.; Mayo, R. A.; Woo, T. K. High Structural Error Rates in “Computation-Ready” MOF Databases Discovered by Checking Metal Oxidation States. *J. Am. Chem. Soc.* **2025**. https://doi.org/10.1021/jacs.5c04914.

2 Methods

The many different computational methods and tools in this thesis will be introduced in this chapter. Due to the complexity of chemical systems, different methods need to be used to collect data in an efficient amount of time with an appropriate level of accuracy. The main methods that will be focused on in this section are Monte Carlo simulations and Density Functional Theory.

2.1 Empirical Forcefield Parameters

One of the simplest ways of viewing a molecular system is with classical force fields. Rather than calculating the entire electronic structure of a molecule or system, the atoms can be treated in a ball-and-spring manner. To do this, a user needs to know what forces are acting on each atom in a system at a given geometry and these are given by an empirically parameterized forcefield. There are many popular forcefields that each are useful for different chemical systems and are used to describe certain systems, such as UFF[1] (general purpose), AMBER[2] and CHARMM[3] (proteins and nucleic acids), TIP3P[4] (bulk water), etc. Once these parameters are collected the potential energy of the system can be calculated as a series of non-bonded and bonded interactions for a given configuration of atoms. A user can then sample multiple configurations for a system to determine what is known as the potential energy surface (PES) which relates the geometry of a system to the potential energy of the system.

For all gas adsorption simulations performed in this thesis, a fixed framework approximation is used meaning, the MOF framework is frozen during the simulations. Though it is possible for a MOFs structure to "flex" during gas adsorption, previous studies have shown this assumption to be a good approximation and time saving.[5–7] The MOF-guest interactions are determined with non-bonded Lennard-Jones potentials and Coulomb potentials using the partial atomic charge calculation described in the following sections.

2.1.1 Lennard-Jones Potential

The Lennard-Jones potential is a form of a potential to describe the dispersion and steric interactions between a pair of atoms as shown in Equation 2.1 as a van der Waals potential

$$E_{VDW} = 4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r}\right)^{12} - \left(\frac{\sigma_{AB}}{r}\right)^{6}\right] \tag{2.1}$$

In equation 2.1, and are parameters (potential well depth and distance where the potential is zero, respectively) of atom pairs *A* and *B*, and *r* is the distance between the two atoms. The $\left(\frac{\sigma_{AB}}{r}\right)^{12}$ is the repulsive term and $\left(\frac{\sigma_{AB}}{r}\right)^{6}$ is the attractive term. Forcefield parameters for and are specific to an atom pair, however if they have not been determined experimentally the Lorentz-Berthelot combining rules are used to calculate the parameters from the individual atom values needed for a given atom pair.

2.1.2 Partial Atomic Point Charges

Electrostatic interactions need to be modeled to properly understand guest-framework interactions in a gas adsorption simulation. Atomic point charges, *{q}*, can be used where electronic charges are assigned to nuclei centres. The electrostatic energy between two atoms can then be calculated with Equation 2.2

$$E_{elec} = \frac{q_A q_B}{4\pi\varepsilon_0 r_{AB}} \tag{2.2}$$

The electrostatic interactions decay much slower than the steric and dispersion interactions, compared to . To prevent the electrostatic energy calculation from being too expensive, the Ewald summation is used to calculate long-range interactions in periodic systems by separating the short and long range components.[8]

The Woo lab uses the Repeating Electrostatic Potential Extracted Atomic Charge Method (REPEAT) which was developed by Campañá et al. in 2009[9] to calculate atomic charges in infinite periodic systems. REPEAT charges are derived from a DFT calculation of the system of interest, such that the atomic charges best reproduce the electrostatic potential from the DFT calculation outside of a certain radius of the atoms, usually the van der Waals (vdW) radii. The DFT computed electrostatic potential (ESP) is evaluated on a grid outside the vdW spheres of the atoms as shown in Figure 2.1 for a molecular system. The charges on atoms are adjusted (or fit) to best reproduce the DFT calculated electrostatic potential on these grid points. The REPEAT method was the first method to allow for the electrostatic potential fitted charges to be determined by periodic DFT calculations. Previous studies have successfully used REPEAT for computational simulations,[5–7,10–12] but the use of quantum mechanical derived charges can still be expensive.

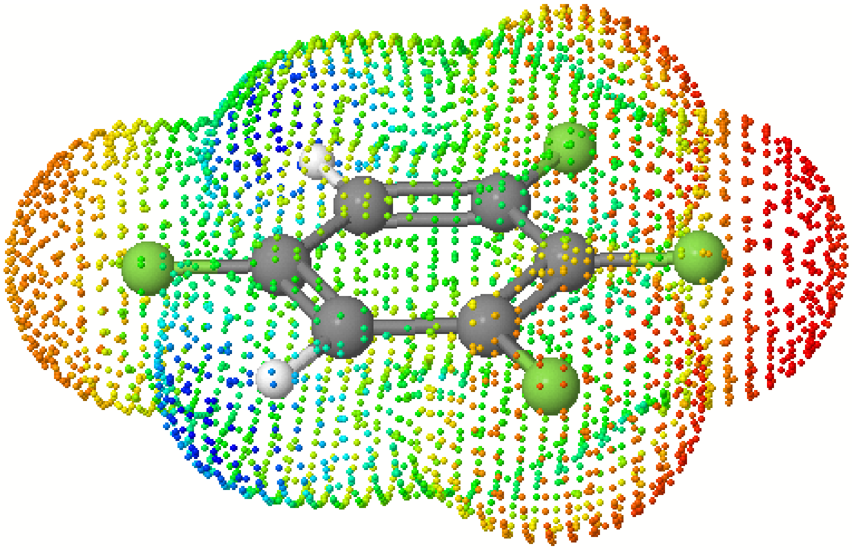

Figure 2.1: Representation of points on the vdW surface of fluorobenzene used to calculate atomic point charges on the molecule’s atoms. The point’s colour reflects the electrostatic potential

2.2 Periodic Boundary Conditions

Due to the crystalline structure of a MOF, a user is able to speed up calculation times by using periodic boundary conditions (PBCs).[13] Rather than viewing the structure as a large singular system, one can reduce the system to its smallest repeating unit, the unit-cell. This unit cell is then surrounded by images of itself in every direction, shown in Figure 2.2, so the boundary atoms are interacting with mirror images of the cell. When an atom exits one of the cell boundaries it enters on the other side meaning the unit cell never loses or gains atoms.

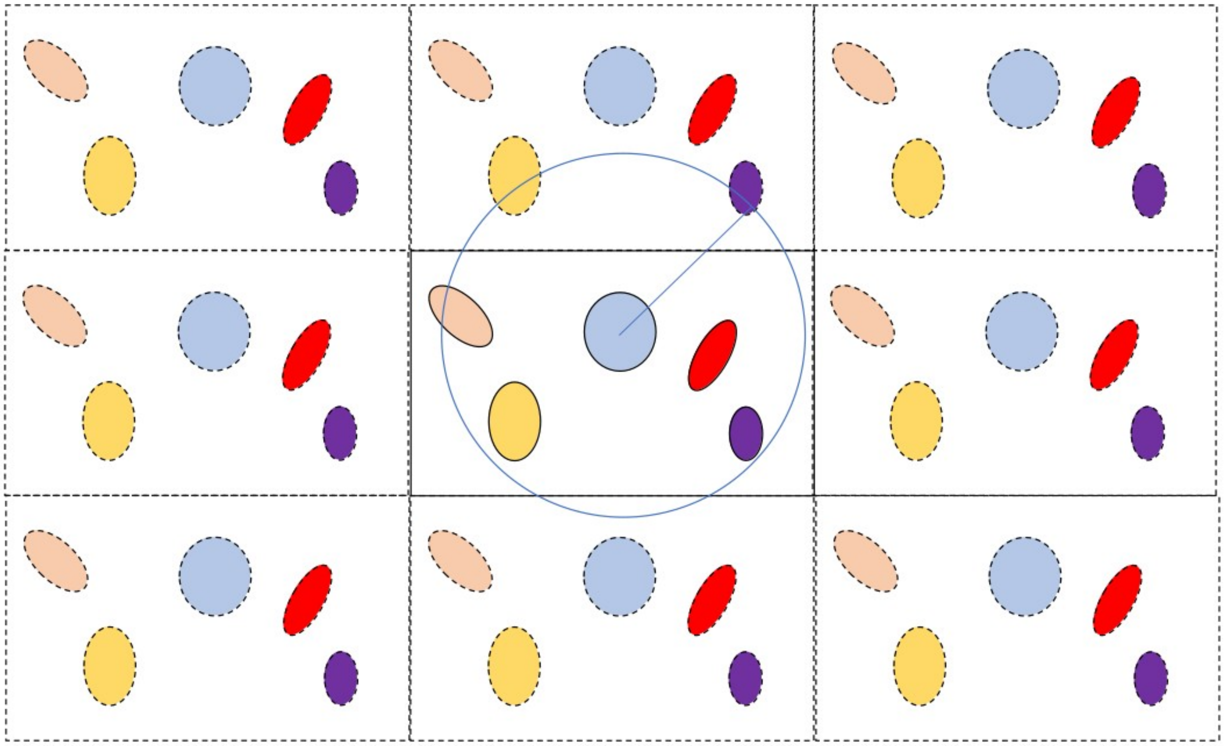

Figure 2.2: Example of periodic boundary conditions. Solid lines demonstrate the primary cell containing various particles, surrounding dashed lines represent the copies of the primary cell, and the blue circle surrounding the primary blue particle represents the cutoff distance.

The next important decision is to determine how the atoms interact and how many periodic images that the interactions are computed for. For short-range interactions (steric and dispersion) modelled by the Lennard-Jones potential, the user can create a cutoff distance that limits the distance at which atoms interact to speed up simulation time drastically. The minimum image convention is also used where only the closest periodic image of an atom pair in the cell can interact pairwise with itself. For the long-range electrostatic interactions, one must typically include the interaction across many periodic images. This is very compute intensive if done without optimization. To speed up the electrostatic interactions, the Ewald summation technique[8] is most commonly used.

2.3 Atomistic Monte Carlo Simulations

Monte Carlo is the most commonly used tool to predict gas uptake capacity of a porous material such as a MOF. With statistical mechanics one can calculate the macroscopic thermodynamic properties from the microscopic simulations of the guest molecules and the MOF framework. To start, one can create a hypothetical isolated system where certain macroscopic properties are set (such as temperature, volume, number of particles, etc.) to fixed values referred to as a macrostate, then the energy of this system can be solved for. This system can have multiple configurations, or microstates, for each macrostate. Sampling different guest configurations (microstates) creates an ensemble, which is a collection of possible microstates generated by perturbing atoms and/or molecules in the system where the thermodynamic properties remain fixed. The macroscopic properties can be calculated from the partition function (Q), which is different for each type of ensemble. There are many types of ensembles but the ones that will be discussed here are the canonical and grand canonical ensemble.

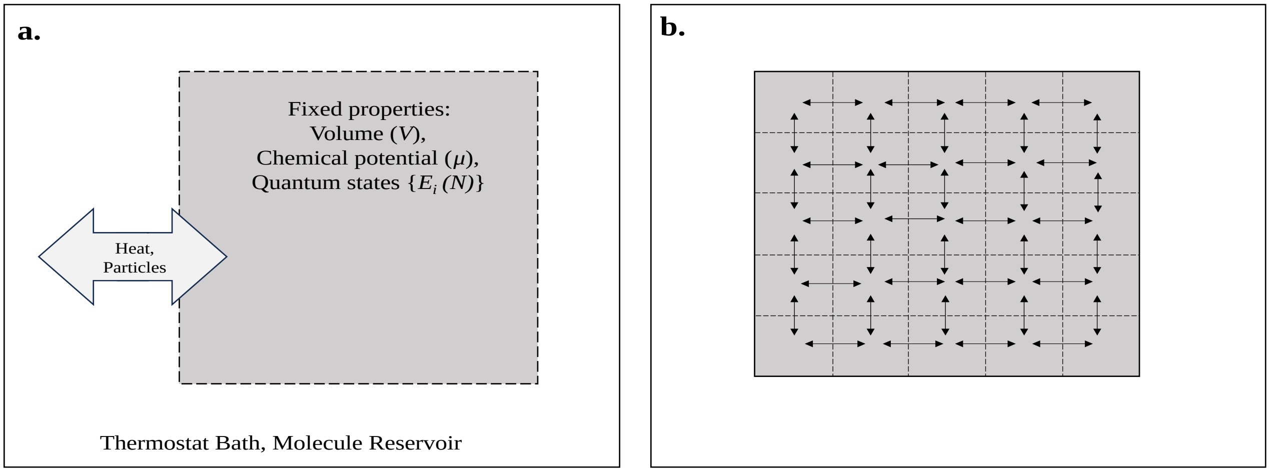

Figure 2.3: Representation of a grand cannonical ensemble system (a) System in a thermostat bath with walls that will allow the transfer of heat and particles, leaving V, μ and E_i fixed. (b) Replicas of the system (the ensemble) are placed next to each other with the ability to transfer heat and particles between the walls.

To begin a Monte Carlo simulation, a researcher needs to create a system with an initial configuration, for a MOF this will generally be the experimental structure with framework atom coordinates fixed. Guest molecules are then usually placed in the simulation with random positions. Different thermodynamic variables can be fixed, such as temperature and pressure, in an equilibrium experiment which correspond to different ensembles. A canonical ensemble fixes the number of particles (*N*), volume (*V*) and temperature (*T*) of a system. The canonical ensemble is sampled through translating or rotating guest molecules. Figure 2.3 shows a grand canonical ensemble where the temperature (T), volume (V) and chemical potential (μ) of a system are fixed.

The goal of a Monte Carlo simulation is to sample the microstates in a way that is representative of the Boltzmann distribution in an efficient way. This cannot be done through random sampling efficiently, so importance sampling needs to be performed with the commonly used Metropolis Monte Carlo[14] method. The algorithm is as follows:

1. Calculate the potential energy in the system's starting configuration.
2. Choose a guest molecule in the system and randomly perturb its position.
3. Calculate the potential energy of the new configuration.
4. If the new potential energy is less than the old potential energy, accept the microstate. If the new potential energy is higher than the old potential energy, the new configuration is only accepted with the following probability $e^{\frac{U_{new}-U_{old}}{kT}}$, (where *U* is the potential energy, *k* is the Boltzmann constant and *T* is the temperature). If the new configuration is not accepted, then we add the original configuration to the ensemble average.
5. This process is repeated until the user determines the simulation is converged. For typical simulations with non-polar gases in MOFs, this is roughly 1 million MC steps.

The above algorithm provides a Boltzmann distribution, a probability distribution of possible states a system can be in, of sampled configurations, where macroscopic properties are determined from the ensemble average of a system such that each configuration from the above procedure is weighted equally.

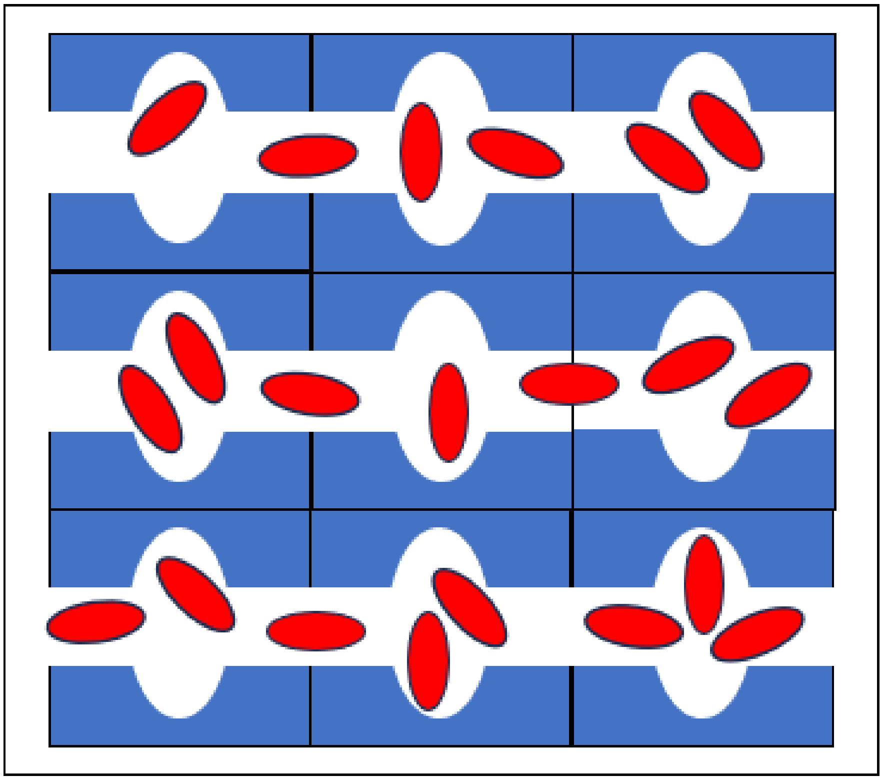

Figure 2.4: Simulation of a gas (red ovals) interacting with a porous solid (blue structure) at a fixed temperature. The system is set to equilibrium where the chemical potential of the gas is equal to that of the adsorbed gas. The potential can be calculated from the ideal gas law thus the free gas doesn't need to be simulated and only the bulk is examined.

