

Covalent Organic Frameworks for Gas Storage: A Neural Network-based Molecular Dynamics Study

Author:

Laura Eugenia Elisabeth Kronenberg, B.Sc.

Master Thesis

Submitted for the Degree

Master of Science in Materials- and Nanosciences

July 2023

Supervisor: assoc. Prof. Dr. Thomas S. Hofer

Institute of General, Inorganic and Theoretical Chemistry

Faculty of Chemistry and Pharmacy

University of Innsbruck

Acknowledgements

First and foremost, I would like to thank my advisor Thomas Hofer for his excellent guidance. Throughout this whole endeavor, he showed me patience, provided knowledge and gave me room to explore.

I would also like to thank Bernhard Kriesche for his expertise and his programming skills, which were essential for the success of this project. Him, Manuel Schuler and Felix Purtscher as well as the Team Hofer as a whole made my time in-office not only a great learning experience but also very enjoyable.

I am very grateful for the ULB Schreibzentrum under leadership of Daniela Rothe - the opportunity to participate in the writing weeks and the master writing group was more than welcome. Thank you to all the wonderful people from my writing group for the shared motivation.

Last but not least, my sincere gratitude goes out to my family and friends for supporting me during every stage of this process.

"For small creatures such as we, the vastness [of the universe] is bearable only through love."

- Carl Sagan

Kurzfassung

Da die Klimakrise, in der wir uns befinden, zu einem großen Teil von klimaschädlichen Gasen in unserer Atmosphäre vorangetrieben wird, sind Methoden, diese Gase aus der Atmosphäre zu entfernen, ein Ansatz, die Klimaerhitzung einzubremsen. Zum Beispiel wird derzeit intensiv an CO₂-Sequestrierung geforscht. Eine für diesen Zweck interessante Materialklasse sind kovalente organische Gerüstverbindungen, die eine hohe Porosität aufweisen. In der vorliegenden Arbeit wurde HEX-COF 1, ein 2D-COF untersucht. Dazu wurden molekulardynamische Simulationen mit dem neuronalen Netzwerkpotential ANI-2x kombiniert. Dieses ermöglicht molekulardynamische Simulationen mit akzeptabler Genauigkeit und deutlich höherer Geschwindigkeit im Vergleich zu quantenmechanischen Methoden. Dadurch werden auch längere Simulationen möglich, mit denen die Diffusion der CO₂-Moleküle in den Poren des COFs analysiert werden kann. Es konnte gezeigt werden, dass HEX-COF 1 in den Simulationen eine ähnliche CO₂-Kapazität hatte, wie in publizierten experimentellen Studien. Weiters konnte festgestellt werden, dass die CO₂-Moleküle wie erwartet ein Diffusionsverhalten von sinkender Diffusivität mit steigender Menge an CO₂-Molekülen pro Pore aufweisen. Außerdem weisen die Ergebnisse darauf hin, dass ANI-2x in der Lage ist, intermolekularen Ladungstransfer zu simulieren, auch wenn es nicht spezifisch auf diese Interaktion trainiert wurde, was daran liegen dürfte, dass es gegen Dichtefunktionaltheorie-Daten trainiert wurde.

Abstract

In the face of today's climate crisis, which shows change at a speed unprecedented in earth's history, ways to mitigate greenhouse gas emissions are highly sought after. One such way might be carbon capture and sequestration, in which CO₂ is extracted and stored. Covalent organic frameworks are a material class with promising properties for this purpose thanks to their high porosity. In this work, HEX-COF 1 as a 2D covalent organic framework was chosen as target system.

With the help of ANI-2x, a neural network potential, molecular dynamics simulations of a 2D covalent organic framework with different concentrations of CO₂ molecules in the pore channels were conducted. Due to the neural network potential enabling long time simulations with an acceptable loss of accuracy compared to quantum mechanical methods, diffusion analysis was made possible. It showed a similar uptake capacity to experimental data and a diffusion behavior following expectations of decreasing diffusivity with increased CO₂ load. Results also suggest that ANI-2x is able to simulate intermolecular charge transfer despite not explicitly being trained for this specific interaction, thanks to its training based on density functional theory data.

Contents

1 Climate Change and CO₂	3
1.1 CO ₂ as Part of System Earth	3
1.1.1 Greenhouse Gases and Radiative Forcing	3
1.1.2 Ocean Acidification	4
1.2 Paleoclimatology: PETM vs. Anthropogenic Climate Change	6
1.3 Carbon Capture and Storage	10
2 Covalent Organic Frameworks	11
2.1 Synthesis of COFs	11
2.2 Uses of COFs	11
2.3 HEX-COF 1 as host for CO ₂	12
3 Methods	14
3.1 Geometry Optimization	14
3.1.1 First order methods of Geometry Optimization	15
3.1.2 Second order methods of Geometry Optimization	16
3.2 Simulations	18
3.2.1 Statistical Mechanics	18
3.2.2 Molecular Dynamics	18
3.2.3 Monte Carlo Methods	21
3.2.4 Non-natural Ensembles: Temperature and Pressure Control	21
3.2.5 Constraints - SHAKE/RATTLE Algorithm	23
3.2.6 Periodic Boundary Conditions	23
3.2.7 Diffusion	24
3.3 Neural Networks	25
3.4 Simulation Protocol and Data Analysis	30
4 Space Groups and Symmetry	32
5 Results	35
5.1 Interaction Energies	35
5.2 Structure of HEX-COF 1	37
5.3 Diffusion Analysis	37
5.4 Consistency of Results	40
6 Conclusion	41

Introduction

Ever since covalent organic frameworks (COFs) were discovered in 2005 by Coté *et al.*,¹ their high customizability for various applications has made them an important field of study. COFs are porous, crystalline polymers, with both 2D and 3D variations existing; this makes them applicable for various purposes. Among the applications for COFs is gas storage,² which is becoming increasingly important in the face of climate change and rising atmospheric greenhouse gas concentrations.³ Carbon capture and storage/sequestration (CCS) could help fight climate change by capturing greenhouse gases such as carbon dioxide, either from exhaust gases from factories and production plants, or directly from the atmosphere. The CO₂ could then be extracted from the host molecules and liquified or brought into a supercritical state in order to be efficiently stored underground.³

One COF that is promising for CO₂ storage is HEX-COF 1, which was first described by Alahakoon *et al.* in 2016. It is "an azine-linked hexaphenylbenzene based covalent organic framework", a 2D COF and has columnar pore channels in which guest molecules such as CO₂ or CH₄ can be stored. For CO₂, it shows an uptake capacity of 20 wt% at 273 K and 12 wt% at 298 K.^{4,5}