To simulate gas adsorption at fixed temperature and partial pressure, we sample the grand canonical ensemble, where the number of particles or guest molecules is allowed to change but the chemical potential of the system is fixed as a thermodynamic variable. The ideal gas has an analytic form for it's chemical potential that is dependent on the mass of the gas and its partial pressure. At equilibrium, we know the chemical potential of the free gas molecules and the adsorbed gas molecules must be equal. Therefore, we can sample the number of guest molecules in the system by randomly inserting and deleting guest molecules in the simulation as shown in Figure 2.4. The probability of accepting the addition/deletion is related to the partial pressure (which is related to the chemical potential) we want to simulate. Again, one must compute the energy of the system after the guest molecule is added or deleted. The probability that a new configuration will be added when a guest molecule is added is given by Equation 2.3, and the probability that on will accept when a new configuration is deleted is given in Equation 2.4.

(2.4)

$$\boldsymbol{ACC_{n+1} = max(1, \frac{Vp}{kT(N+1)} e^{-\frac{\Delta U}{kT}})} \tag{2.3}$$

$$\boldsymbol{ACC_{n-1} = max(1, \frac{NkT}{Vp} e^{-\frac{\Delta U}{kT}})} \tag{2.4}$$

Where *V* is the volume, *p* is the pressure, *N* is the number of guests and *U*, *k* and *T* are the same as above. Again, if an addition/deletion move is rejected, the old configuration is added to the ensemble average.

The total number of guest molecules in the simulation averaged over all GCMC steps can be used to compute the amount of gas adsorbed by the MOF at a given pressure and temperature. The assumption is that all molecules in the simulation cell are adsorbed. Then running simulations over different pressures, one can obtain the gas adsorption isotherm.[15-17] Water is much harder to simulate as it is difficult to obtain force fields for due to its polar nature, and its ability to interact strongly with the MOF framework itself.[18-20]

The GCMC simulations completed in this thesis were performed by the code FastMC, a modified version of the MD package DL_POLY[21], created by members of the Woo Lab[22]. GCMC simulations are used for creating isotherms to predict gas uptake and to predict binding sites.

2.3.1 Guest Atom Localization Algorithm (GALA)

GALA is another program developed by the Woo lab group with the purpose of locating binding sites in a MOF from the probability distributions of the guest molecules generated from GCMC simulations. A GCMC simulation not only can be used to determine the amount of gas adsorbed at a given pressure, but it also generates probability distributions of the guest molecules. The binding sites can be determined from the location of the maxima in the probability distributions. However, due to the stochastic nature of the GCMC simulations, the probability distributions generated are usually quite noisy giving rise to thousands of maxima in a typical simulation of a MOF. Thus, the group developed the GALA method to automatically smooth the probability distributions and identify the location of the binding sites from the noisy probability distributions. GALA will also provide the binding energy at each site based on the calculation which can be separated into the van der Waals and electrostatic components.

2.4 Density Functional Theory

Based on first principles in quantum mechanics, Density Functional Theory, calculations have developed into an extremely powerful tool to accurately compute the potential energy surface of a compound. This is particularly true when a chemical reaction occurs or there is not an accurate empirical force field available. For the simulation of MOFs, DFT calculations are widely used.

As with Monte Carlo simulations, the periodic nature of MOFs can be used to reduce the cost of DFT calculations. For this thesis the DFT calculations were calculated with the Vienna Ab Initio Software Package (VASP),[23–26] a DFT package commonly used when performing calculations on periodic structures which uses plane wave basis functions. The calculations were done with the PBE[27] GGA[28] functional combined with the projector augmented-wave (PAW) method[29] to treat core regions of space. PBE is combined with dispersion corrections to examine the potential energy surface of a MOF. DFT calculations are used to optimize geometries of a structure, compute reaction and binding energies, compute electronic properties such as band gaps, etc.

The first focus on DFT in this thesis is to calculate the partial atomic charges through REPEAT to perform accurate GCMC simulations as mentioned in previous sections. The second use is to optimize a MOF's framework atoms from experimental single crystal X-ray diffraction (SCXRD) results. Hydrogen atom positions can not be accurately determined by SCXRD, therefore the hydrogen atom positions typically need to be added algorithmically or manually to represent the proper structure. The SCXRD file can also contain forms of disorder such as duplicate atoms appearing and over- or undercoordinated atoms. Once the errors are fixed in a crystal structure editor, the MOF framework will undergo a DFT geometry optimization to ensure accurate positions of the framework atoms.

2.5 Determination of Geometric Properties of MOFs

For machine learning and to develop structure property relationships, it is important to compute the geometric properties of a MOF. The Zeo++[30] code is currently the most popular tool to rapidly calculate the geometric properties of a MOF.

Zeo ++ can compute many different geometric features of a MOF given the atomic coordinates of the system. Density (g/cm^3) is gathered as it gives a general idea to how porous the material is as a high density generally means the material has low porosity and vice versa. When analysing a MOF channel, there are two diameters that are needed, these are the pore limiting diameter and the largest cavity diameter shown in Figure 2.5. The pore limiting diameter refers to the smallest diameter that a channel reaches, and the largest cavity diameter refers to the largest diameter a channel reaches.

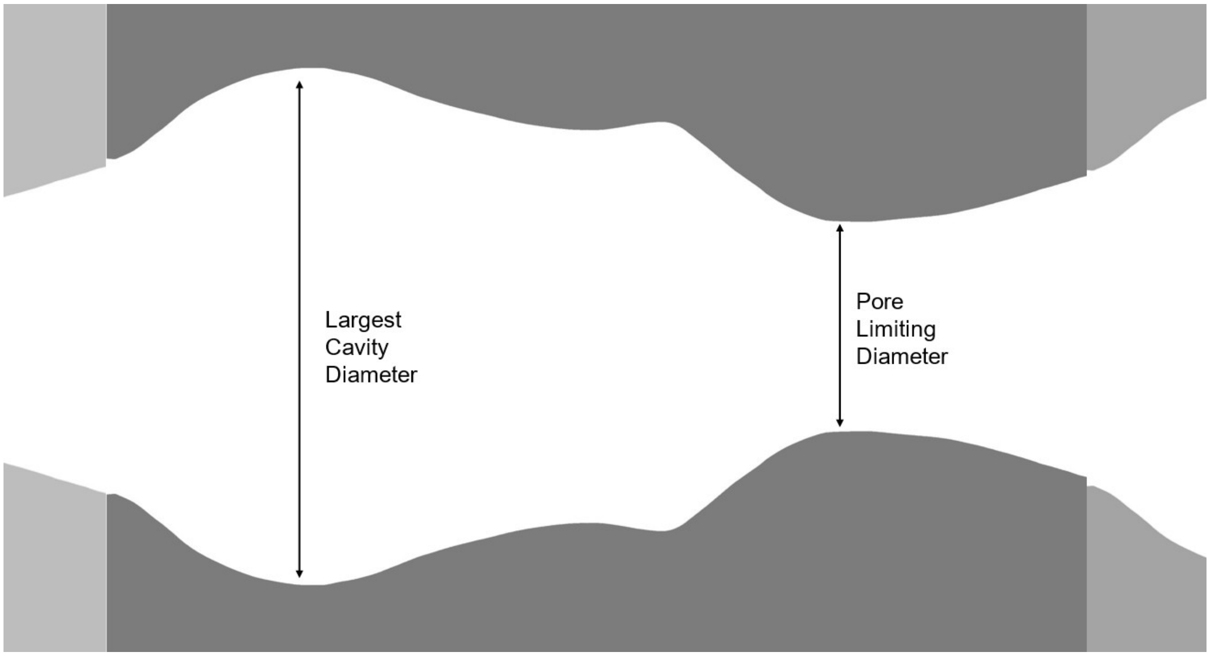

Figure 2.5: 2D representation of a pore in a periodic porous material.

The following properties can have different values based on the probe used and what value of the probe is being used. There are three different ways to calculate volume in a pore, either the entire volume, Figure 2.6 (b), or by calculating what is accessible by a probe, either from the centre or the edge, Figure 2.6 (c) and (d) respectively. These are known as the van der Waals surface, accessible surface area and Connoly surface area, respectively. The volume is also calculated based on the surface drawn by the probes, an example of how these look is on Figure 2.6. The geometric volume refers to the volume of the entire channel and the accessible volume can be calculated from the centre of the probe, or the edge of the probe referred to as probe-centred and probe occupiable volume respectively. The void fraction is the fraction of void areas to the solid part of the materials.

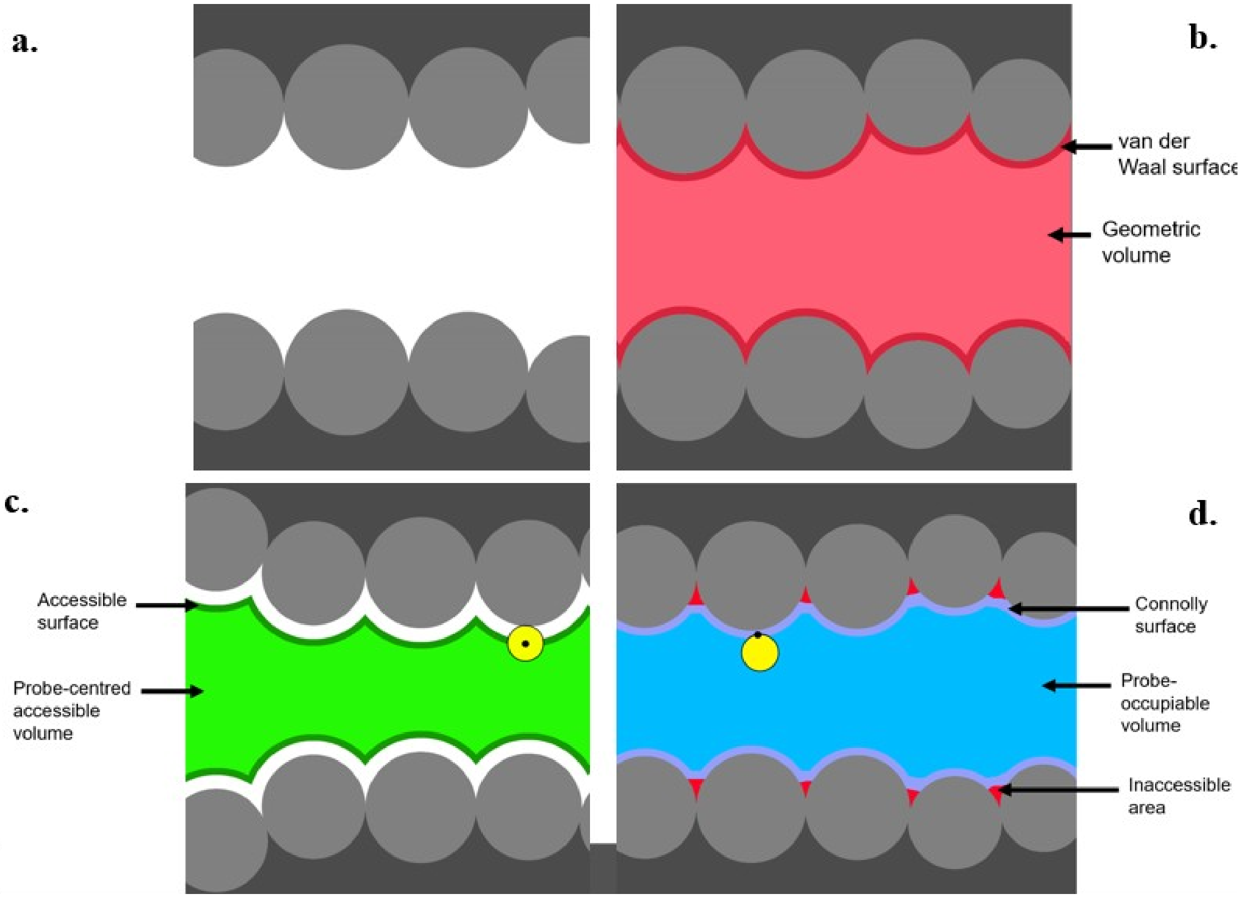

Figure 2.6: (a.) Image of the pore in a porous material where the surface area and volume are calculated based on different probe methods. (b.) pore where the surface area and volume are calculated based van der Waals surface of the pore's atoms. (c.) pore where the surface area and volume are calculated based on the centre of a probe (yellow circle). (d.) pore where the surface area and volume are calculated based on the edge of a probe (yellow circle).

2.6 Machine Learning

This work does not directly include any machine learning (ML) but does feature the generation of geometric descriptors for the use in machine learning. This section will have a brief overview of machine learning to understand why certain data was collected in a future section.

Material science has taken advantage of many different types of machine learning but supervised learning is most commonly used.[31] The goal of supervised learning is to determine a trend in data from a set of easy to calculate properties, chemical and/or geometric features referred to as descriptors, to a set of properties more difficult to calculate. There have been successful studies using machine learning including for gas separation and storage.[32–34]

2.6.1 Descriptors

A descriptor, in the context of machine learning, is any property used to describe the system in question. In the case of MOFs, or any chemical system, descriptors can be observable properties such as working capacity or selectivity based on previous calculations or experimental data, and non-physical properties such as atomic charge. The diverse nature of MOF structures requires the used of many descriptors for a machine learning model to properly distinguish the differences between two materials. Ward et al.[35] identified 145 descriptors for ML on MOFs with many shown in the Figure 2.7.

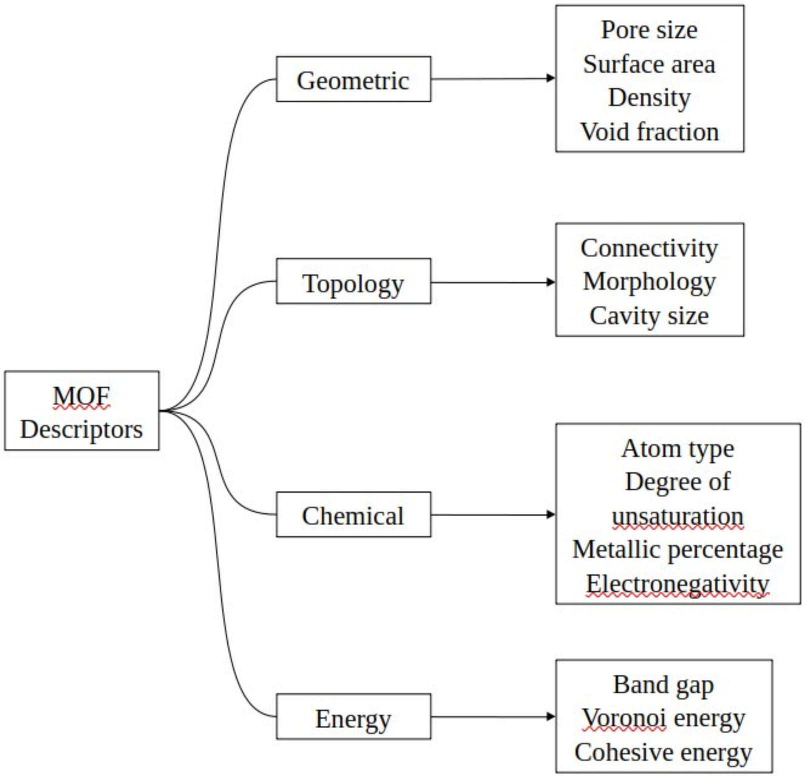

Figure 2.7: Machine learning descriptors for MOFs[36]

To speed up the calculation processes, descriptors must be selected based on the relative importance to the desired output data. Anderson et al.[37] performed a study to determine the relative importance of descriptors when studying working capacity, selectivity and loading for gas molecules and mixtures. The generally important descriptors include largest cavity diameter, density, topology, etc. which makes knowledge of geometric descriptors important.