In recent years, machine learning and artificial intelligence has made significant improvements and is now being used broadly across different disciplines.^{6–8} In the context of theoretical chemistry, neural network potentials (NNPs) have shown great promise for calculating the potential energy surface of a system.⁹ Supramolecular systems such as COFs are computationally intense to calculate with quantum chemical methods, which would be preferable to classical molecular mechanics methods due to their accuracy in terms of complex interactions. The high computational effort becomes especially critical when attempting to study diffusion of gas molecules within host materials , since this goal requires long-time simulations in order to accurately calculate diffusion coefficients. This is where NNPs are able to bridge the gap, allowing long-time simulations while keeping computational effort comparatively low. The ANI-2x NNP was chosen for this work, since it is trained on all elements necessary for the COF studied.¹⁰

Chapter 1 will first provide an overview of the global climate change/crisis we are currently facing. It will examine the motivation behind capturing and storing CO₂ by providing an outline of some important issues it is connected to, notably its role as a greenhouse gas and in ocean acidification. Following this, the current climate crisis will be put into the context of paleoclimate to make clear that the quickness of this of change is unprecedented in earth's history. Lastly, a short synopsis of carbon capture and storage/sequestration (CCS) will be provided in order to assess its potential for climate change mitigation and better understand the purpose of this thesis.

Chapter 2 will introduce covalent organic frameworks (COFs), first generally, then in the context of gas storage. After a summary of the synthesis of COFs, their many uses will be discussed. Then, HEX-COF 1 specifically as the COF studied in this thesis will be described in more detail.

Chapter 3 will discuss the applied methodology, starting with geometry optimization before progressing to time-dependent simulations. In the context of simulations, statistical mechanics as their basis will be briefly summarized and then Monte Carlo methods as well as molecular dynamics will be discussed. Then, advanced ensembles and methods to obtain them, such as thermostats and manostats will be described. The next section will then focus on constraints of molecular geometries in order to enable more efficient simulations. Following this, periodic boundary conditions as another feature of many simulations will be introduced. To conclude the simulations section, diffusion and its analysis in computational chemistry using simulation trajectories is reviewed. Next, neural networks and neural network potentials with a special focus on the ANI framework of training neural network potentials and the resulting NNP ANI-2x are outlined. Lastly, the simulation protocol used for the investigations in this thesis is given and discussed.

Chapter 4 will give a short overview of space groups and symmetry to provide better context for the discussion of the results.

Results will be discussed in chapter 5, starting with the computed interaction energies of the layers. Then, HEX-COF 1's structure will be described and compared to that of the original publication. Next, the results of the diffusion analysis will be covered. Lastly, the consistency between sampling runs is put to the test and described.

Finally, the conclusion will attempt to summarize this thesis and provide an outlook.

1 Climate Change and CO₂

The earth and its atmosphere are a delicate system that is constantly in flux. Throughout our planet's history, climate has changed multiple times, influencing all organisms living in this system. Various proxies can be used to reconstruct past atmospheres and climates, such as ice cores and the air bubbles entrapped in it as well as the ice composition itself, chemical signatures in rocks and fossils, and sedimentary rocks and fossils themselves as a mirror of how living conditions were at the time of deposition. Through these different proxies, it is possible to piece together a climate record. It shows that throughout earth's history, factors such as atmospheric composition, earth position relative to the sun, plate tectonics, biosphere and hydrosphere have continually changed, always influencing each other in a never ending feedback loop.¹¹⁻¹⁴

1.1 CO₂ as Part of System Earth

In order to better understand the influence of CO₂ on earth's climate and organisms, the following section will provide a short introduction to two important mechanisms as examples, since it is beyond the scope of this thesis to discuss this complex topic in its entirety. First, greenhouse gases and radiative forcing will be described, then, ocean acidification will be focused on.

1.1.1 Greenhouse Gases and Radiative Forcing

Already in 1896, Svante Arrhenius postulated that the amount of CO₂ in the atmosphere has an influence on earth's climate. He wrote: "*if the quantity of carbonic acid increases in geometric progression, the augmentation of the temperature will increase nearly in arithmetic progression*".¹⁵ Carbonic acid at the time was the name for carbon dioxide. Over a century later, his observations still hold – while the exact numbers have changed slightly, the tenor of rising atmospheric CO₂ concentrations increasing radiative forcing and earth's temperature remain true. Equation 1 expresses this relation, with ΔF as radiative forcing, α as a constant, C_0 as the CO₂-concentration at the beginning of the observation period and C as the CO₂-concentration at the end of the observation period.¹⁶ According to the IPCC's 6th assessment report in 2021, CO₂ concentrations have risen by 47% between 1750 and 2019: from 278.3 ± 2.9 ppm in 1750 and 285.5 ± 2.1 ppm in 1850¹⁷ to 409.9 ± 0.9 in 2019 and about 418 ppm as of 2023 (2019 and 2023 values from NOAA, National Oceanic & Atmospheric Administration). The global mean temperature has risen by 1.15 (1.02 to 1.28) °C between 1850-1900 and 2022.¹⁸

$$\Delta F = \alpha \ln \left(\frac{C}{C_0} \right) \quad (1)$$

Without greenhouse gases like CO₂, CH₄ (methane) and N₂O (nitrous oxide), the average temperature at earth's surface would be -18°C (255 K). This can be calculated using equations 2 - 3. Equation 2 is the Stefan-Boltzmann-law, which describes the outgoing longwave radiation; πB stands for the emission of

radiation at temperature T by a blackbody and σ is the Stefan-Boltzmann constant.^{19,20}

$$\pi B = \sigma T^4 \quad (2)$$

From this, equation 3 can be derived, which allows calculation of the earth's equivalent blackbody temperature. With an albedo of $\mathcal{A} = 0.3$ and incoming shortwave radiation of $F_s = 1372 W m^{-2}$, one can calculate that T_e , earth's effective radiation temperature of earth would be $-18^\circ C$.^{19,20}

$$T_e = \left[\frac{(1 - \mathcal{A})F_s}{4\sigma} \right]^{\frac{1}{4}} \quad (3)$$