2.7 References

(1) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035. https://doi.org/DOI: 10.1021/ja00051a040.

(2) Love, O.; Galindo-Murillo, R.; Zgarbová, M.; Šponer, J.; Jurečka, P.; Cheatham, T. E. Assessing the Current State of Amber Force Field Modifications for DNA─2023 Edition. *J. Chem. Theory Comput.* **2023**, *19* (13), 4299–4307. https://doi.org/10.1021/acs.jctc.3c00233.

(3) MacKerell, A. D.; Banavali, N.; Foloppe, N. Development and Current Status of the CHARMM Force Field for Nucleic Acids. *Biopolymers* **2000**, *56* (4), 257–265. https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W.

(4) Pálinkás, G.; Kálmán, E.; Kovács, P. Liquid Water. *Mol. Phys.* **1977**, *34* (2), 525–537. https://doi.org/10.1080/00268977700101881.

(5) Daglar, H.; Keskin, S. Computational Screening of Metal-Organic Frameworks for Membrane-Based CO2/N2/H2O Separations: Best Materials for Flue Gas Separation. *J. Phys. Chem. C* **2018**, *122* (30), 17347–17357. https://doi.org/10.1021/acs.jpcc.8b05416.

(6) Basdogan, Y.; Keskin, S. Simulation and Modelling of MOFs for Hydrogen Storage. *CrystEngComm* **2015**, *17* (2), 261–275. https://doi.org/10.1039/c4ce01711k.

(7) Düren, T.; Bae, Y. S.; Snurr, R. Q. Using Molecular Simulation to Characterise Metal–Organic Frameworks for Adsorption Applications. *Chem. Soc. Rev.* **2009**, *38* (5), 1237–1247. https://doi.org/10.1039/b803498m.

(8) Ewald, P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253–287.

(9) Campañá, C.; Mussard, B.; Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **2009**, *5* (10), 2866–2878. https://doi.org/10.1021/ct9003405.

(10) Burtch, N. C.; Jasuja, H.; Dubbeldam, D.; Walton, K. S. Molecular-Level Insight into Unusual Low Pressure CO2 Affinity in Pillared Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2013**, *135* (19), 7172–7180. https://doi.org/10.1021/ja310770c.

(11) Sutrisno, A.; Terskikh, V. V.; Shi, Q.; Song, Z.; Dong, J.; Ding, S. Y.; Wang, W.; Provost, B. R.; Daff, T. D.; Woo, T. K.; Huang, Y. Characterization of Zn-Containing Metal-Organic Frameworks by Solid-State 67Zn NMR Spectroscopy and Computational Modeling. *Chem. - A Eur. J.* **2012**, *18* (39), 12251–12259. https://doi.org/10.1002/chem.201201563.

(12) Cho, H. S.; Tanaka, H.; Lee, Y.; Zhang, Y. B.; Jiang, J.; Kim, M.; Kim, H.; Kang, J. K.; Terasaki, O. Physicochemical Understanding of the Impact of Pore Environment and Species of Adsorbates on Adsorption Behaviour. *Angew. Chemie - Int. Ed.* **2021**, *60* (37), 20504–20510. https://doi.org/10.1002/anie.202107897.

(13) Makov, G.; Payne, M. Periodic Boundary Conditions in Ab Initio Calculations. *Phys. Rev. B* **1995**, *51* (7), 4014–4022. https://doi.org/10.1103/PhysRevB.51.4014.

(14) Gubernatis, J. E. Marshall Rosenbluth and the Metropolis Algorithm. *Phys. Plasmas* **2005**, *12* (5), 1–5. https://doi.org/10.1063/1.1887186.

(15) Lamia, N.; Jorge, M.; Granato, M. A.; Almeida Paz, F. A.; Chevreau, H.; Rodrigues, A. E. Adsorption of Propane, Propylene and Isobutane on a Metal-Organic Framework:

Molecular Simulation and Experiment. *Chem. Eng. Sci.* **2009**, *64* (14), 3246–3259. https://doi.org/10.1016/j.ces.2009.04.010.

(16) Wang, S. Comparative Molecular Simulation Study of Methane Adsorption in Metal - Organic Frameworks. *Energy and Fuels* **2007**, *21* (2), 953–956. https://doi.org/10.1021/ef060578f.

(17) Dzubak, A. L.; Lin, L. C.; Kim, J.; Swisher, J. A.; Poloni, R.; Maximoff, S. N.; Smit, B.; Gagliardi, L. Ab Initio Carbon Capture in Open-Site Metal-Organic Frameworks. *Nat. Chem.* **2012**, *4* (10), 810–816. https://doi.org/10.1038/nchem.1432.

(18) Paranthaman, S.; Coudert, F. X.; Fuchs, A. H. Water Adsorption in Hydrophobic MOF Channels. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 8123–8129. https://doi.org/10.1039/b925074c.

(19) Erucar, I.; Keskin, S. Computational Assessment of MOF Membranes for CH4/H2 Separations. *J. Memb. Sci.* **2016**, *514*, 313–321. https://doi.org/10.1016/j.memsci.2016.04.070.

(20) Datar, A.; Witman, M.; Lin, L. C. Improving Computational Assessment of Porous Materials for Water Adsorption Applications via Flat Histogram Methods. *J. Phys. Chem. C* **2021**, *125* (7), 4253–4266. https://doi.org/10.1021/acs.jpcc.0c11082.

(21) Smith, W.; Forester, T. R. DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graph.* **1996**, *14* (3), 136–141. https://doi.org/10.1016/S0263-7855(96)00043-4.

(22) Boyd, P. G. Computational High Throughput Screening of Metal Organic Frameworks for Carbon Dioxide Capture and Storage Applications, University of Ottawa, 2016.

(23) Kresse, G.; Furthmüller, J. Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mat. Sci.* **1996**, *6*, 15.

(24) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47* (1), 558–561. https://doi.org/10.1103/PhysRevB.47.558.

(25) Kresse, G.; Hafner, J. Ab Initio Molecular-Dynamics Simulation of the Liquid-Metalamorphous- Semiconductor Transition in Germanium. *Phys. Rev. B* **1994**, *49* (20), 14251–14269. https://doi.org/10.1103/PhysRevB.49.14251.

(26) Ferretti, A.; Calzolari, A.; Bonferroni, B.; Di Felice, R. Maximally Localized Wannier Functions Constructed from Projector-Augmented Waves or Ultrasoft Pseudopotentials. *J. Phys. Condens. Matter* **2007**, *19* (3). https://doi.org/10.1088/0953-8984/19/3/036215.

(27) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(28) Perdew, J. P.; Ernzerhof, M.; Burke, K. Rationale for Mixing Exact Exchange with Density Functional Approximations. *J. Chem. Phys.* **1996**, *105* (22), 9982–9985. https://doi.org/10.1063/1.472933.

(29) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50* (24), 17953–17979. https://doi.org/10.1103/PhysRevB.50.17953.

(30) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.

(31) Chong, S.; Lee, S.; Kim, B.; Kim, J. Applications of Machine Learning in Metal-Organic Frameworks. *Coord. Chem. Rev.* **2020**, *423*, 213487.

https://doi.org/10.1016/j.ccr.2020.213487.

(32) Raza, A.; Sturluson, A.; Simon, C. M.; Fern, X. Message Passing Neural Networks for Partial Charge Assignment to Metal-Organic Frameworks. *J. Phys. Chem. C* **2020**, *124* (35), 19070–19082. https://doi.org/10.1021/acs.jpcc.0c04903.

(33) Borboudakis, G.; Stergiannakos, T.; Frysali, M.; Klontzas, E.; Tsamardinos, I.; Froudakis, G. E. Chemically Intuited, Large-Scale Screening of MOFs by Machine Learning Techniques. *npj Comput. Mater.* **2017**, *3* (1), 1–6. https://doi.org/10.1038/s41524-017-0045-8.

(34) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO 2 Adsorption Properties of Metal–Organic Frameworks. *J. Phys. Chem. C* **2020**, acs.jpcc.0c06334. https://doi.org/10.1021/acs.jpcc.0c06334.

(35) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016**, *2* (March), 1–7. https://doi.org/10.1038/npjcompumats.2016.28.

(36) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation. *J. Chem. Inf. Model.* **2021**, *61* (5), 2131–2146. https://doi.org/10.1021/acs.jcim.1c00191.

(37) Anderson, R.; Rodgers, J.; Argueta, E.; Biong, A.; Gómez-Gualdrón, D. A. Role of Pore Chemistry and Topology in the CO2 Capture Capabilities of MOFs: From Molecular Simulation to Machine Learning. *Chem. Mater.* **2018**, *30* (18), 6325–6337. https://doi.org/10.1021/acs.chemmater.8b02257.

3 Computational Study of $Ni_3(pzdc)_2(ade)_2$

3.1 Abstract

The work done on this project was in collaboration with Professor Stylianou's lab from Oregon State University to conduct a computational study on a MOF that they had developed, $[Ni_3(pzdc)_2(ade)_2(H_2O)_4]$(pyrazole-3,5-dicarboxylic acid (H_3pzdc), and adenine (ade)). Their work was able to experimentally determine properties of the MOF such as open Ni(II) sites were generated, a surface area of 160 m^2/g was calculated with the Brunauer-Emmett-Teller (BET) method, an isotherm at 77K demonstrated high H_2 uptake at low pressures, and a 2.3 ± 0.2 kcal/mol binding energy. The goal of this author's work was to determine the activated structure of $[Ni_3(pzdc)_2(ade)_2(H_2O)_4]$ and to determine the mechanism of binding through computational simulations. The activated structure was determined computationally by a comparison to the experimental PXRD patterns with 15 Å^3 smaller volume than the non-activated structure. It was determined that the method of binding was physisorption through DFT structure optimization. To study the binding further, the original structure had its metal centre swapped and hydrogen adsorption simulations were run to compare uptake. The findings showed that the pore size and shape contributed more to uptake than metal centre. This work has been published in ACS Applied Materials & Interfaces in the paper "Hydrogen Adsorption in Ultramicroporous Metal−Organic Frameworks Featuring Silent Open Metal Sites" where I performed all computational calculations presented.[1]

3.2 Introduction

Excess use of fuels that emit greenhouse gases has led to the damage of the environment requiring a switch to fuels sources that have benign products. Hydrogen has been considered a possible alternative to traditional fuels due to the absence of greenhouse gases production and having a high gravimetric energy density (142 MJ/kg).[2] The ability to store hydrogen at useful volumetric densities is a challenge that has yet to be solved as working with gases is difficult to do without using a pressurized environment that can be expensive and dangerous.[3]

Porous solids such as carbon nanotubes, zeolites, and MOFs, can be used to store gases.[4] A porous solid would be able to store a gas due to the accessible pores densifying the gas in a safer way than a pressurized container. When determining the viability of a porous solid for a specified application, it is important to understand the interactions between the solid and the guests.[5] This understanding allows researchers to rationally design and synthesize materials for specific uses.[6–8]

It is important to go over properties of a porous solid that lead to good adsorption when selecting a material for H_2 storage. An ideal porous solid would have a reversible adsorption-desorption cycle with high hydrogen uptake and operate in a safe temperature and pressure range. Ideally the binding mode would be through physisorption rather than chemisorption due to the larger energy requirements to remove H_2.[9] Hydrogen uptake at 77 K and 1 bar is the most used method to determine hydrogen uptake which will give results that can be easily referenced against other materials. Deliverable adsorption capacity is the difference in the amount of gas remaining in a porous solid at lowest desorption pressure and the amount of gas in the solid at storage pressure.[10] Other important properties are heat of adsorption, charge and discharge kinetics, and reusability. It is important to understand the effects pore structure have on hydrogen adsorption to understand why a material is a strong candidate for adsorption.

Previous studies into uptake of hydrogen in porous materials shows a trend between the surface area and adsorption capacity. This trend is referred to as Chahine's rule where the excess gravimetric capacity increases as the surface area increases[9] (1 wt% per 500m^2/g, peaks at

$3000 m^2/g$). A study of carbide-derived carbons with controllable pore sizes demonstrated a deviance in this trend. The pore size has also shown shows effect on the uptake as smaller pores, 6-7 Å, have higher hydrogen uptake than pores beyond 7 Å.[11–13]

MOFs provide a great platform to optimize the above properties due to their tunability when being designed. Pore size and surface area can be optimized by searching through a series of metal centres and organic linkers to create a MOF with varying geometric parameters optimal for different purposes. Further functionalization through the use of open metal centres and the addition of functional groups can increase the binding affinity for hydrogen to a MOF. Previous studies show MOFs with pores less than 7 Å have a high affinity for H_2.

My work is the computational study on a MOF that was experimentally synthesized and characterized by the Materials Discovery Lab at Oregon State University for its notable H_2 adsorption at low pressures. The material is composed of Ni(II), pyrazole-3,5-dicarboxylic acid (H_3pzdc), and adenine (ade) to form $[Ni_3(pzdc)_2(ade)_2]$, a 3D porous MOF with pores smaller than 7 Å shown in Figure 3.1. The Woo lab was contacted to perform a computational study of the material to examine the interactions between the MOF and the guest to understand the high uptake of hydrogen in the MOF. The material's uptake was studied at 77 K and 1 bar and geometric properties were determined computationally and experimentally. Below is this author's work on the computational study of BFF to explain the high uptake of hydrogen in the MOF.

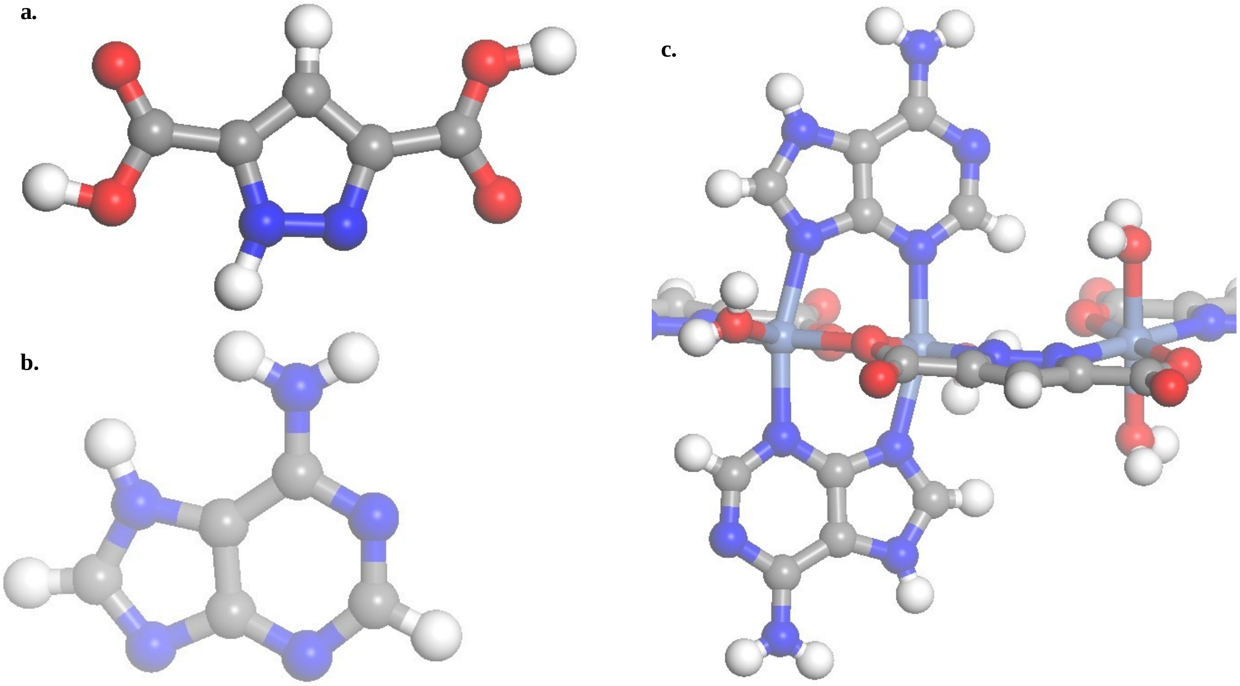

Figure 3.1: The SBUs of BFF, (a) pyrazole-3,5-dicarboxylic acid and (b) adenine, with the completed structure of $[Ni_3(pzdc)_2(ade)_2(H_2O)_4]$ (c). Atoms are coloured as follows: N-dark blue, H-white, O-red, C-grey, Ni-light blue.

3.3 Computational Methods

The DFT calculations were performed with the Vienna Ab initio Simulation Package (VASP) version 5.4.4[14], with the PBE[15] functional and the projector augmented-wave[16] pseudopotentials. The dispersion correction used was the D3 correction by Grimme et al.[17] for all DFT calculations. All calculations were done unrestricted. A planewave cutoff of 400 eV was used for fixed unit cell calculations and 520 eV was used when the unit cell was geometry optimized.

Two separate GCMC simulation programs were used for different purposes. An in house Grand Canonical Monte Carlo (GCMC) simulation program developed by the Woo lab group called FastMC[18] was used. Another group developed program called GALA[19], was used to identify the binding sites of the guest hydrogen molecules from the probability distributions generated by the GCMC simulations. The Lennard-Jones potential accounted for the steric and dispersion interactions and the framework parameters were gathered from the UFF force field[20] whereas the guest H_2 parameters used a five site model from Belof et al.[21] The fixed partial atomic charge method was used to model the electrostatic interactions and REPEAT[22] was used to assign the partial charges based on single-point DFT calculations of the MOF. Adsorption isotherms for H_2 were performed at 77.0 K for BFF. For the FastMc calculations, 320000 MC steps were used per isotherm point with a super cell of 4 X 3 X 3 with a Lennard-Jones cut off radius of 12.5 Å.

Zeo ++[23] is a program that can rapidly compute geometric properties of porous structures. In this study version 0.3.0 of Zeo ++ was used and all calculations used the high accuracy flag and a probe radius of 1.2 Å to resemble a hydrogen molecule.

3.4 Experimental Results

Understanding the MOF's experimental properties are important to understanding the results from the computational study. This section will briefly review the experimental work done by our collaborators from the Stylianou's lab to create, characterize and activate the MOF.[1] Synthesis was performed by mixing $NiCO_3$, H_3pzdc and ade in water and heated to 130 ^{0}C under reflux for 4 hours. The crystal structure was obtained by single crystal X-ray diffraction. 1D chains are formed from the ade and pzdc^{3-} ligands coordinating to Ni^{2+} atoms shown in Figure 3.2 (Ni_A, Ni_B). These 1D chains create a zig-zag pattern that is porous due to the inefficient packing of the chains with voids that account for 19.1% of the unit cell volume calculated by Olex2.[24]

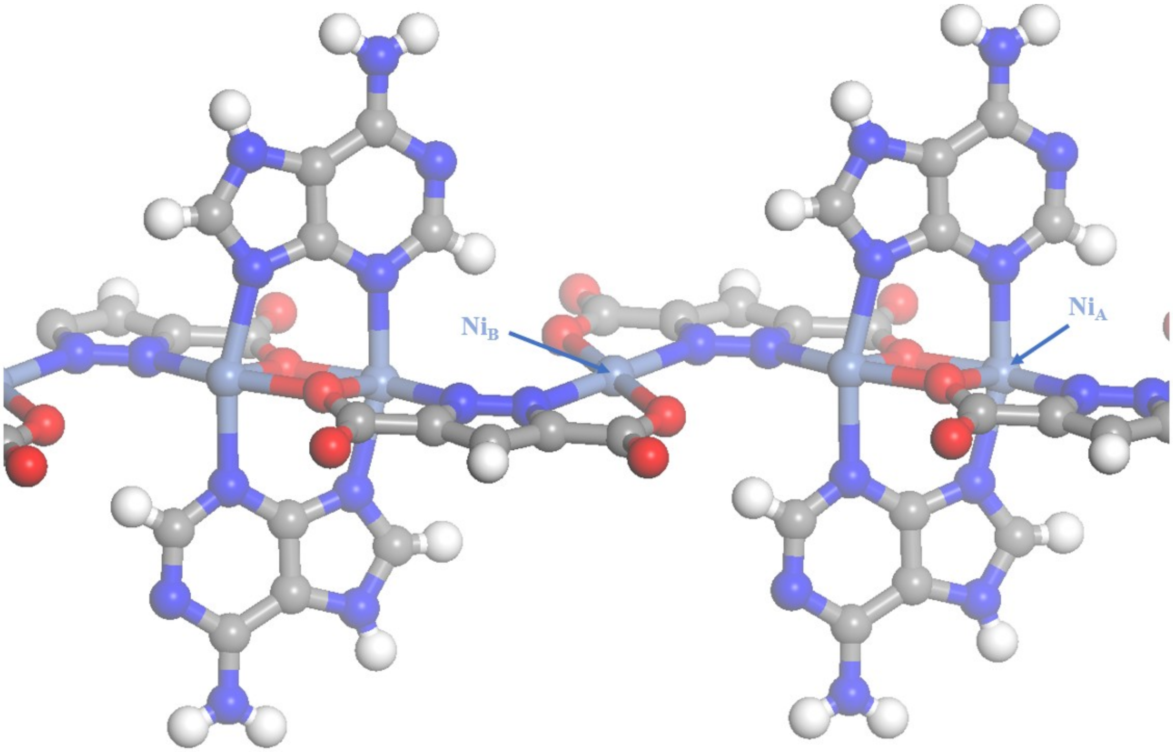

Figure 3.2: Structure of [$Ni_3(pzdc)_2(ade)_2$] one-dimensional chains. Atoms are coloured as follows: N-dark blue, H-white, O-red, C-grey, Ni-light blue.

The material is activated by exposing it to vacuum (10^{-4} mbar) at 100 ^{0}C, [$Ni_3(pzdc)_2(ade)_2(H_2O)_4$]·2.18$H_2O$ loses all of water molecules in both the pores and bound to the Ni^{2+} atoms, leaving the nickel binding sites unsaturated. Powder X-ray diffraction patterns were taken for both the activated and unactivated structure with a Rigaku miniFlex benchtop X-ray diffractometer. The structural change in the material is seen as the activated structure's PXRD pattern seen in green in Figure 3.3 is different from the unactivated pattern in pink. A nitrogen isotherm was performed at 77K and 1 bar with a Micrometrics 3Flex apparatus. demonstrated the permanent porosity and was used to calculate the BET geometric properties; a surface area of 160 m^2/g, pore volume of 0.0769 cm^3/g and a pore size of 6.7 Å.

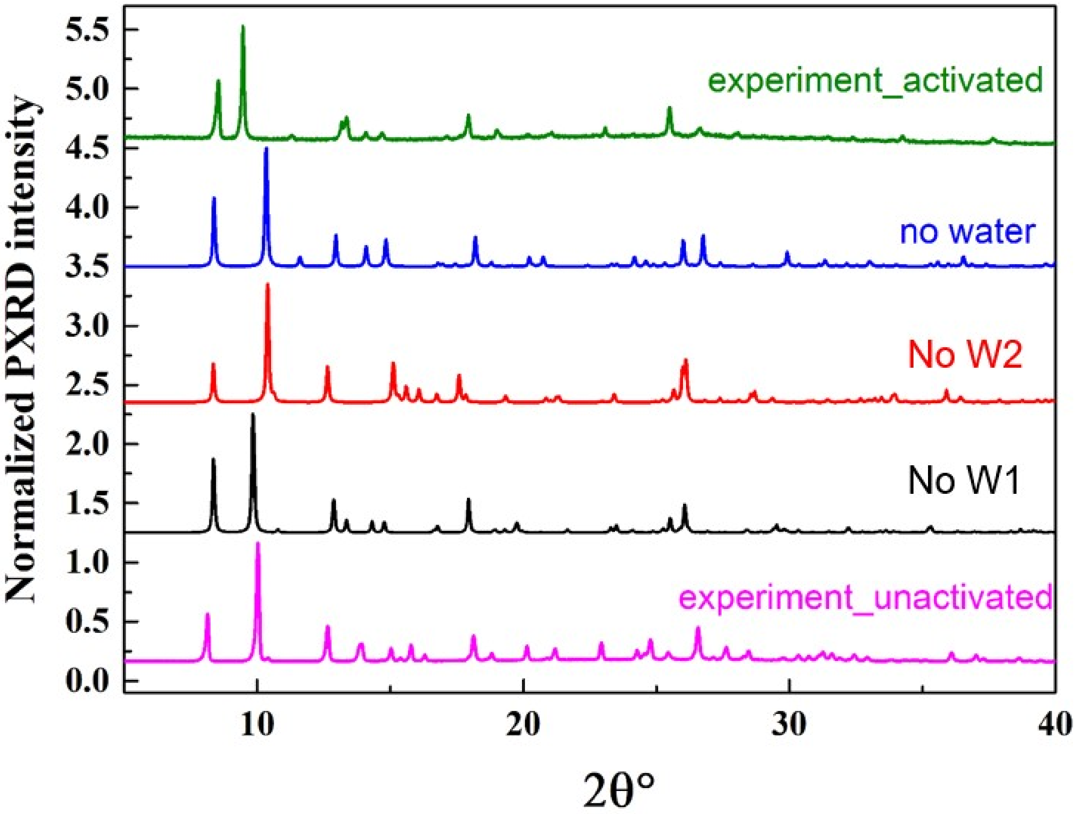

Figure 3.3: Computationally generated PXRD patterns (no W1, no W2, no water) compared to experimentally acquired PXRD patterns (experiment_activated, experiment_unactivated).[1]

A hydrogen isotherm, shown in Figure 3.4 (a), experimentally performed at 77 K and 0.8 bar shows the uptake saturation at ~ 0.15 bar and 2.11 mmol/g. This isotherm was performed with a closed-cycle helium refrigerated cryostat. The heat of adsorption (Q_{st}) was determined by taking the isotherm at separate temperatures, Figure 3.4 (b), and using the Clausius-Clapeyron equation to get a Q_{st} of 2.3 ± 0.24 kcal/mol.

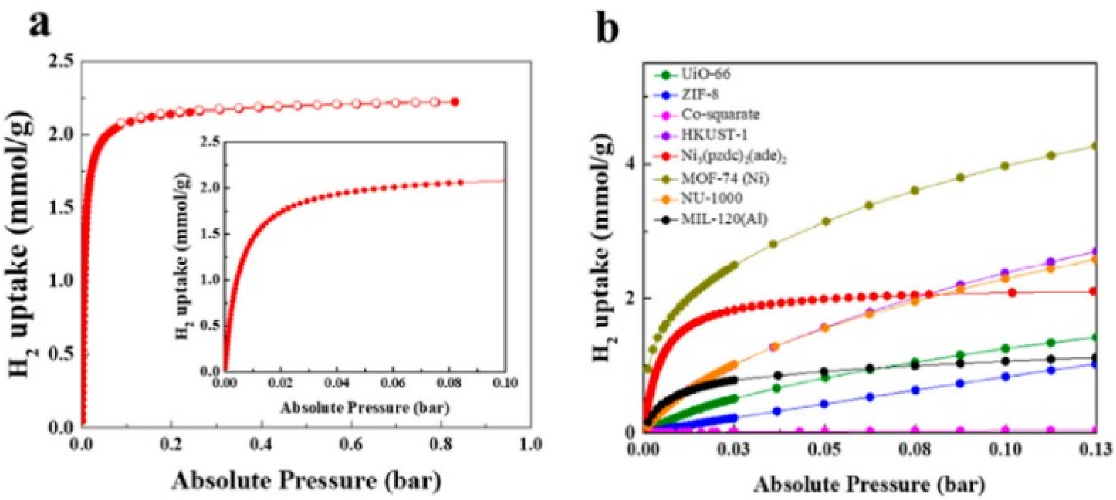

Figure 3.4: These isotherms are all experimentally produced by Chiu *et al.* "(a) H_2 adsorption isotherm at 77 K and 0.8 bar of [$Ni_3(pzdc)_2(ade)_2$]; inset: A zoom-in of the low-pressure H_2 adsorption (0 to 0.1 bar)." (b) "Comparison of H_2 adsorption of [$Ni_3(pzdc)_2(ade)_2$] with other benchmark MOFs at 77 K in the low-pressure region of up to 0.13 bar. Filled symbols: adsorption; empty symbols: desorption." Reprinted with permission from [1]. Copyright 2023 American Chemical Society.

In comparison to other materials, BFF has a higher binding energy than porous carbon materials and MOFs lacking open metal sites with Q_{st} between 1.0-1.7 kcal/mol,[25] and a slightly lower binding energy than MOFs with open metal sites with Q_{st} between 2.4-2.9 kcal/mol.[26] Low pressure hydrogen uptake was also compared to other benchmark MOFs which were synthesized by the Materials Discovery Lab listed in Figure 3.4 (b). BFF outperformed all MOFs besides MOF-74-Ni(II) for uptake at low pressure with MOF-74's open metal sites believed to be the reason for the higher uptake.

It is important to determine the binding mode to further understand the high hydrogen uptake of BFF. Carbon monoxide was loaded into BFF due to its ability to π-bond unsaturated metal centres[27,28] and the structure was analysed by Fourier transform infrared (FT-IR) spectroscopy. New peaks appear on the FT-IR spectrum, Figure 3.5 (a), as CO is loaded at 2134 and 2124 cm^{-1} which align with CO bonds on Ni(II) sites.[28,29] The two separate bands are due to CO interacting with two different metal sites and not due to different stretches in a theoretical Ni(II)

$(CO)_2$ group due to changes in saturation behaviour. The same experiment was performed with H_2 being loaded in place of CO. The associated peaks from H-H stretching and Ni-H_2 interactions at 4125 and 4035 cm^{-1}[30,31] do not appear in Figure 3.5 (b). The lack of binding peaks suggests the open metal sites do not bind to the guest H_2 molecules, and the strong adsorption is due to the optimal pore size of BFF.

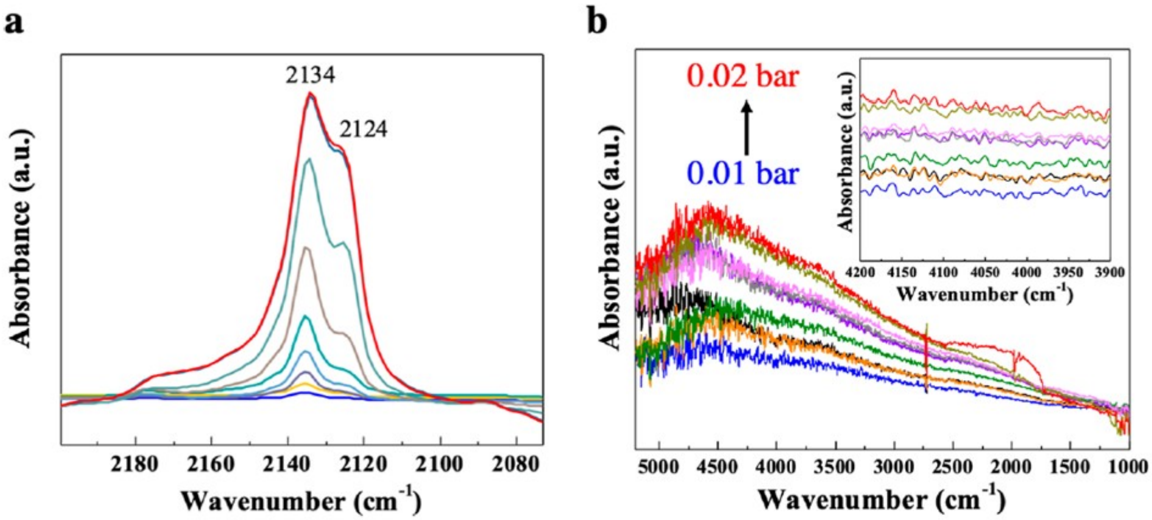

Figure 3.5: These Fourier transform infrared spectra are all experimentally produced by Chiu *et al* . “In situ static transmission FT-IR spectra during (a) dosing of CO at pressure from 0 to 6.6 mbar at 298 K and (b) dosing of H_2 from 10 to 20 mbar at 77 K. Two well-resolved peaks (2134 and 2124 cm^{-1}) appeared with increasing CO dosing, which correspond to CO bound to the Ni open metal sites forming two different Ni(II)(CO) complexes. No peaks are observed during H_2 dosing, indicating that H_2 does not bind with the open Ni^{2+} sites in [$Ni_3(pzdc)_2(ade)_2$] with the formation of Ni(II)−H_2 species.” Reprinted with permission from [1]. Copyright 2023 American Chemical Society.