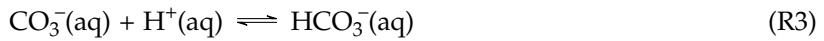
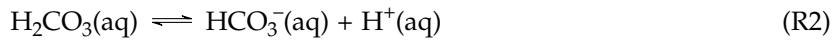
However, since a portion of the longwave radiation the earth emits is absorbed by water vapor, clouds and greenhouse gases, the radiative energy gets "trapped" within earth's atmosphere. The mechanism behind this is that the solar radiation consists mainly of shortwave radiation, while the radiation earth emits is mostly longwave radiation. Greenhouse gases in the atmosphere are almost transparent to the solar radiation, while they absorb the longwave radiation emitted by earth.¹⁹⁻²¹ Radiative forcing is thus defined as "*the reduction in net upward [terrestrial radiation] flux at the tropopause (i.e. above most of the atmosphere which interacts with [terrestrial radiation] produced by the components in question, expressed as a global average in $W m^{-2}$ "*", as McIlveen writes.²¹

The IPCC in their 2021 report on climate change's physical science basis report that the total anthropogenic effective radiative forcing (ERF) relative to 1750 is positive with a value of $2.72 W m^{-2}$. The ERF describes a climate driver's strength, with positive ERF resulting in warming. Greenhouse gases have a positive ERF and thus lead to warming, while aerosol cooling leads to cooling. Of the greenhouse gases, CO₂ is the largest contributor to the ERF with the increase of CO₂-concentrations compared to 1750 causing an ERF of $2.156 \pm 0.259 W m^{-2}$.²² While water vapor is one of the most common and strongest greenhouse gases in the atmosphere, modelling has shown that the anthropogenic water vapor emitted into the atmosphere causes an ERF of only -0.1 to $0.05 W m^{-2}$. Comparing this to the above mentioned ERF for CO₂, human-caused water vapor has a negligible effect. The reason behind this is that water vapor usually does not reach the upper troposphere and instead may even lead to a cooling effect due to increased cloud formation and evaporative cooling.²³

1.1.2 Ocean Acidification

Between 1751 and 2004, the average surface ocean pH sank from 8.25 to 8.14²⁴ - this process is called ocean acidification. Aqueous CO₂ that was taken up by the ocean will first react with the water to carbonic acid (H₂CO₃(aq), see reaction R1), which proceeds to dissociate into a hydrogen ion (H⁺) and bicarbonate (HCO₃⁻, reaction R2). The hydrogen ions do not exist as singular ions, they are actually transferred to water molecules which become oxonium ions (H₃O⁺). The H⁺ may also react with carbonates already dissolved in the seawater to form bicarbonate (reaction R3).²⁴⁻²⁶





The $\text{p}K$ values for these reactions are $\text{p}K_1 = 6.35$ (reaction R2) and $\text{p}K_2 = 10.33$ (reaction R3).^{27,28} Seawater is of course not pure H_2O but has different salts such as Na^+ or Cl^- dissolved in it, leading to an average salinity of 3.5% or 35 g/L. A decrease in salinity, temperature or pressure results in a decrease in the $\text{p}K$ values.²⁷

The different carbonate species together comprise the total dissolved inorganic carbon (DIC) in seawater, see equation 4. In order to make things easier, $[\text{CO}_2]$ and $[\text{HCO}_3^-]$ are sometimes expressed together as $[\text{CO}_2^*]$. HCO_3^- is the dominant carbonate form in seawater.^{24–26}

$$\text{DIC} = [\text{CO}_2] + [\text{H}_2\text{CO}_3] + [\text{HCO}_3^-] + [\text{CO}_3^{2-}] \quad (4)$$

Another important concept in seawater carbonate chemistry is Ω , the carbonate saturation state, with K_{sp} denoting the thermodynamic solubility product of CaCO_3 . The equation for Ω is given in equation 5. At $\Omega < 1$, CaCO_3 is preferentially dissolved, at $\Omega > 1$, it is (thermodynamically) stable.^{26,27}

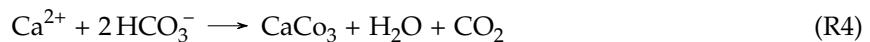
$$\Omega = \frac{[\text{Ca}^{2+}] \cdot [\text{CO}_3^{2-}]}{K_{sp}} \quad (5)$$

With the rising atmospheric CO_2 levels, more and more CO_2 is being absorbed by seawater. Generally speaking, the uptake of CO_2 mainly happens in the mid to high latitudes due to CO_2 's increased solubility in cooler seawater, while outgassing happens in the warmer tropical regions.^{24,25} To illustrate this, CO_2 solubility at 5°C is 0.2774 g CO_2 /100 mL H_2O while at 15°C it is 0.1970 g CO_2 /100 mL H_2O and at 25°C it is 0.1449 g CO_2 /100 mL H_2O .²⁹ Warm, carbon-poor seawater will stay on top while the cool, carbon-rich layers slide underneath - this is the solubility pump. Additionally, biological activity in the high latitudes takes up a lot of CO_2 through photosynthesis, subsequently dying and falling to the bottom, which is known as the biological pump bringing carbon down into the deep sea. On average, there is only a slight net surplus of CO_2 being taken up into the sea of about 2 Pg carbon/year, while about 90 Pg C/year go both directions. The 2 Pg C/year are much smaller than some regional variations. Unfortunately, the anthropogenic increase in CO_2 levels in the atmosphere may cause the balance to tip especially in some of the southern oceans to reverse the direction, so they will take up CO_2 in the future instead of outgassing.^{24,25}

The reason this is an issue is strongly tied to marine ecosystems. Since marine organisms have evolved in and adapted to specific conditions, the rapid changes brought about by nowadays' climate change pose challenges but also opportunities. In their 2019 paper, Hurd *et al.* described the different effects of changing seawater carbonate chemistry and ocean acidification on various organisms. In fish, a rising seawater $p\text{CO}_2$ will lead to an increase in the blood $p\text{CO}_2$, leading to the fish using more HCO_3^- from

the surrounding water and eliminating H⁺ in order to avoid the negative effects of a lower blood plasma pH (acidosis). This has several consequences: the otoliths (calcium carbonate fragments in the inner ear which are important for the vestibular system) will grow bigger in fish raised in a high pCO₂ environment. However, these fish may also begin to suffer ill effects to their cognitive and sensory capacities caused by the influence of higher blood HCO₃⁻-levels on the GABA_A receptor.²⁶

In shelled mollusks like bivalves and gastropods, the effects of ocean acidification are probably the most obvious: while calcification as described in reaction R4 may be heightened by the increased availability of HCO₃⁻, the outer shell may dissolve faster than it can be produced if Ω < 1 and net calcification becomes negative.²⁶