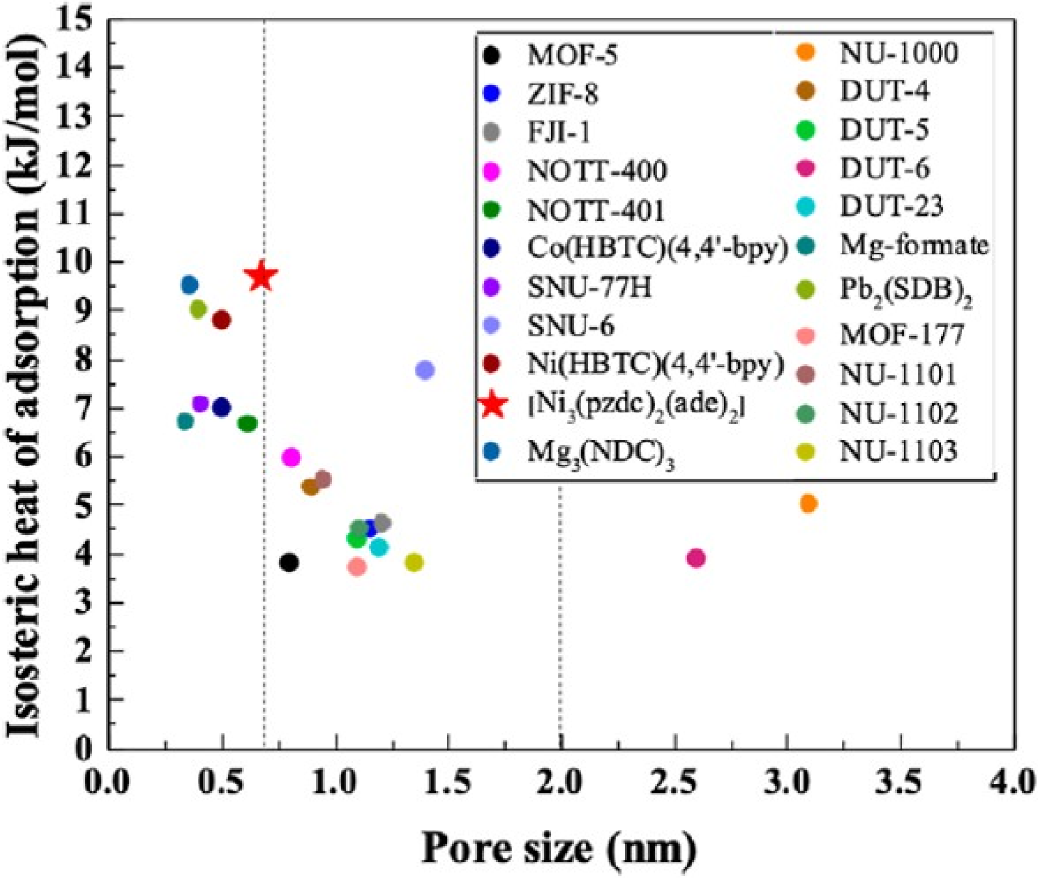

Figure 3.6: Heats of adsorption plotted against pore size produced by Chiu *et al.* "Isosteric heat of adsorption (Q_{st}) is plotted as a function of the pore size for Ni$_3$(pzdc)$_2$(ade)$_2$] and MOFs without open metal sites." Reprinted with permission from [1]. Copyright 2023 American Chemical Society.

Since the hydrogen is not considered bonded to the nickel, a comparison can be made to the Q_{st} of MOFs which do not feature open metal sites. Figure 3.6 demonstrates that MOFs without open metal sites with pores less than 7 Å have high affinities to H_2. BFF features the highest heat of adsorption for a MOF without open metal centres.

3.5 Results and Discussion

To complement the experimental results, a computational study was performed to understand the mode of binding.

Table 3.1: Comparison of BFF unit cell parameters derived from the experimental unactivated results, the experimental activated results and the computational activated results.

Unit Cell Parameter	Unactivated	Experimental Activated	Computational Activated
a (Å)	7.332	7.437	7.419
b (Å)	9.916	10.049	9.913
c (Å)	11.832	10.898	11.658
α (°)	67.278	97.191	67.888
β (°)	80.572	76.263	73.138
γ (°)	72.566	72.011	71.898
Unit cell Volume (Å^3)	755.8913	735.00	740.2532

The crystal structure of BFF was determined in its unactivated form. This structure is then activated, a process that ideally involves removing solvent from the MOF without damaging the structure. TGA (thermogravimetric analysis) concluded that during activation, BFF loses mass equivalent to two water molecules per unit cell of the MOF. The MOF's activated crystal structure could not be determined, only its PXRD pattern, therefore DFT calculations were used to help determine the activated structure. We were initially given unit cell parameters for the activated structure based on the PXRD pattern, however the experimental activated unit cell was very different from the unactivated structure as seen in Table 3.1, particularly the α angle. It was decided to let VASP relax the unactivated unit cell parameters upon geometry optimization to compare to the experimental value. This was labeled as the computational activated unit cell parameters and were used going forward.

Since there were four water molecules bound in the unit cell of BFF but only two were removed, DFT optimizations of the structure were performed to determine the most stable structure with two water molecules removed. Figure 3.7 shows the unactivated BFF structure with the four coordinated water molecules circled bound to their respective nickel atoms. Ni2 features two water molecules bound to it (W2 and W2') and Ni1 has one water molecule bound to it (W1) with a mirror image (W1') leaving four coordinated water molecules per unit cell. There are six total combinations to test for with the two water molecules remaining: W1/W1', W2/W2', W1/W2, W1/W2', W1'/W2 and W1'/W2'. VASP was used to optimize each combination with two sets of calculations; one where the unit cell parameters the same as the experimental activated unit cell and one where the unit cell parameters are relaxed by VASP. The most energetically stable combination, as seen in Table 3.2, was when the two waters were removed from Ni2, and the unit cell was relaxed as it is 20 kcal/mol lower in energy than the other combinations. Ni2 has an oxidation state of 2+, matching the structure where it is in a square planar configuration. When the waters are removed from Ni1 and Ni1' it develops an odd geometric shape leading it to be the least energetically stable. For these reasons, it can be expected that removing the non-symmetric water solvents (W1/W2, W1/W2', W1'/W2 and W1'/W2') will not result in stronger energetic stability which is confirmed by the results. Figure 3.3 shows the PXRD patterns for relevant experimental and computational structures. Stylianou's lab compared the computational PXRD results to experimental to determine that the W2/W2' removed PXRD pattern aligns with the experimental activated PXRD pattern. For these reasons, this form of BFF was accepted as the activated structure and was used from here on out to perform simulations. The activated structure unit cell parameters are in Table 3.1.

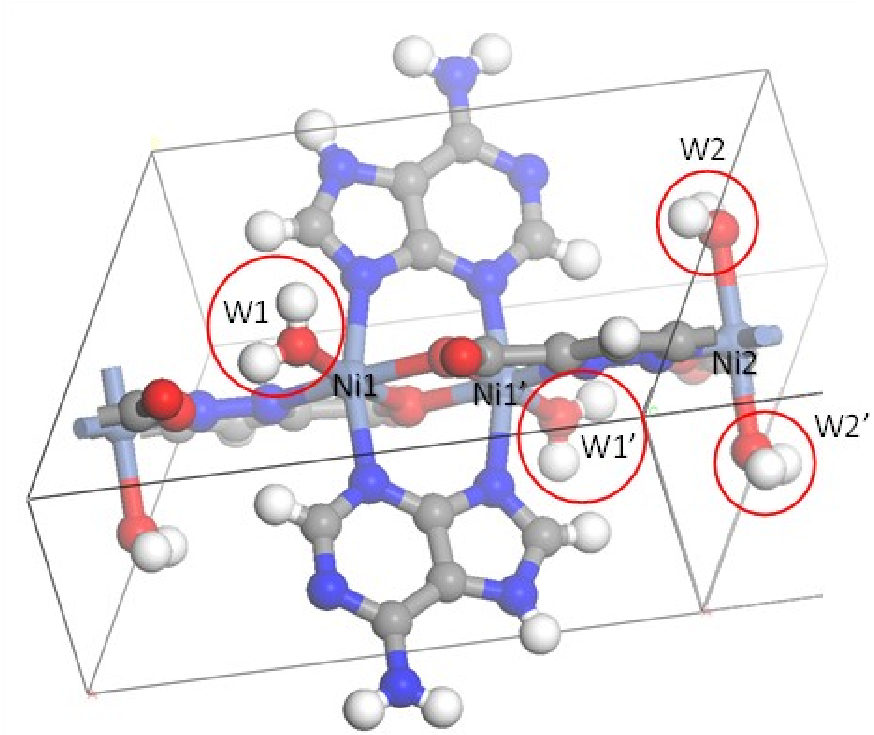

Figure 3.7: Crystal structure of **BFF** with the coordinated water molecules circled in red. Atoms are coloured as follows: N-dark blue, H-white, O-red, C-grey, Ni-light blue.

Table 3.2: Relative energy of different BFF isomers determined by DFT calculations. Results separated by using the experimental unit cell parameters and the DFT calculated unit cell parameters.

Water removed	W1* /W2*	W1 /W2*	W1* /W2	W1 /W2	W1	W2
Experimental Relative Energy (kcal/mol)	-29.65	-23.85	-23.88	-25.07	-52.27	0
Computational Relative Energy (kcal/mol)	-16.53	-5.43	-5.26	-3.46	0	-28.46

Hydrogen adsorption GCMC simulations were performed on the activated BFF structure with the parameters specified in section 3.3 and the results are shown in Figure 3.9. The dark blue simulated H_2 isotherm compares well with the experimental isotherm in brown. Though the simulated isotherm overestimated the saturation uptake of H_2 there is good agreement between the two plots. Heats of adsorption were also calculated with the simulated value being 2.15 kcal/mol which is in good agreement with the experimental value being 2.3 kcal/mol.

Experimental evidence suggested that the H_2 was physically binding in BFF rather than chemically and with good agreement in the simulated and experimental isotherms, we aimed to determine the location of the binding sites from the GCMC simulations. The GCMC simulations not only estimate the gas adsorption at a given pressure, they also provide probability distribution maps of where the guest molecules are in the framework. However, the probability distributions are noisy, with many local maxima and isolating the binding sites from the raw distributions can be difficult. The Woo lab has developed a tool called GALA that smooths the probability distributions and can identify the binding sites in an automated fashion. GALA determined the H_2 binding sites from the GCMC simulations to the centre of the pore above the square planar nickel, as shown in Figure 3.8. We also investigated the nature of the H_2 binding sites through DFT calculations as they are considered more accurate than the force fields used in the GCMC simulations. These simulations involved placing an H_2 molecule close to the open metal centre, ~ 2 Å away from the Ni2 site, and optimizing the guest atoms' positions. All of these simulations led to the guest returning to a position close to where GALA had predicted the binding site, ~ 4 Å away from the nearest open metal centre. This distance is beyond what is expected for hydrogen chemisorption onto an open nickel site, between 1.4 – 1.7 Å.[32,33] There was one exception where the guest hydrogen was tested as separate atoms near the metal site where a single atom remains near the nickel but, the over all structure had a higher energy than when the hydrogen was a molecule. Thus, the molecule is considered physiosorbed and not chemisorbed.

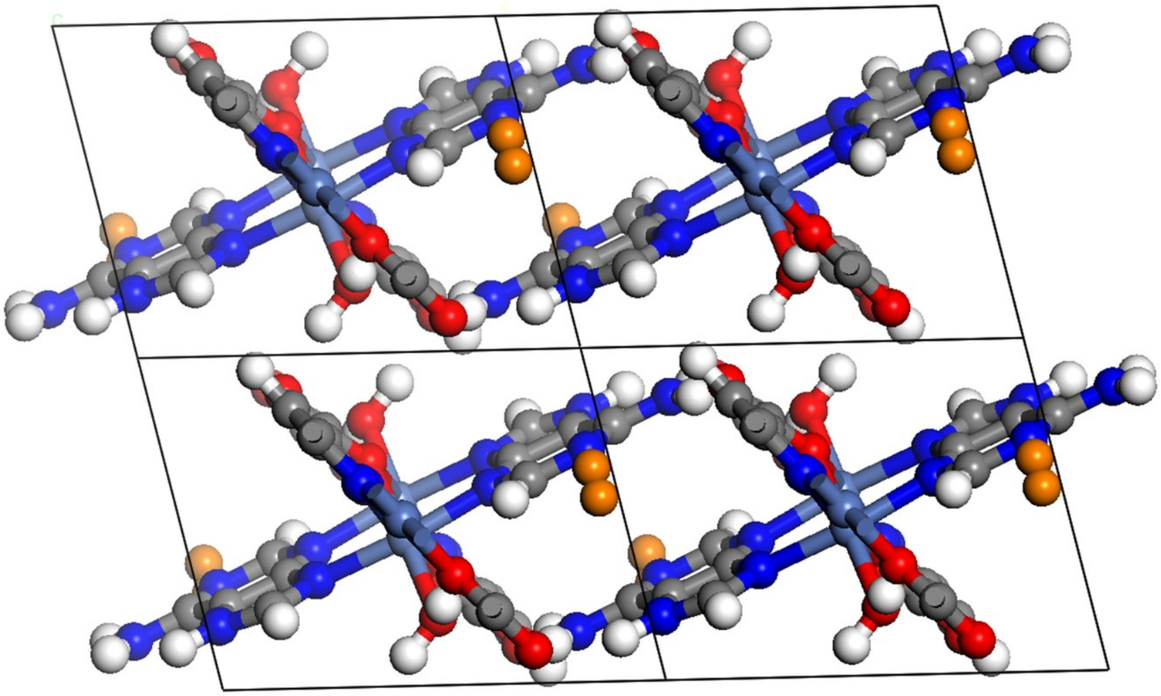

Figure 3.8: View of the activated form of BFF showing the H_2 binding site (orange) identified from GCMC simulations.

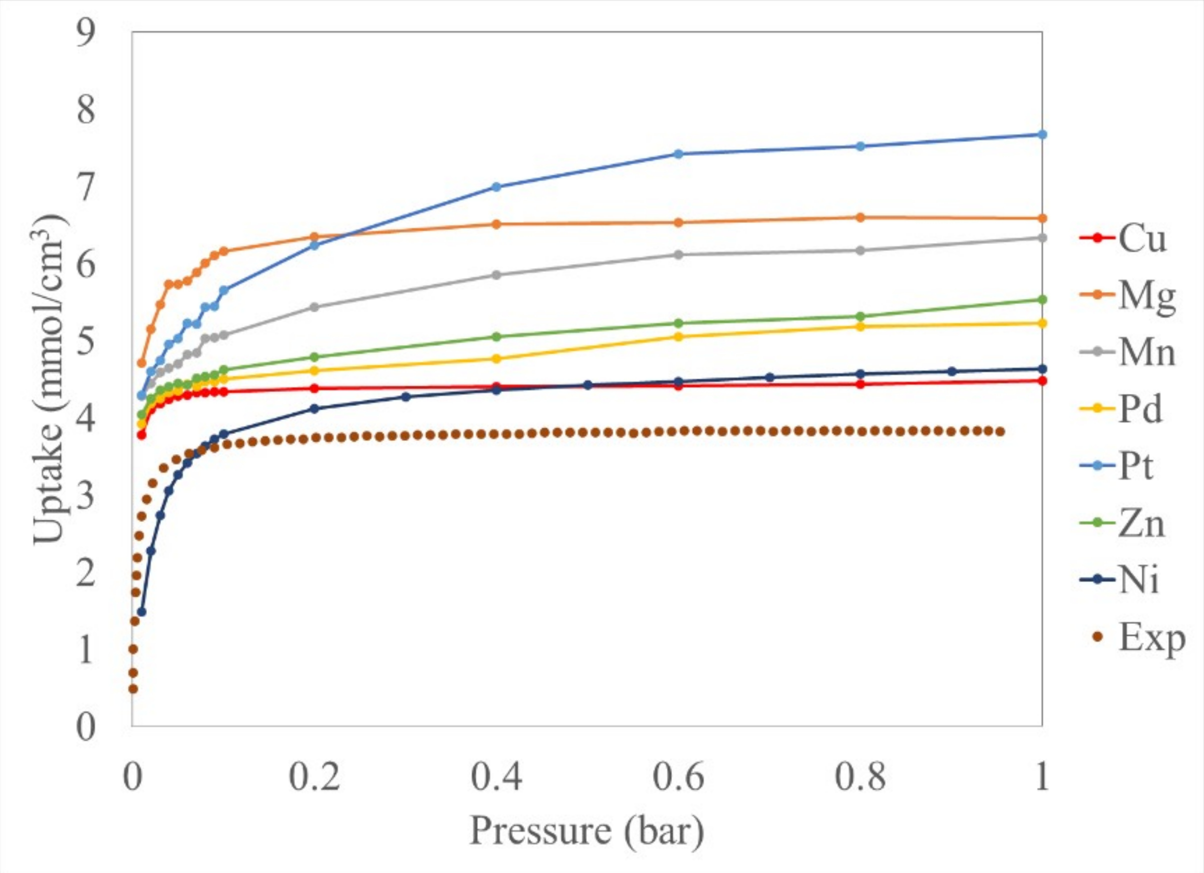

Figure 3.9: GCMC simulated and experimental H_2 gas adsorption isotherms at 77 K and 1 bar of **BFF** compared to GCMC simulated isotherms of various metal substituted **BFF** analogues.

Other metals were examined to see their effects on uptake in the structure and to highlight promising structures for hydrogen storage. To do this, all nickel atoms were replaced with metals that could also adopt a square planar configuration with a neutral charge, these atoms being Pd, Pt, Cu, Zn. Each structure was then geometry optimized for both positions and unit cell parameters at the DFT level.

Table 3.3 shows the geometric properties calculated by Zeo ++, and Figure 3.9 shows the GCMC simulated isotherms of the DFT optimized metal swapped MOFs. Figure 3.9 shows all the metal swaps show the same steep uptake around 0.04 bar further suggesting that the uptake is due to the pore structure. Platinum is notable for having the highest saturation uptake. Platinum being the largest atom, has the smallest pore diameter seen in table 3.3. Zeo ++ calculates the pore sizes based on a static representation of the MOF, therefore even if the pore limiting diameter is smaller than the hydrogen molecule it is possible for hydrogen to enter the MOF. The metal swaps feature similar pore diameters however there is a fairly large difference in accessible surface areas. Further experimental study would need to be done to determine whether these MOFs can be made in lab.

Table 3.3: Comparison of geometric properties of activated BFF upon metal atom substitution by analogous transition metals using a probe radius of 1.2 Å.

Metal	Density (g/cm^3)	Unit Cell Volume (Å^3)	Largest Cavity Diameter (Å)	Pore Limiting Diameter (Å)	Accessible Gravimetric Surface Area (m^2/g)
Ni (BFF)	1.77	740.3	4.0	2.9	473.3
Cu	1.79	745.6	3.5	2.6	433.5
Pd	2.01	771.6	4.3	2.9	407.0
Pt	2.65	750.3	3.3	2.1	0
Zn	1.73	775.9	4.4	2.9	550.0

3.6 Conclusion

The computational study of BFF performed provided insight into the nature of hydrogen adsorption of BFF. The activated structure was determined through DFT geometry optimization with two of the four water molecules removed. The lowest energy structure was significantly more stable than other structures and was therefore used as the structure for the activated complex. GCMC simulations were then performed using this activated complex to simulate the hydrogen gas adsorption isotherm. The computed and experimental isotherms in figure 3.9 were in agreement with one another, with the computed isotherm slightly over-estimating the saturation uptake. Importantly, the sharp uptake at low pressure was reproduced. The computed heat of adsorption was also in excellent agreement with the experimental value. The H_2 binding sites were then extracted from the GCMC simulations, showing that the H_2 binding site is not near the metal centres near the centre of the pore. DFT calculations on the binding sites confirmed this result. Calculations on replacing Ni in BFF were also performed. Isotherm simulations suggested that other metals in a BFF like structure would have a higher H_2 adsorption capacity.

3.7 References

(1) Chiu, N. C.; Compton, D.; Gładysiak, A.; Simrod, S.; Khivantsev, K.; Woo, T. K.; Stadie, N. P.; Stylianou, K. C. Hydrogen Adsorption in Ultramicroporous Metal-Organic Frameworks Featuring Silent Open Metal Sites. *ACS Appl. Mater. Interfaces* **2023**. https://doi.org/10.1021/acsami.3c12139.

(2) Mazloomi, K.; Gomes, C. Hydrogen as an Energy Carrier: Prospects and Challenges. *Renew. Sustain. Energy Rev.* **2012**, *16* (5), 3024–3033. https://doi.org/10.1016/j.rser.2012.02.028.

(3) Schlapbach, L.; Züttel, A. For Mobile Applications. *Nature* **2001**, *414* (November), 353–358. https://doi.org/10.1038/35104634.

(4) Zhao, D.; Wang, X.; Yue, L.; He, Y.; Chen, B. Porous Metal-Organic Frameworks for Hydrogen Storage. *Chem. Commun.* **2022**, *58* (79), 11059–11078. https://doi.org/10.1039/d2cc04036k.

(5) Lorzing, G. R.; Gosselin, A. J.; Trump, B. A.; York, A. H. P.; Sturluson, A.; Rowland, C. A.; Yap, G. P. A.; Brown, C. M.; Simon, C. M.; Bloch, E. D. Understanding Gas Storage in Cuboctahedral Porous Coordination Cages. *J. Am. Chem. Soc.* **2019**, *141* (30), 12128–12138. https://doi.org/10.1021/jacs.9b05872.

(6) Chiu, N. C.; Loughran, R. P.; Gładysiak, A.; Vismara, R.; Park, A. H. A.; Stylianou, K. C. Wet Flue Gas CO2 Capture and Utilization Using One-Dimensional Metal-Organic Chains. *Nanoscale* **2022**, *14* (40), 14962–14969. https://doi.org/10.1039/d2nr04156a.

(7) Idrees, K. B.; Li, Z.; Xie, H.; Kirlikovali, K. O.; Kazem-Rostami, M.; Wang, X.; Wang, X.; Tai, T. Y.; Islamoglu, T.; Stoddart, J. F.; Snurr, R. Q.; Farha, O. K. Separation of Aromatic Hydrocarbons in Porous Materials. *J. Am. Chem. Soc.* **2022**, *144* (27), 12212–12218. https://doi.org/10.1021/jacs.2c03114.

(8) Denning, S.; Majid, A. A. A.; Lucero, J. M.; Crawford, J. M.; Carreon, M. A.; Koh, C. A. Metal-Organic Framework HKUST-1 Promotes Methane Hydrate Formation for Improved Gas Storage Capacity. *ACS Appl. Mater. Interfaces* **2020**, *12* (47), 53510–53518. https://doi.org/10.1021/acsami.0c15675.

(9) Chen, Y. P.; Bashir, S.; Liu, J. L. Nanostructured Materials for Next-Generation Energy Storage and Conversion: Hydrogen Production, Storage, and Utilization. *Nanostructured Mater. Next-Generation Energy Storage Convers. Hydrog. Prod. Storage, Util.* **2017**, 1–349. https://doi.org/10.1007/978-3-662-53514-1.

(10) Gómez-Gualdrón, D. A.; Wang, T. C.; García-Holley, P.; Sawelewa, R. M.; Argueta, E.; Snurr, R. Q.; Hupp, J. T.; Yildirim, T.; Farha, O. K. Understanding Volumetric and Gravimetric Hydrogen Adsorption Trade-off in Metal-Organic Frameworks. *ACS Appl. Mater. Interfaces* **2017**, *9* (39), 33419–33428. https://doi.org/10.1021/acsami.7b01190.

(11) Gogotsi, Y.; Dash, R. K.; Yushin, G.; Yildirim, T.; Laudisio, G.; Fischer, J. E. Tailoring of Nanoscale Porosity in Carbide-Derived Carbons for Hydrogen Storage. *J. Am. Chem. Soc.* **2005**, *127* (46), 16006–16007. https://doi.org/10.1021/ja0550529.

(12) Yushin, G.; Dash, R.; Jagiello, J.; Fischer, J. E.; Gogotsi, Y. Carbide-Derived Carbons: Effect of Pore Size on Hydrogen Uptake and Heat of Adsorption. *Adv. Funct. Mater.* **2006**, *16* (17), 2288–2293. https://doi.org/10.1002/adfm.200500830.

(13) Gogotsi, Y.; Portet, C.; Osswald, S.; Simmons, J. M.; Yildirim, T.; Laudisio, G.; Fischer, J. E. Importance of Pore Size in High-Pressure Hydrogen Storage by Porous Carbons. *Int. J. Hydrogen Energy* **2009**, *34* (15), 6314–6319.

https://doi.org/10.1016/j.ijhydene.2009.05.073.
(14) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **1996**, *54* (16), 11169. https://doi.org/10.1103/PhysRevB.54.11169.
(15) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
(16) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758.
(17) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15). https://doi.org/Artn 154104Doi 10.1063/1.3382344.
(18) Boyd, P. G. Computational High Throughput Screening of Metal Organic Frameworks for Carbon Dioxide Capture and Storage Applications, University of Ottawa, 2016.
(19) Burns, T. D. Pores to Process: The In Silico Study of Metal-Organic Frameworks from Crystal Structure to Industrial Pressure Swing Adsorption for Postcombustion Carbon Capture and Storage. **2022**.
(20) Rappe, A. K.; Casewit, C. J. J.; Colwell, K. S. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035. https://doi.org/10.1021/ja00051a040.
(21) Belof, J. L.; Stern, A. C.; Space, B. An Accurate and Transferable Intermolecular Diatomic Hydrogen Potential for Condensed Phase Simulation. *J. Chem. Theory Comput.* **2008**, *4* (8), 1332–1337. https://doi.org/10.1021/CT800155Q.
(22) Campañá, C.; Mussard, B.; Woo, T. K. Electrostatic Potential Derived Atomic Charges for Periodic Systems Using a Modified Error Functional. *J. Chem. Theory Comput.* **2009**, *5* (10), 2866–2878. https://doi.org/10.1021/ct9003405.
(23) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.
(24) Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H. OLEX2: A Complete Structure Solution, Refinement and Analysis Program. *J. Appl. Crystallogr.* **2009**, *42* (2), 339–341. https://doi.org/10.1107/S0021889808042726.
(25) Hirscher, M.; Panella, B. Hydrogen Storage in Metal-Organic Frameworks. *Scr. Mater.* **2007**, *56* (10), 809–812. https://doi.org/10.1016/j.scriptamat.2007.01.005.
(26) Sumida, K.; Brown, C. M.; Herm, Z. R.; Chavan, S.; Bordiga, S.; Long, J. R. Hydrogen Storage Properties and Neutron Scattering Studies of Mg2(Dobdc)—a Metal-Organic Framework with Open Mg2+ Adsorption Sites. *Chem. Commun.* **2011**, *47* (4), 1157–1159. https://doi.org/10.1039/c0cc03453c.
(27) Hadjiivanov, K. I.; Vayssilov, G. N. Characterization of Oxide Surfaces and Zeolites by Carbon Monoxide as an IR Probe Molecule. *Adv. Catal.* **2002**, *47*, 307–511. https://doi.org/10.1016/S0360-0564(02)47008-3.
(28) Jaegers, N. R.; Khivantsev, K.; Kovarik, L.; Klas, D. W.; Hu, J. Z.; Wang, Y.; Szanyi, J. Catalytic Activation of Ethylene C-H Bonds on Uniform D8 Ir(i) and Ni(Ii) Cations in

Zeolites: Toward Molecular Level Understanding of Ethylene Polymerization on Heterogeneous Catalysts. *Catal. Sci. Technol.* **2019**, *9* (23), 6570–6576. https://doi.org/10.1039/c9cy01442j.

(29) Ivanova, E.; Mihaylov, M.; Thibault-Starzyk, F.; Daturi, M.; Hadjiivanov, K. FTIR Spectroscopy Study of CO and NO Adsorption and Co-Adsorption on Pt/TiO2. *J. Mol. Catal. A Chem.* **2007**, *274* (1–2), 179–184. https://doi.org/10.1016/j.molcata.2007.05.006.

(30) Kapelewski, M. T.; Runčevski, T.; Tarver, J. D.; Jiang, H. Z. H.; Hurst, K. E.; Parilla, P. A.; Ayala, A.; Gennett, T.; Fitzgerald, S. A.; Brown, C. M.; Long, J. R. Record High Hydrogen Storage Capacity in the Metal-Organic Framework Ni2(m-Dobdc) at Near-Ambient Temperatures. *Chem. Mater.* **2018**, *30* (22), 8179–8189. https://doi.org/10.1021/acs.chemmater.8b03276.

(31) Lyu, H.; Zhang, Q.; Wang, Y.; Duan, J. Unified Meso-Pores and Dense Cu2+ Sites in Porous Coordination Polymers for Highly Efficient Gas Storage and Separation. *Dalt. Trans.* **2018**, *47* (13), 4424–4427. https://doi.org/10.1039/c8dt00512e.

(32) Yang, H.; Whitten, J. L. Dissociative Adsorption of H2 on Ni(111). *J. Chem. Phys.* **1993**, *98* (6), 5039–5049. https://doi.org/10.1063/1.464958.

(33) Behm, R. J.; Christmann, K.; Ertl, G.; Van Hove, M. A.; Weinberg, W. H. Chemisorption Geometry of Hydrogen on a Ni(111) Surface. *Surf. Sci.* **1979**, *89* (1–3), 403. https://doi.org/10.1016/0039-6028(79)90626-5.

4 Calculation of Geometric Descriptors for a MOF Database

4.1 Abstract

The Woo lab was involved in a project to create a large computationally ready database composed of hypothetical and experimental MOFs (~280,000 structures) with geometric properties, descriptors and DFT derived electrostatic fitted partial atomic charges pre-calculated. The charges are needed to perform GCMC simulations to simulate gas interaction with the MOF framework. The descriptors chosen can be used to build machine learning models to predict relevant properties for MOF such as gas uptake or selectivity. The focus of this author's work on this project is the collection of relevant geometric properties for the database. The distribution of these descriptors are then compared to the existing design space to show the improved diversity of this database which is an important metric to consider when using the database to build machine learning models. This work has been published in ACS, Chemistry of Materials in the paper "ARC−MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning" where this author performed all computational calculations presented.[1] As of Nov. 2025, this database has been cited in >100 publications since its publication in January 2023.

4.2 Introduction

Metal Organic Frameworks (MOFs) are a class of porous material that has gained interest over recent years for their interaction with gases. MOFs are made up of organic and inorganic structural building units (SBUs) that come together to form a periodic structure. This makes MOFs potentially very tunable as different SBUs can be used in combination to create MOFs with specific properties. These properties have led to MOF research for usages for many applications including batteries,[2,3] water harvesting[4,5] and drug delivery.[6,7]

MOFs unique properties led to research into applications for gas separation and storage. The effects of climate change from the increased release of greenhouse gases into the atmosphere have become noticeable over the last decades[8,9] leading to more research into the application of MOFs for greenhouse gas collection. MOFs are ideal candidates for gas interactions as they have small pore sizes and large pore volumes. Record breaking surface areas are ~ 7000 m^2/g for experimental MOFs and ~14600 m^2/g for theoretical MOFs.[10]

The staggering number of MOFs that can be made make it impractical for experimental chemists to search for top performing materials in a reasonable amount of time. It is far more time efficient to use computational techniques to create databases of hypothetical and experimentally discovered MOFs and perform simulations on the materials to identify ideal materials for a specific goal. The existing MOF databases of note are the CoRE[11], CSD MOF subset[12], QMOF[13] and the Mujumdar[14] databases all with different sets of MOFs. The CoRE, CSD MOF subset and QMOF are databases of experimental materials whereas the Mujumdar database is composed of hypothetical MOFs with the latter containing MOFs.

Hypothetical MOFs are materials that are rationally designed by a chemist that have not been created experimentally. There are two ways that an hMOF can be made, either “top-down” or “bottom-up”. MOFs created from the “top-down” approach use existing topologies and generate the crystal structure from adding SBUs to the network as used in the Boyd-Woo database, whereas MOFs created from the ‘bottom-up” approach start with existing SBUs to create a connected crystal network as used in the Wilmer database. Hypothetical databases, such as the Wilmer database, have been shown to have poor diversity with respect to inorganic SBUs.[14,15]

To simulate adsorption properties to predict experimental performance, simulations need to be performed to create an isotherm, generally atomistic GCMC simulations. To perform a GCMC simulation, interatomic potentials and electrostatic potentials need to be generated. Guest host interactions in these simulations are typically computed with so-called van der Waals potentials which account for the steric and dispersion interactions and the electrostatic interactions. While the van der Waals potentials are believed to be generally transferable between MOFs, the partial atomic charges needed to model electrostatic interactions are not. This means that the atomic charges used in simulations need to be recomputed for each MOF that is simulated. DFT derived charges can be computed using the REPEAT method to generate electrostatic potential fitted charges.

The structure of a MOF has many different geometric features that can be calculated to characterize the material. Previous studies successfully used geometric properties to create machine learning models to predict methane storage[16], CO_2/CH_4 selectivity[17] and CO_2/H_2 selectivity[18]. Having a diverse set of MOFs with a range of geometric properties available in one database will be useful for anyone looking to perform machine learning. Moosavi et al.[19] trained an ML model on three sets of data with training sets from existing databases were compared to a set of data considered diverse of the two examined databases. The diverse set of data was able to provide an improved model to the other training sets.