Corals and coralline algae are affected in two ways: photosynthetic rates are directly proportional to rising seawater DIC, however, only at values that are unrealistic for predicted ocean acidification. Secondly, and not surprisingly, the net calcification sinks with increasing ocean acidification. Some species of non-calcifying seaweeds may respond to increased CO₂-levels in seawater with higher growth rates, while others show no change or even a decrease. Phytoplankton are another important group of marine organisms, including diatoms (microalgae with a silica shell) and coccolithophores (single-celled organisms that build a shell out of carbonate crystals). In diatoms, some species may grow faster in CO₂-rich waters, since the energy cost of pH-regulation may decrease as more H⁺ are available to counter the OH⁻ produced during the diatom's metabolic activities. In coccolithophores, photosynthesis generates OH⁻ while using HCO₃⁻, causing a need for H⁺ uptake, while calcification requires release of H⁺, further driving ocean acidification.²⁶

1.2 Paleoclimatology: The Paleocene-Eocene Thermal Maximum vs. Anthropogenic Climate Change

One prominent example of climate change in earth's history and the complex interweaving of different factors is the Paleocene-Eocene Temperature Maximum (PETM), which will be discussed in greater detail due to the interesting comparison to current anthropogenic climate change. The PETM is known as a hyperthermal, a brief (on a geologic timescale, meaning no longer than a few 10,000 years) period in earth's history with markedly raised temperatures. It was discovered in 1991 by Kennett and Stott when they studied cores from a drill site off the coast of Antarctica. They noticed not only an extinction layer in the core without bioturbation of the sediments but upon isotope analyses of foraminifera, they were able to determine isotopic excursions of both δ¹³C and δ¹⁸O, showing warming of the ocean.³⁰ The PETM was determined to have started about 56 million years ago and had a duration of about 200,000 years (200 ka). Massive amounts of carbon were released into the atmosphere over a timespan of about 8–10 ka.^{31,32} The carbon source is still under debate (see below). The global average temperature increased by 5–8°C, with flora and fauna facing immense challenges concerning their geographical

distribution, forcing rapid evolution and relocation to higher latitudes. Concerning extinctions, only some benthic foraminifera went extinct, while most other organisms survived.³¹

Possible sources for the released carbon include destabilization and subsequent release of methane from deep-sea methane clathrates, thermogenic methane from heating up North Atlantic Cretaceous-Paleocene mudstones due to magmatic intrusions, thawing permafrost in Antarctica; less likely are huge wildfires and tectonic changes causing shallow seas to dry and release carbon. The amount of carbon released is estimated to be about what we humans are releasing currently. However, the duration of past and expected anthropogenic carbon release is about 15 times shorter than of the PETM (about 8 ka). The recovery from the PETM took about 83 ka after about 113 ka of disturbed worldwide climate. Recovery was likely driven by biological sequestration of carbon and subsequent covering by soil and peat, followed by higher carbonate preservation in the ocean.³¹

In contrast to this naturally occurring climate change on the paleocene-eocene border is today's climate change; see table 1 for a comparison. The Intergovernmental Panel on Climate Change (IPCC) released their Sixth Assessment Report: The Physical Science Basis in 2021, warning in chapter 5 that it "*is unequivocal that the increases in atmospheric carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) since the pre-industrial period are caused by human activities*".¹⁷ The amount of carbon released is currently at 670 ± 65 GtC (1850-2021) which is still lower than the estimated 2000-6000 GtC during the PETM. However, the current release of carbon has occurred over a timespan of only a little over 170 years, while the carbon release during the PETM is estimated to have taken about 20 000 years. Moreover, to really drive this point home, 70% (470 GtC) of the carbon released during the last 180 years was since 1960, and 33% (220 GtC) since the 2000. The yearly amount released between 2012 and 2021 was 10.8 ± 0.8 GtC, including the 0.2 GtC yr^{-1} cement carbon sink.³³ If carbon emissions are not reduced, and humankind continues at this rate, the 2000 GtC, the lower estimate of the carbon amount from the PETM, will be reached in about 123 years. This would bring the total timespan to 294 years for the same amount that took 20 000 years to be released in the past.

In 2008, Zachos *et al.* wrote that "*by the year 2400, it is predicted that humans will have released about 5,000 gigatonnes [sic] of carbon (Gt C) to the atmosphere since the start of the industrial revolution if fossil-fuel emissions continue unabated and carbon-sequestration efforts remain at current levels*";³² from today's perspective, taking the numbers from the 2022 global carbon budget by Friedlingstein *et al.* of 670 ± 65 GtC having been released since 1850,³³ that leaves 4330 ± 65 GtC from 2021-2400. Dividing this number by 379 (the number of years from 2021 to 2400), a value of $11.6 \pm 0.2 \text{ GtC yr}^{-1}$ (without the cement carbon sink) is obtained. This is only slightly higher than the current $10.8 \pm 0.8 \text{ GtC yr}^{-1}$ between 2012 and 2021. Since total anthropogenic emissions of carbon per year are projected to amount to 11.1 GtC yr^{-1} in 2022,³³ Zachos *et al.*'s projections seem realistic if emissions are not reduced.

The global temperature change during the PETM was probably between +5 to +8°C. As of 2022, global mean temperature has increased by 1.15°C compared to pre-industrial levels.¹⁸ Comparing the PETM's and anthropogenic climate change's biotic effects, the PETM caused many organisms to shift their geographic distribution towards cooler regions at the poles,³¹ much like we are seeing today as Pörtner *et al.* have pointed out in the Technical Summary of "Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change".³⁴ As already mentioned, extinction-wise, some benthic foraminifera went extinct during the PETM, but otherwise there were no mass extinctions.³¹ The IPCC so far recognizes three extinctions due to anthropogenic climate change: the Bramble Cay melomys (*Melomys rubicola*), the cloud-forest restricted golden toad (*Incilus periglenes*) and the white subspecies of the lemuroid ring-tail possum (*Hemibelideus lemuroides*).³⁵

Another issue that was not present at the time of the PETM but is becoming concerning is the rising potential for conflicts and war as the climate crisis renders parts of the earth unlivable. Water scarcity and resulting droughts one hand and extreme floodings will not only displace people but also have major impacts on food and water security. Some diseases and other conditions are also projected to have a higher incidence with warming climate: malaria, heat deaths, child undernutrition and diarrhea will cause 250,000 more deaths by 2050, with most of them occurring in Africa.³⁶

Table 1: Comparison between PETM and current anthropogenic climate change.