The Woo lab developed the ARC-MOF (*ab initio* REPEAT charge) database,[1] containing ~280,000 experimental and hypothetical MOF structures with precomputed DFT derived partial atomic charges and descriptors to use for building machine learning models. Three different descriptor classes are included for a user to perform machine learning on the database. These classes are geometric, atomic-property-weighted radial distribution functions (AP-RDFs), and revised autocorrelation (RAC). The focus on this chapter is on the geometric properties as that was this author's contribution to the project.

4.3 Methods

If the ARC-MOF DB is to be used for high-throughput screening, its contents should only contain well made structures that are considered chemically feasible. A series of checks were performed to ensure the quality of the accepted MOFs. One of the filtering methods determined whether the oxidation state of the metal atoms is reasonable. A code developed by another member of the Woo Lab group was used to calculate the oxidation state on each metal centre of a MOF and a material was discarded if the oxidation state was impossible, non-integer or extremely rare. Unit cells deemed unrealistically small were also excluded which meant any unit cell vector smaller than 3.22 Å, the smallest unit cell vector from an experimentally synthesized MOF, were excluded. The final check was to look for hypercoordinated main-group elements not bound to the metal centre. Structure graphs were generated without the presence of metal atoms and any MOF with hypervalent carbon atoms, oxygen atoms, halogen atoms, and hydrogen atoms were excluded from the ARC-MOF DB.

Zeo ++[20] is a program that can rapidly compute geometric properties of porous structures. In this study version 0.3.0 of Zeo ++ was used and all calculations used the high accuracy flag and a probe radius of 1.86 Å to resemble a nitrogen molecule. To ensure the calculations were done efficiently, each value was calculated with a different number of Monte Carlo steps; accessible surface area and probe-occupiable volume were calculated with 2000 MC steps and accessible volume was calculated with 50000 MC steps. Zeo++ was used to calculate the geometric descriptors in this chapter.

The descriptors used for the diversity analysis are six geometric properties calculated from Zeo ++as well as the RAC descriptors for metal centre, ligand chemistry and functional group chemistry. The details of the RAC calculations will not be included as they were not this student's work. The large amount of data requires a dimensional reduction to aid in data visualization which in this work is done by the uniform manifold approximation and projection (UMAP) technique[21] with the use of the RAPIDS cuML Python library.[22] An explanation of the how UMAP reduces the dimensionality was not this authors work, therefore only a description of how to interpret the data will be included. UMAP was used to reduce the six-dimension geometric descriptors provided in Table 2, to two-dimensions to create a plot. Two sets of data were compared, the "entire" design space and the ARC-MOF subset.

UMAP plots only offer a qualitative analysis of the data, therefore three quantitative metrics also used by Moosavi et al.[15] were used to analyse the data. Disparity (D) measures the spread of two datasets compared to each other. In this case it is the ratio between the area occupied by ARC-MOF and the area occupied by the entire design space. Variety (V) and balance (B) require the formation of clusters based on descriptors and are not dependent on the results of the UMAP plots. The hbdscan Python library[23] used HBDSCAN[24] to cluster the MOFs based on descriptor. Variety is a comparison of how many subset clusters occupy the total number of clusters and will equal one if there is one subset structure in each cluster. Balance is an absolute metric computed with Pielou's evenness which equals unity for a balanced data set.

4.4 Database Composition

The ARC-MOF database is composed of structures from 15 different sources all filtered by the criteria mentioned in the methods section. This section will briefly go over the databases labeled DB*x* where x is a number between zero and fifteen, and refers to a different source database. Table 4.1 features the information collected from each database but was not created by this author.

Table 4.1: Sources for the construction of the ARC-MOF database with accompanying labels and number of MOFs provided by each database.[1]

Label	References	Total no. of MOFs	no. of MOFs after Structure Check	no. of MOFs in ARC-MOF (%)
DB0	Boyd et al.	358,398	263,218	203,025 (72.2)
DB1	Lan et al.	303,992	181,885	23,267 (8.3)
DB2	Colón et al.	13,514	3920	199 (0.1)
DB3	Anderson et al.	426	358	123 (0.0)
DB4	Gómez-Gauldrón et al.	204	48	25 (0.0)
DB5	Chung et al.	51,163	27,022	22,366 (8.0)
DB6	Li et al.	11,555	10,944	9092 (3.2)
DB7	Majumdar et al.	23,891	12,316	6955 (2.5)
DB8	Anderson and Gómez-Gauldrón	126	122	8 (0.0)
DB10	Anderson et al.	105	78	22 (0.0)
DB12	CoRE 2019			
DB13	Bao et al.	8629	6180	5165 (1.8)
DB14	CSD MOF			
DB15	this work	7708	2841	2146 (0.8)
Total		806,520	521,381	279,610

Though there are 15 databases sourced for ARC-MOF, only the relevant databases will be discussed to explain the distribution plots presented. DB0 was the previous database made by the Woo lab named the Boyd-Woo database[25] and is a collection of hypothetically generated MOFs made by from the ToBasCCo algorithm.[26] Table 1 shows that DB0 makes up the majority, 72.2% of the database, therefore it is unnecessary to compare the distribution of geometric parameters between DB0 and ARC-MOF as the normalization of the data will cause their plots to look nearly identical. DB7, created by Majumdar et al.[14], is a hypothetical database made with inorganic linkers not commonly seen in other hypothetical databases, which makes it important when performing a diversity analysis. DB12 is the CoRE 2019 database[27] is a database of experimentally characterized MOFs that are ready for computational analysis originally from the CSD. DB14 is the CSD-MOF collection[12], a subset of materials from the CSD that are considered MOFs. This was further reduced to only non-disordered MOFs, MOFs with hydrogen atoms, and only MOFs that did not have codes matching the CoRE database to reduce duplicates. DB15 uses structures created by Pormake, a top-down MOF construction method. Of these selected MOFs, each database had its structures sorted by number of atoms in a unit cell to perform DFT calculations on. MOFs that were too large to complete a DFT calculation were not included in ARC-MOF. Additionally, if a MOF did not have a valid metal oxidation state it was also removed.

4.5 Results and Discussion

The distribution of geometric properties of MOFs contained in the ARC-MOF DB was compared to four other existing databases: CoRE 2019,[27] the CSD mof subset,[12] QMOF,[13] and the Majumdar[14] databases. The plots shown in Figure 4.1 are normalized histograms of the databases to compare the diversity. Since the minima and maxima values of geometric properties are hard to tell from Figure 4.1, Table 4.2 lists them. Since ARC-MOF is much larger than the other databases compared, a normalized distribution may suggest that ARC-MOF does not contain any MOFs with certain properties near extremes because the percentage is small. Figure 4.2 plots the absolute number of MOFs on each database as a function of different property values.

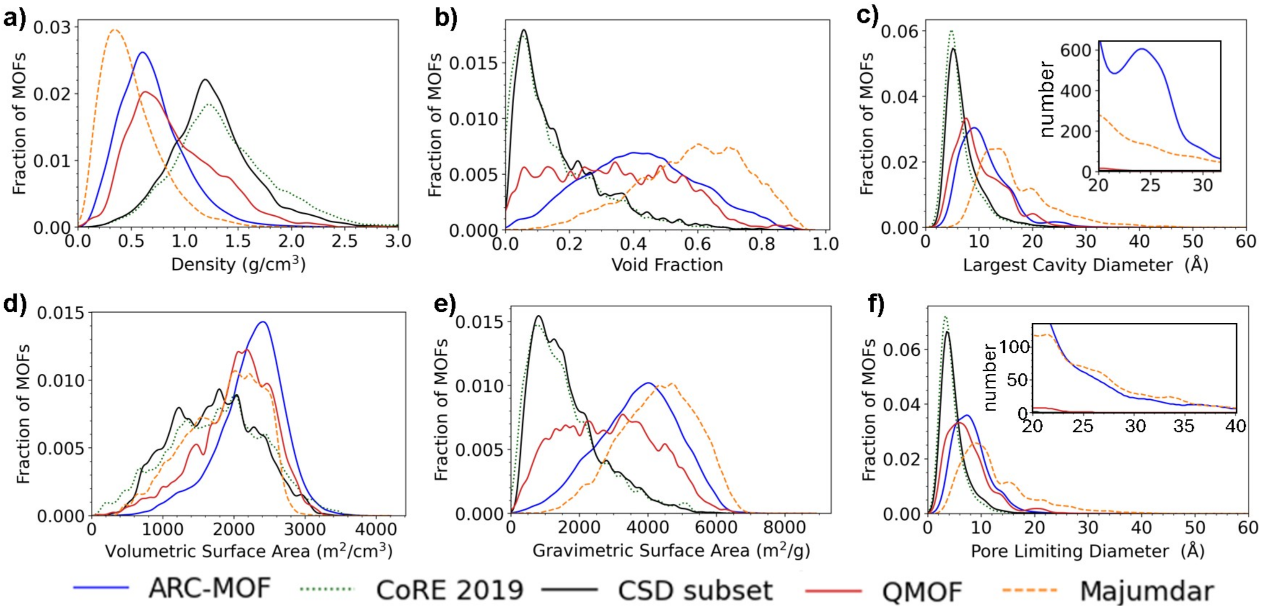

Figure 4.1: Comparison of geometric property normalized distribution between existing databases and ARC-MOF. Structures with zero surface area, void fractions, and pore diameters are not plotted, but are included in the distribution calculation

Table 4.2: Max and Min of Relevant Geometric Properties in ARC-MOF Compared to Existing Databases[1]

Database		Density (g/cm^3)	Surface Area (m^2/cm^3)	Surface Area (m^2/cm^3)	Volume Fraction	Largest Cavity Diameter (Å)	Pore Limiting Diameter (Å)
ARC-MOF	Max	6.20	3475	10,219	0.95	83.1	81.1
	Min	0.02	0	0	0	1.6	0.1
CoRE	Max	4.16	3150	8309	0.89	71.6	71.5
	Min	0.06	0	0	0	2.7	1.1
CSD	Max	4.06	3153	6621	0.80	71.6	71.5
	Min	0.13	0	0	0	2.7	0.5
QMOF	Max	2.88	2878	7438	0.88	33.7	30.1
	Min	0.08	0	0	0	1.9	0.9
Majumdar	Max	2.09	2724	8013	0.92	56.6	54.2
	Min	0.05	0	0	0	4.1	3.0

Figure 4.1 shows the databases containing hypothetical structures (ARC, Majumdar) have similar distributions of geometric parameters to each other, as do the experimental databases (CSD and CoRE). The hypothetical MOF databases feature more porous MOFs with larger surface areas, pore diameters, and void fractions which is consistent with the lower densities seen in the hypothetical databases. Since hMOFs are constructed *in silico* constructed MOFs may be unstable and have unphysically large pore sizes. This is the reason for the large skew in distribution between the experimental and hypothetical databases. It is important to note as Figure 4.1 shows the normal distribution of each database, where ARC-MOF appears to lack MOFs of large pores, the absolute distribution inset plots show ARC containing more MOFs of all pore sizes than QMOF. Table 4.2

demonstrates that ARC-MOF contains the largest maximum and smallest minimum of each geometric property.

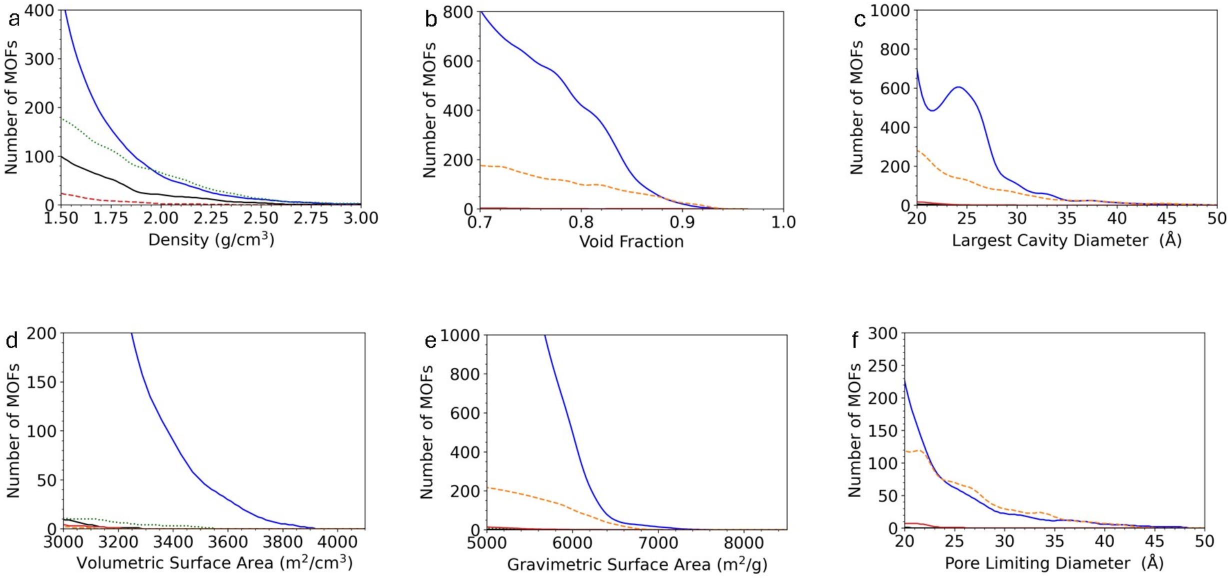

Figure 4.2: Distribution of geometric parameters from the databases in the legend. This is in comparison to the normal distribution to demonstrate the large number of MOFs in ARC-MOF compared to other databases.

4.5.1 Diversity Analysis

The following work on the diversity analysis was done by Jake Burner, a PhD. student in the Woo lab group using the data that I computed.[1]

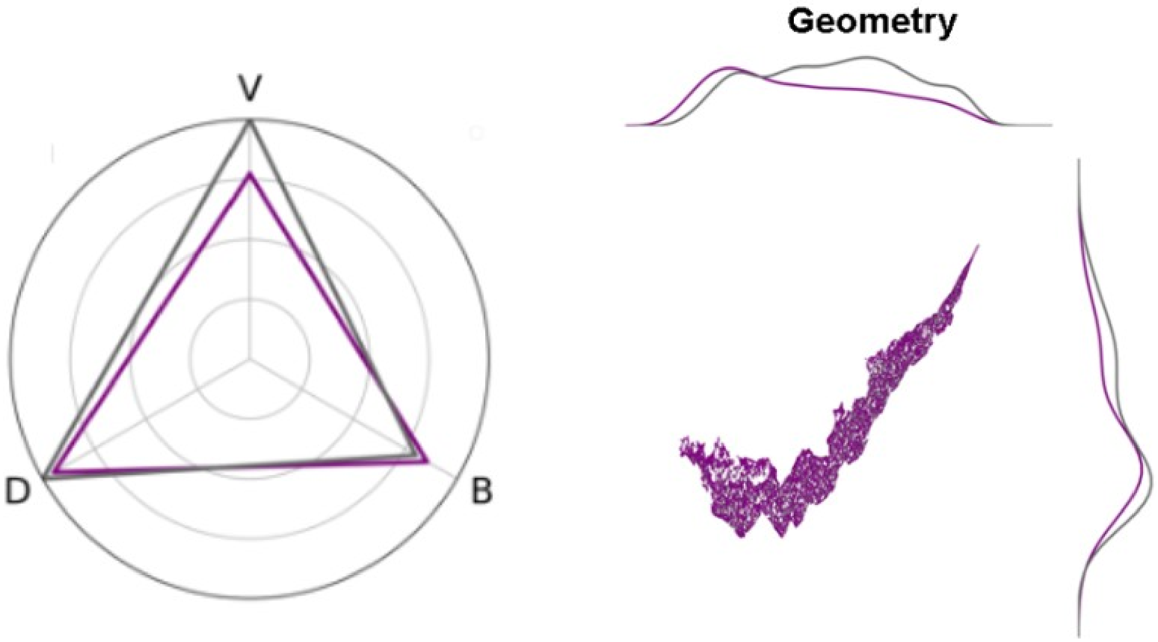

Figure 4.3: “Two-dimensional UMAP projection of descriptors of 50,000 random MOFs and corresponding radar plots showing diversity metrics for the geometry descriptors. Structures present in ARC−MOF (∼280 K MOFs) are represented by purple points, which is overlaid on the entire design space (∼480 K MOFs), represented by gray points. The diversity metrics shown on the radar plots are variety (V), disparity (D), and balance (B), where gray again represents the entire set of MOFs and color represents the ARC−MOF subset. Only MOFs with non-zero accessible surface area are shown.”[1]

The goal of the UMAP plots is to show the coverage of ARC-MOF over the entire design space. As mentioned in the methods section, the entire design space is represented by a random 50K of the 480K MOFs that make up the entire design space. The UMAP plot in Figure 4.3 feature a scatterplot representing the reduced dimensionality of the geometric descriptors with the purple points and lines representing the ARC-MOF data and the grey points and lines representing the entire design space. There are kernel density estimate plots along the axes of each plot to quantitatively measure the balance of the dimension-reduction by showing the density of points along an axis. The radar plots show the diversity metric results (disparity, variety, and balance).