	PETM	Anthropogenic Climate Change
Time	Paleocene-Eocene border (~56 Ma ago)	Beginning of industrial era (~1850) ~180 years thus far, end open
Timespan of carbon release	<20 ka	Ongoing
Total duration	200 ka	
Carbon source	Debated, possible sources include methane clathrates, thawing permafrost and volcanic activity	Anthropogenic; mainly burning of fossil fuels and change in land use
Amount of carbon released	Between 2000-6000 GtC	670 GtC between 1850-2021, with an average of 10.8 ± 0.8 GtC / yr between 2012-2021 +1.15 °C as of 2022 (compared to pre-industrial temperatures)
Temperature change	+5-8 °C	Shift towards poles in geographical distribution of many species due to warming, several animal species already extinct. At 1.5 °C warming, 9% of species projected to go extinct. Potential for global conflict and war.
Biotic effects	Shift towards poles in geographical distribution of many organisms due to warming, some species of benthic foraminifera extinct	

1.3 Carbon Capture and Storage

The concept to capture CO₂ and store it somewhere safe or even utilize it for our needs has attracted increasing interest. In the case of CO₂, this process is known as carbon capture and storage/sequestration (CCS).³⁷ It has long been known that biologic sequestration is a possibility, since CO₂ is used by plants in the process of photosynthesis, yielding sugar and oxygen from CO₂ and water with the help of sunlight. Geological sequestration is a newer idea, in which CO₂ is captured and stored as a gas, liquid or in supercritical form in various underground formations, such as coal mines, oil fields or saline aquifers.³⁷ In the 2005 report “Carbon Dioxide Capture and Storage”, the IPCC postulates that CCS may enable us to capture 220-2,200 Gt of CO₂ until 2100, based on the assumption that atmospheric CO₂ levels could settle between 450-750 ppmv (parts per million by volume) CO₂.³ Viewing CCS as part of the entire mitigation process, the IPCC estimates that CCS would account for 15-55% of it. This of course depends on how many facilities would actually capture CO₂ as part of their processing chain, which in turn depends on the costs of CCS – the acceptable price range begins at less than 25 US\$ per ton of CO₂.³ Several companies are working in CO₂ sequestration nowadays, among them the Carbfix/Climeworks cooperation in Iceland³⁸ and the Quest Carbon Capture and Storage Project in Alberta, Canada.³⁹ Both of these examples use amine scrubbing (see below).

There are different options to capture CO₂ from the gas phase (known as carbon dioxide scrubbing), however, only one is currently being used at industrial scales: amine gas treating or scrubbing. It works by utilizing amines in an aqueous solution binding to CO₂ at lower temperatures and releasing them at higher temperatures through water vapor stripping. The water vapor is then condensated and CO₂ remains, which can then be treated further. Several problems with amine gas treating exist, notably the high energy cost and the degradation of the amines when they come into contact with O₂ and eventually also CO₂ degradation.⁴⁰

Nanoporous compounds such as zeolites, metal organic frameworks (MOFs) and COFs have been proposed as alternatives. These porous, crystalline materials have the ability to adsorb and store CO₂. However, their moisture sensitivity can cause them to stop being effective after a few cycles of CO₂ capture and desorption, since the gases that are to be scrubbed often contain some degree of moisture. MOFs have the additional problem of requiring metal ions, putting further strain on the already critical global demand for these metals.

2 Covalent Organic Frameworks

Covalent organic frameworks (COFs) were discovered by Coté *et al.* in 2005.¹ In their description of COFs, Coté and coworkers focus on the fact that they are porous, easy to produce, highly customizable and stable. In this, they are very similar to MOFs, however, COFs are only composed of light elements such as H, B, C, N and O. This is their biggest advantage over MOFs, which require metal ions or metal ion clusters together with an organic linker. This distinguishing feature makes COFs lighter and more environmentally sustainable. COFs usually have a high specific surface area due to their porosity, exceeding even those of zeolites and porous silicates, which have thus far been highly appreciated in technology for this reason.^{1,2}

2.1 Synthesis of COFs

As already mentioned, COFs allow high modularity and customizability. This is achieved through a principle similar to toy building blocks: monomers of different shapes with reactive sites are joined together according to a topological “blueprint” - each monomer attaching to the polymer backbone does so in a specific position with covalent bonds. Through this process, a network with empty spaces between the monomers is generated - these empty spaces are the coveted pores.²

Forming such networks with planar monomers yields 2D-COFs with layers that may be stacked atop one another using noncovalent interactions such as parallel stacking or van der Waals forces. Combining two different monomer geometries ([1 +1], so-called “knots” and “linkers”) allows the building of symmetric topologies, however, much greater diversity is afforded in combining one knot with different linkers, leading to multiple-component (MC) design and asymmetric topologies which are much more complex.²

3D-COFs are made possible by the addition of non-planar monomers, enabling the growth of the network into the third dimension. Orthogonal or tetrahedral (T_d) monomers serve as knots between C_1 , C_2 , C_3 , C_4 or T_d linkers.²

2.2 Uses of COFs

Thanks to their diversity in geometry, size and composition, COFs are currently one of the most promising substance classes with wide technological applications. COFs may be used as semiconductors, in luminescence and sensor applications, energy storage, mass transport (drug delivery), gas and vapor storage as well as separation (“molecular sieving”), as well as different types of catalysis.² Gas and vapor storage will be discussed in greater detail, as the COF studied in this work shows highly promising gas adsorption capabilities.

Several gases need efficient storage for different reasons: ammonia, hydrogen and methane as energy resources, while CO₂ as one of the main greenhouse gases needs to be captured and stored as part of carbon capture and sequestration.

As already mentioned, the customizability of COFs is one of their biggest advantages. To illustrate this point further: it is possible to tailor COFs to a specific purpose, which in the context of gas storage means a pore size and volume, linker and knot composition (functional groups) and surface area ideally suited to the goal at hand. In the case of CO₂, this means adding CO₂-philic units, making use of its quadrupole moment and improving electrostatic hydrogen bonding. However, it is important to figure out where to stop, as too many functional groups may take away space for CO₂ molecules. Examples of groups that aid CO₂ uptake are azines, amines, ionic liquids, and carboxylic acids.² CO₂ uptake capacities at 273 K and 1 bar pressure range from 31 mg g⁻¹ for iPrTAPB-TFPB⁴¹ to 218 mg g⁻¹ for TpPa-COF (Microwave),⁴² both of which are imine-linked. Among the azine-linked COFs, the lowest and highest CO₂ uptake value reported as of 2020 in a review by Geng *et al.* is 126 mg g⁻¹ for Acac-CTF-5-400⁴³ and 217 mg g⁻¹ for COF-JLU2,⁴⁴ respectively.