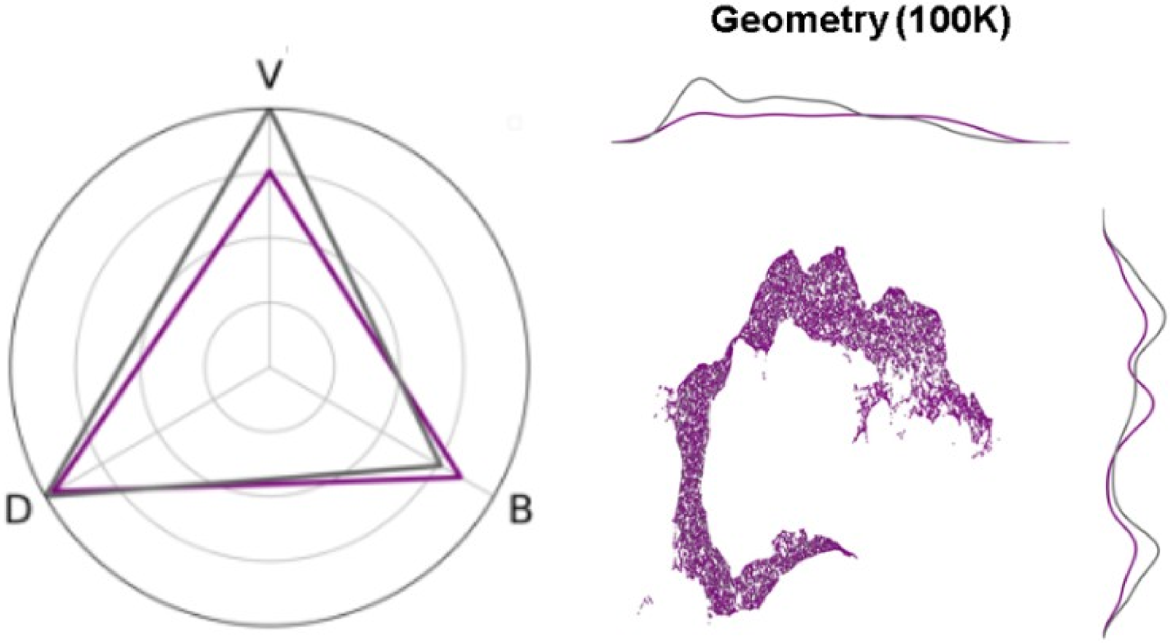

Figure 4.4: “Two-dimensional UMAP projection of descriptors of a varying number of MOFs sampled using farthest point sampling and corresponding radar plots showing diversity metrics for geometry based on geometric descriptors. Structures present in ARC−MOF (∼280 K MOFs) are represented by purple points, which is overlaid on the entire design space (∼480 K MOFs), represented by gray points. The diversity metrics shown on the radar plots are variety (*V*), disparity (*D*), and balance (*B*), where gray again represents the entire set of MOFs and color represents the ARC−MOF subset. Only MOFs with non-zero accessible surface area are shown.”[1]

These plots demonstrate that ARC-MOF is diverse with respect to the entire design space. This conclusion comes form the lack of large grey regions in the scatter plots meaning ARC-MOF covers the same space that the entire design space covers. The radar plots also show good variety and balance for the geometry descriptors.

The other descriptors didn’t have as good balance as the geometric descriptors and needed correction. Farthest point sampling[28] was used to correct the data imbalance by iteratively selecting points that lead to optimal data set coverage. Figure 4.4 shows the UMAP plot for the balanced subset of 100K MOFs which did not change the diversity of the data set. Geometry was farthest-point sampled for 100K MOFs demonstrating the well-balanced nature of these descriptors. The goal of a balanced subset is to be used for machine learning to prevent the introduction of biases from repeating data.

4.6 Conclusions

Geometric properties were calculated and collected for MOFs in the ARC database. The distribution of these properties was then compared to existing databases to demonstrate how large and diverse ARC is compared to existing databases. ARC MOF had distributions of geometric properties covering the ranges seen in both hypothetical and experimental databases and a greater number of MOFs in each section. A brief analysis of the diversity of the geometric features of the MOFs was performed using the UMAP dimensionality reduction tool. ARC-MOF is already being used in high-throughput screening and machine learning studies.

4.7 References

(1) Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P. G.; Maley, S.; Gibaldi, M.; Simrod, S.; Ogden, V.; Woo, T. K. ARC–MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning. *Chem. Mater.* **2023**, *35* (3), 900–916. https://doi.org/10.1021/acs.chemmater.2c02485.

(2) Zhang, X.; Chen, A.; Zhong, M.; Zhang, Z.; Zhang, X.; Zhou, Z.; Bu, X. H. *Metal–Organic Frameworks (MOFs) and MOF-Derived Materials for Energy Storage and Conversion*; Springer Singapore, 2019; Vol. 2. https://doi.org/10.1007/s41918-018-0024-x.

(3) Ye, Z.; Jiang, Y.; Li, L.; Wu, F.; Chen, R. *Rational Design of MOF-Based Materials for Next-Generation Rechargeable Batteries*; 2021; Vol. 13. https://doi.org/10.1007/s40820-021-00726-z.

(4) Kalmutzki, M. J.; Diercks, C. S.; Yaghi, O. M. Metal–Organic Frameworks for Water Harvesting from Air. *Adv. Mater.* **2018**, *30* (37), 1–26. https://doi.org/10.1002/adma.201704304.

(5) Kim, H.; Rao, S. R.; Kapustin, E. A.; Zhao, L.; Yang, S.; Yaghi, O. M.; Wang, E. N. Adsorption-Based Atmospheric Water Harvesting Device for Arid Climates. *Nat. Commun.* **2018**, *9* (1), 1191. https://doi.org/10.1038/s41467-018-03162-7.

(6) Hasan, M. N.; Bera, A.; Maji, T. K.; Pal, S. K. Sensitization of Nontoxic MOF for Their Potential Drug Delivery Application against Microbial Infection. *Inorganica Chim. Acta* **2021**, *523* (March), 120381. https://doi.org/10.1016/j.ica.2021.120381.

(7) Mallakpour, S.; Nikkhoo, E.; Hussain, C. M. Application of MOF Materials as Drug Delivery Systems for Cancer Therapy and Dermal Treatment. *Coord. Chem. Rev.* **2022**, *451*, 214262. https://doi.org/10.1016/j.ccr.2021.214262.

(8) Washington, D. Climate Watch Historical GHG Emissions.

(9) Mitchell, J. F. B. The “Greenhouse” Effect and Climate Change. **1989**, No. 89, 115–139.

(10) Farha, O. K.; Eryazici, I.; Jeong, N. C.; Hauser, B. G.; Wilmer, C. E.; Sarjeant, A. a; Snurr, R. Q.; Nguyen, S. T.; Yazaydın, a Ö.; Hupp, J. T. Metal-Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit? *J. Am. Chem. Soc.* **2012**, *134* (36), 15016–15021. https://doi.org/10.1021/ja3055639.

(11) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal-Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985–5998. https://doi.org/10.1021/acs.jced.9b00835.

(12) Li, A.; Perez, R. B.; Wiggin, S.; Ward, S. C.; Wood, P. A.; Fairen-Jimenez, D. The Launch of a Freely Accessible MOF CIF Collection from the CSD. *Matter* **2021**, *4* (4), 1105–1106. https://doi.org/10.1016/j.matt.2021.03.006.

(13) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine Learning the Quantum-Chemical Properties of Metal–Organic Frameworks for Accelerated Materials Discovery. *Matter* **2021**, *4* (5), 1578–1597. https://doi.org/10.1016/j.matt.2021.02.015.

(14) Majumdar, S.; Moosavi, S. M.; Jablonka, K. M.; Ongari, D.; Smit, B. Diversifying Databases of Metal Organic Frameworks for High-Throughput Computational Screening.

ACS Appl. Mater. Interfaces **2021**, *13* (51), 61004–61014. https://doi.org/10.1021/acsami.1c16220.

(15) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal-Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11* (1), 4068. https://doi.org/10.1038/s41467-020-17755-8.

(16) Fernandez, M.; Woo, T. K. T. K.; Wilmer, C. E. C. E.; Snurr, R. Q. R. Q. Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks. *J. Phys. Chem. C* **2013**, *117* (15), 7681–7689. https://doi.org/10.1021/jp4006422.

(17) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Quantitative Structure–Property Relationship Models for Recognizing Metal Organic Frameworks (MOFs) with High CO2Working Capacity and CO2/CH4Selectivity for Methane Purification. *Eur. J. Inorg. Chem.* **2016**, *2016* (27), 4505–4511. https://doi.org/10.1002/ejic.201600365.

(18) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High Co 2 Working Capacity and Co 2 /h 2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123* (7). https://doi.org/10.1021/acs.jpcc.8b10644.

(19) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the Diversity of the Metal-Organic Framework Ecosystem. *Nat. Commun.* **2020**, *11* (1), 1–10. https://doi.org/10.1038/s41467-020-17755-8.

(20) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149* (1), 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.

(21) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv e-prints* **2018**. https://doi.org/10.48550/arxiv.1802.03426.

(22) Nolet, C. J.; Lafargue, V.; Raff, E.; Nanditale, T.; Oates, T.; Zedlewski, J.; Patterson, J. Bringing UMAP Closer to the Speed of Light with GPU Acceleration. *ArXiv e-prints* **2020**. https://doi.org/10.48550/arxiv.2008.00325.

(23) Wang, Y.; Yu, S.; Gu, Y.; Shun, J. Fast Parallel Algorithms for Euclidean Minimum Spanning Tree and Hierarchical Spatial Clustering. In *Proceedings of the 2021 International Conference on Management of Data*; ACM: New York, NY, USA, 2021; pp 1982–1995. https://doi.org/10.1145/3448016.

(24) Campello, R. J. G. B.; Moulavi, D.; Zimek, A.; Sander, J. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Trans. Knowl. Discov. from Data* **2015**, *10* (1). https://doi.org/10.1145/2733381.

(25) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; Reimer, J. A.; Navarro, J. A. R.; Woo, T. K.; Garcia, S.; Stylianou, K. C.; Smit, B. Data-Driven Design of Metal–Organic Frameworks for Wet Flue Gas CO2 Capture. *Nature* **2019**, *576* (7786), 253–256. https://doi.org/10.1038/s41586-019-1798-7.

(26) Boyd, P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical

Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18* (21), 3777–3792. https://doi.org/10.1039/C6CE00407E.

(27) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; Slater, B.; Siepmann, J. I.; Sholl, D. S.; Snurr, R. Q. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *J. Chem. Eng. Data* **2019**, *64* (12), 5985–5998. https://doi.org/10.1021/acs.jced.9b00835.

(28) Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137–148. https://doi.org/10.1080/00401706.1969.10490666.

5 Conclusions

5.1 Summary

This thesis is composed of two separate projects involving computational studies of MOFs. Chapter 3 was a computational study of the H_2 binding in the MOF, BFF, synthesized by Chiu et al.[1] The Stylianous lab at Oregon State University synthesized a MOF, $[Ni_3(pzdc)_2(ade)_2(H_2O)_4]$ with high hydrogen uptake at low pressures. Given its structure, we performed atomistic simulations to understand its high H_2 adsorption properties. BFF's non-activated crystal structure was determined however, only the PXRD pattern for the activated structure was obtained. As a result, DFT level geometry optimizations were performed on the activated structure. TGA analysis suggested that two water molecules were removed upon activation. The activated structure was determined by optimizing the MOF structure after removing every possible combination of water sites to determine the lowest energy unit cell and the most similar PXRD pattern to experimental. We then performed GCMC simulations on our candidate activated structure to generate simulated adsorption isotherms. These were found to be similar to the experimental isotherms. Hydrogen molecules were then added to the unit cell and the structure was optimized to provide insights into the mode of H_2. Finally, a metal centre swap was performed to study the effects the metal centre had on hydrogen uptake.

The second project involved the generation of geometric properties for the ARC-MOF database of computation ready experimental and hypothetical MOF structures.[2] One goal of the ARC-MOF DB was to create a database with precomputed descriptors for machine learning applications. For this work Zeo++ was used to determine the relevant geometric properties for the database. The calculated properties were as follows: pore diameters, accessible surface area, accessible volume, probe-accessible volume, and density. The geometric properties were then compared to previous databases to determine the distribution and diversity of the database in terms of geometric parameters.

5.2 Ongoing Work

BFF's large uptake was unexpected due to the low gravimetric surface area which was counter to Chahine's rule. Knowledge of this allows for the rational design of materials for hydrogen adsorption to meet the Department of Energy's on-board hydrogen storage for vehicles at ambient temperatures. One review suggests increasing the binding affinity of the MOFs will enhance effectiveness at ambient temperatures.[3] For example, a vanadium based MOF featured strong binding, 21 kJ/mol, resulting in a gravimetric uptake of 1.64 wt % and a volumetric uptake of 10.7 g/L. Combining this knowledge with optimal pore structures may lead to successful MOF based hydrogen storage solutions.

The ARC-MOF database needs to be updated with new structures to ensure it represents the MOF design space. To improve the diversity of the database, a wider variety of metal SBUs need to be included to improve training of machine learning models. There have been some studies since the publication of the paper. Arjmandi et al.[4] published a paper using the ARC-MOF database to predict water uptake. A diverse set of 2600 MOFs were used to create a structure function relationships were connected from GCMC simulations with interpretable machine learning. The GCMC simulation was from water uptake at 100 % and 30 % relative humidity revealing the importance of evaluating materials at real world conditions. The machine learning study reveals the water uptake capacity is controlled by factors such as adsorption energetics, local electrostatics and framework density. Mohamed et al.[5] performed a study include stability metrics in a high-throughput screening study to determine top performing MOFs for CO_2 capture. The adsorption data of 15,219 hMOFs from ARC-MOF were used to initially filter the list then MOFs then stability metrics were determined to select for the stable hMOFs for CO_2 capture. Hopefully there will be many more studies done with the ARC-MOF database.

5.3 References

(1) Chiu, N. C.; Compton, D.; Gładysiak, A.; Simrod, S.; Khivantsev, K.; Woo, T. K.; Stadie, N. P.; Stylianou, K. C. Hydrogen Adsorption in Ultramicroporous Metal-Organic Frameworks Featuring Silent Open Metal Sites. *ACS Appl. Mater. Interfaces* **2023**. https://doi.org/10.1021/acsami.3c12139.

(2) Burner, J.; Luo, J.; White, A.; Mirmiran, A.; Kwon, O.; Boyd, P. G.; Maley, S.; Gibaldi, M.; Simrod, S.; Ogden, V.; Woo, T. K. ARC–MOF: A Diverse Database of Metal-Organic Frameworks with DFT-Derived Partial Atomic Charges and Descriptors for Machine Learning. *Chem. Mater.* **2023**, *35* (3), 900–916. https://doi.org/10.1021/acs.chemmater.2c02485.

(3) Chen, Z., Kirlikovali, K. O., Idrees, K. B., Wasson, M. C., & Farha, O. K. (2022). Porous materials for hydrogen storage. In *Chem* (Vol. 8, Issue 3, pp. 693–716). Elsevier Inc. https://doi.org/10.1016/j.chempr.2022.01.012

(4) Arjmandi, M., Aytaç, E., Khayet, M., & Hilal, N. (2026). Next-generation MOFs for atmospheric water harvesting: The role of machine learning techniques. In *Coordination Chemistry Reviews* (Vol. 548). Elsevier B.V. https://doi.org/10.1016/j.ccr.2025.217211

(5) Mohamed, S. A., Zhao, D., & Jiang, J. (2023). Integrating stability metrics with high-throughput computational screening of metal–organic frameworks for CO2 capture. *Communications Materials*, *4*(1). https://doi.org/10.1038/s43246-023-00409-9