2.3 HEX-COF 1 as host for CO₂

HEX-COF 1 was first described as "*an azine-linked hexaphenylbenzene based covalent organic framework*" by Alahakoon *et al.* in 2016. The authors report that their "*aldehyde functionalized HEX monomer with six-fold symmetry*" was polymerized with hydrazine linkers, therefore ending up with a COF that has triangle shaped pores.

HEX-COF 1 has a pore size of about 1 nm, owing to the small azine linkers. The layers are arranged in such a way that pore canals are formed, though interlayer adhesion is described as weak in the publication, due to the missing face-to-face parallel-stacking as is usual in COF monomers.⁴

Synthesis of HEX was achieved by Alahakoon *et al.* in 2016 via a cobalt-catalyzed cyclotrimerization, with 4,4'-diformyltolan as precursor. Subsequently, HEX was polymerized to HEX-COF 1 in a "*solvent system of 2:1 mesithylene:dioxane with 0.1 mL of 6 M aqueous AcOH catalyst for 72 h at 120 °C*".⁴⁵

HEX-COF 1 was analyzed in the original publication using powder X-ray diffraction (PXRD), showing weak and broad peaks, suggesting poor long-range order. This is in contrast to the pore size distribution with a defined peak at 11 Å and the high surface area of 1214 m² g⁻¹, which points towards HEX-COF 1 being an ordered material. From a model of HEX-COF 1, a simulated PXRD pattern results in hexagonal P6/m unit cell parameters of 17.8 Å × 17.8 Å × 5.5 Å.⁴

Nitrogen adsorption measurements at 77 K and analysis with the Brunauer-Emmett-Teller (BET) model

show that the surface area is $1214 \text{ m}^2\text{g}^{-1}$. From this, the pore size distribution was modelled with a non-local density functional theory (NLDFT) carbon slit pore model, showing a majority of the pores to be 11 \AA wide.

Gas adsorption of CO_2 was also evaluated, demonstrating a strong affinity of HEX-COF 1 with CO_2 – at 273 K, almost 20 wt% were adsorbed, at 298 K, about 12 wt%, which is 200 mg g^{-1} and 120 mg g^{-1} , respectively. With CH_4 , about 2 wt% were able to be adsorbed at 273 K and about 1 wt% at 298 K. The authors argue that these great adsorption capabilities may be caused by the small pore sizes in combination with the polar azine functional groups.^{2,4,5}

3 Methods

3.1 Geometry Optimization

Optimizing a molecule's geometry is often the first step in computational chemistry, since it is not as computationally demanding as other methods. It involves searching the potential energy surface for minimum structures. Several methods are available, each with its own set of advantages and disadvantages. Before delving further into these methods, the potential energy surface shall be described in more detail, as it is of great importance.

The potential energy surface (PES) can be seen as a sort of "landscape" of the potential energy of a number of atoms, describing all possible atomic conformations. Minima in this hypersurface denote energetically ideal structures, where the potential energy is as low as possible. Saddle points are also of interests, as they describe transition states between minima.

When taking the entire configuration space as Q , each atom and each molecule has a configuration vector \mathbf{Q} with its coordinates. Moving atoms would thus imply adding or subtracting a $\Delta\mathbf{Q}$ vector to/from \mathbf{Q} . Considering a simple molecule such as CO₂ as an example, the configuration vector in Cartesian coordinates would be given as:

$$\mathbf{Q}_{\text{CO}_2} = \begin{pmatrix} \mathbf{r}_C \\ \mathbf{r}_{O_1} \\ \mathbf{r}_{O_2} \end{pmatrix} = \begin{pmatrix} x_C \\ y_C \\ z_C \\ x_{O_1} \\ y_{O_1} \\ z_{O_1} \\ x_{O_2} \\ y_{O_2} \\ z_{O_2} \end{pmatrix}$$

Concerning the complexity of the PES, it has $3N - 6$ coordinate dimensions ($3N - 5$ in case of linear molecules) and degrees of freedom, with N as the number of atoms; the potential energy is depicted in an additional dimension, so the depiction of the PES would end up having $3N - 5$ dimensions. For example, a diatomic molecule thus has only one degree of freedom ($3 \cdot 2 - 5 = 1$) and would be shown as a 2D plot. It is obvious that accurately depicting the PES for larger molecules becomes very difficult very quickly. However, the PES is a useful concept to describe geometry optimization, as the latter is concerned with finding minima of the PES and thus the optimal structure of the molecules.^{45,46}

3.1.1 First order methods of Geometry Optimization

First order methods employ the negative gradients of the potential energy surface in order to locate minima.

Steepest descent

The steepest descent method is based on simply setting the negative gradient ($d = -g_k$) as the descent direction d_k and descending in that direction with a step length α , see equation 6. This usually brings quick results in the early stage of an optimization, but quickly deteriorates in its effectiveness. Strictly speaking, this method will reach the minimum after an infinite number of steps. Another issue is associated to the determination of the step length α , as will be discussed below.^{45,47,48}

$$\mathbf{Q}' = \mathbf{Q} + \alpha \mathbf{d}_k \quad (6)$$

Conjugate gradient

For conjugate gradient, the step direction is determined by combining the direction from the previous iteration with a comparison factor γ , according to equation 7. When comparing the formula for conjugate gradient (equation 8) to the one for steepest descent (equation 6), it becomes obvious that conjugate gradient with a γ of 0 is essentially the same as steepest descent. The "conjugate" in the name already gives a clue as to what relation the gradients have to each other: they obey the conjugacy condition $\mathbf{d}_k^T \mathbf{d}_j = 0$ for $k \neq j$. This leads to the fact that the conjugate gradient method as described here needs a maximum of n steps if n is equal to the dimensions of space.⁴⁷⁻⁴⁹

$$\mathbf{d}_k^{cg} = \mathbf{d}_k + \gamma \mathbf{d}_{k-1}^{cg} \quad (7)$$

$$\mathbf{Q}' = \mathbf{Q} + \alpha \mathbf{d}_k^{cg} \quad (8)$$

Several options to choose the γ value are available, among the most commonly used being Fletcher-Reeves (FR) and Polak-Ribière (PR).⁴⁷⁻⁴⁹

$$\gamma_{k+1}^{FR} = \frac{\mathbf{d}_{k+1}^T \mathbf{d}_k + 1}{\mathbf{d}_k^T \mathbf{d}_k} \quad (9)$$

$$\gamma_{k+1}^{PR} = \frac{\mathbf{d}_k^T (\mathbf{d}_k - \mathbf{d}_{k-1})}{\mathbf{d}_{k-1}^T \mathbf{d}_{k-1}} \quad (10)$$

Determining the step length α_k

As already mentioned, choosing an ideal step length is an issue in geometry optimizations. Both a too small and a too large step length are problematic and hamper progress. Two main options exist to optimize the step length, namely line search and trust region approaches.

Line search works by minimizing the one-dimensional auxiliary function seen in equation 11 with α_k

as the step length, x_k as the current point and d_k as the direction as defined earlier.

$$h(\alpha_k) = f(x_k + \alpha_k d_k) \quad (11)$$

In case of finding the optimal solution, $h(\alpha_k)$ is minimized exactly. Otherwise, it is approximated if it decreases enough.^{47,49}

The *trust region* approach is based on fitting a model function m_k (see equation 12) to the function f at the point x_k and searching for m_k 's minimum only in a certain area close to x_k - this is the trust region. It is evaluated by calculating ρ according to equation 13 in order to determine whether the fit is good ($\rho \approx 1$, often between 0.9 and 1.1), in which case the trust region is expanded. If ρ falls below 0.5 or goes above 1.5, the trust region is decreased.^{47,49}

$$m_k = f(x_k) + \mathbf{g}_k^T \boldsymbol{\alpha} \quad (12)$$

$$\rho = \frac{\Delta V_{actual}}{\Delta V_{predicted}} = \frac{f(x_k) - f(x_k + \alpha_k)}{m_k(0) - m_k(\alpha)} \quad (13)$$

3.1.2 Second order methods of Geometry Optimization

Second order methods work with the first and second derivatives of the potential energy function, which means they use information about the function's curvature.

Newton-Raphson Method

The Newton-Raphson method involves calculating the Hessian (matrix of second derivatives) and thus obtaining information about the potential energy surface's curvature.

In order to use it for minimization in the one-dimensional case, $f(x_{k+1})$ is approximated according to equation 14 and this yields the iterative scheme outlined in equation 15.⁴⁵⁻⁴⁹

$$f(x_{k+1}) \approx f(x_k) + (x_{k+1} - x_k)f'(x_k) + \frac{1}{2}(x_{k+1} - x_k)^2 f''(x_k) \quad (14)$$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad (15)$$

To be able to describe systems in higher dimensions, equation 14 is changed to 16, where \mathbf{p} is a search vector and \mathbf{H}_k is the Hessian.

$$f(\mathbf{x}_k + \mathbf{p}_k) \approx f(\mathbf{x}_k) + \mathbf{g}(\mathbf{x}_k)^T \mathbf{p}_k + \frac{1}{2} \mathbf{p}_k^T \mathbf{H}(\mathbf{x}_k) \mathbf{p}_k \quad (16)$$

$$\mathbf{H}_k \mathbf{p}_k = -\mathbf{g}_k \quad (17)$$

$$x_{k+1} = x_k - \mathbf{H}^{-1} \mathbf{g}_k \quad (18)$$

As can be seen, equation 18 implies that using the Newton-Raphson method involves solving a set of linear equations at each step in order to update the Hessian. This is unfortunately a computationally costly process of at least $\mathcal{O}(n^3)$.⁴⁵⁻⁴⁹

Quasi-Newton Methods

Quasi-Newton methods iteratively build \mathbf{B}_k^{-1} , an approximation of the inverse Hessian. The iteration process from equation 18 thus turns into equation 19, while \mathbf{B}_{k+1}^{-1} must fulfill the *quasi-Newton condition* in equation 20.

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{B}_k^{-1} \quad (19)$$

$$\mathbf{B}_{k+1} \mathbf{s}_k = \mathbf{y}_k \quad (20)$$

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \quad (21)$$

$$\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k \quad (22)$$

Among the most-used quasi-Newton methods is the BFGS-method, named after Broyden, Fletcher, Goldfarb and Shanno, who developed it in 1970.⁵⁰⁻⁵³ The BFGS update formula is given in equation 24, with \mathbf{U}_k denoting a symmetric and positive-definite (all eigenvalues are positive) update matrix of rank 2.

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \mathbf{U}_k(\mathbf{s}_k, \mathbf{y}_k, \mathbf{B}_k) \quad (23)$$

Expanding on this, the updates are calculated following equation 24, usually implemented to directly calculate the inverse of \mathbf{B}^{-1} according to equation 25.

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k^T}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \quad (24)$$

$$\mathbf{B}_k^{-1} = \left(\mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) \mathbf{B}_k^{-1} \left(\mathbf{I} - \frac{\mathbf{y}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \quad (25)$$

The BFGS search vector \mathbf{p}_k , similar to equation 18, follows equation 26.

$$\mathbf{p}_k = -\mathbf{B}_k^{-1} \mathbf{g}_k \quad (26)$$

The approximation for the Hessian at the start is usually simply set as the identity matrix.

A limited-memory version of the BFGS method is often used - also in this work - in order to save on memory, called L-BFGS.⁵⁴ In this, only the data obtained from the last few iterations is stored and used for the next step. To be more precise, only n numbers of pairs of $\mathbf{s}_k, \mathbf{y}_k$ are stored, with n usually between 3-10, sometimes also up to 20. Once a new pair of $\mathbf{s}_k, \mathbf{y}_k$ is calculated, the oldest one left is removed from the set.^{47,49,54}

3.2 Simulations

Simulations are different from geometry optimization methods in that they make it possible to study the system's behavior at temperatures different from 0 K. Two main families of simulation approaches are available: Monte Carlo (MC) and Molecular Dynamics (MD). It is also possible to combine the two in MC/MD hybrid applications.

3.2.1 Statistical Mechanics

Statistical mechanics make it possible to examine *ensembles*, representative system states with the help of statistics and probability. According to the *ergodic theory*, an average of a sufficient number of system states is equivalent to the respective time average. Different constraints may be imposed on the system, resulting in different types of statistical ensembles: a constant particle number N , volume V and energy E lead to the *microcanonical ensemble (NVE)*, while a constant particle number N , volume V and temperature T gives the *canonical ensemble (NVT)*. In the *grand canonical ensemble (μVT)*, the chemical potential μ , volume V and temperature T are constant. Another important ensemble is the *isothermal-isobaric ensemble (NpT)*, in which particle number N , pressure p and temperature T remain fixed.⁴⁸

From the energy of the configurations of these system states, a *state function* Q can be derived, making it possible to extract thermodynamic properties such as the *entropy S* and the *Gibbs free energy G*. The energy U of the configurations is made up of the *kinetic energy* E_{kin} and the *potential energy* E_{pot} .

$$U = E_{kin} + E_{pot} \quad (27)$$

In order to compute the kinetic energy, equation 28 can be used, with N as the number of particles, \mathbf{p}_i as the linear momentum, m_i as particle mass, \mathbf{v}_i as particle velocity, k_b as the Boltzmann constant and T as temperature. The pressure p can be calculated according to the virial theorem of Clausius, equation 29, with \mathbf{F}_{ij} as the force and \mathbf{r}_{ij} as the distance between two particles.⁵⁵

$$E_{kin} = \sum_{i=1}^N \frac{|\mathbf{p}_i|^2}{2 \cdot m_i} = \sum_{i=1}^N \frac{m_i \cdot |\mathbf{v}_i|^2}{2} = \frac{3}{2} \cdot N \cdot k_b \cdot T \quad (28)$$

$$p = \frac{1}{V} \left[N \cdot k_b \cdot T - \frac{1}{3} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mathbf{F}_{ij} \cdot \mathbf{r}_{ij} \right] \quad (29)$$

3.2.2 Molecular Dynamics

Molecular dynamics (MD) is a method to simulate the time-dependent motion of atoms based on a set of starting coordinates and particle velocities in combination with an interatomic potential. With the help of the second Newtonian equation of motion, the propagation of the system is calculated and subsequently evaluated; the new particle positions and velocities can then be used as input for the next simulation loop, as depicted in figure 1. Different options are available to integrate Newton's

equations of motion, for example the velocity Verlet algorithm, which was used in this study and will be described in more detail at a later point. The time steps Δt for MD-based simulations are quite small, with femtoseconds (10^{-15} s) being a typical order of magnitude.

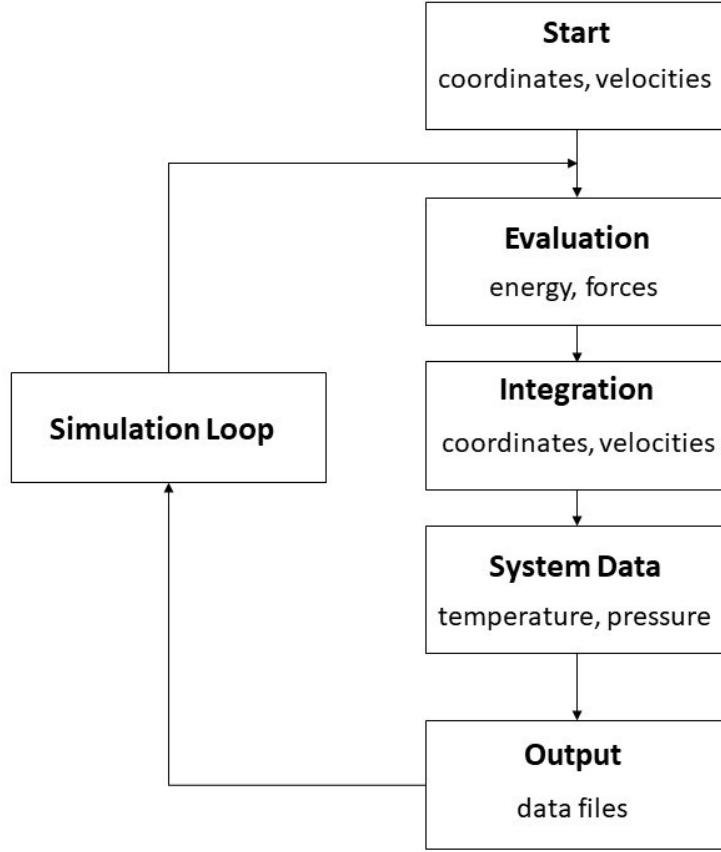


Figure 1: Schematic drawing of the simulation loop. The *integration* step refers to the integration of the equations of motion.

An interesting phenomenon in MD simulations is the fact that they are theoretically deterministic - running two MD simulations with the exact same starting conditions should in theory yield the same results. However, due to differences in computer architecture and rounding errors, MD simulations are oftentimes non-deterministic and produce different outcomes from the same input i.e. they are chaotic.⁴⁸ As already mentioned, the input required to start an MD simulation are the atom coordinates and their velocities. The forces on each particle are then calculated according to equation 30 using the first derivative of the potential energy with respect to x_j^i being the particle positions of atoms in three-dimensional space:

$$\mathbf{F}_j^i = -\frac{\partial E_{pot}(x_1^i, x_2^i, \dots, x_n^i)}{\partial x_j^i} \quad (30)$$

Incorporating the mass m_j into this equation, one arrives at Newton's second equation of motion:

$$\mathbf{F}_j^i = -\frac{\partial E_{pot}(x_1^i, x_2^i, \dots, x_n^i)}{\partial \mathbf{x}_j^i} = m_j \frac{\partial^2 \mathbf{x}_j^i}{\partial t^2} \quad (31)$$

In order to solve this differential equation, different approaches have been developed: the Euler method comprises the simplest option, however, it also does not perform too well. In contrast, Runge-Kutta and predictor-corrector methods deliver good results, but they are computationally more demanding. Between these extremes lie the Verlet and velocity Verlet methods as well as the leapfrog algorithm, the latter being very similar to the velocity Verlet algorithm. The latter was used as the time integration algorithm in this study, therefore it will be described in more detail. It is based on four steps, described by equations 32 to 35, with \mathbf{v} as the velocity, t as time, \mathbf{F}_t as the forces, \mathbf{x}_t as particle positions at time t , m as mass, and E_{pot} as the potential energy; it is also depicted schematically in figure 2.^{46–48}

$$\mathbf{v}_{t+\Delta t/2} = \mathbf{v}_t + \frac{\Delta t}{2} \frac{\mathbf{F}_t}{m} \quad (32)$$

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \Delta t \cdot \mathbf{v}_{t+\Delta t/2} \quad (33)$$

$$\mathbf{F}_{t+\Delta t} = -\frac{\partial E_{pot}}{\partial \mathbf{x}_t} \quad (34)$$

$$\mathbf{v}_{t+\Delta t} = \mathbf{v}_{t+\Delta t/2} + \frac{\Delta t}{2} \frac{\mathbf{F}_{t+\Delta t}}{m} \quad (35)$$

The first step (equation 32) is to calculate the velocity after half the time step $\mathbf{v}_{t+\Delta t/2}$. Next, the particle positions after a full time step $\mathbf{x}_{t+\Delta t}$ are calculated (equation 33). In the third step, the forces at the full time step $\mathbf{F}_{t+\Delta t}$ are computed using equation 34. Finally, the fourth step is to calculate the particle velocities after the full time step $\mathbf{v}_{t+\Delta t}$ (equation 35).

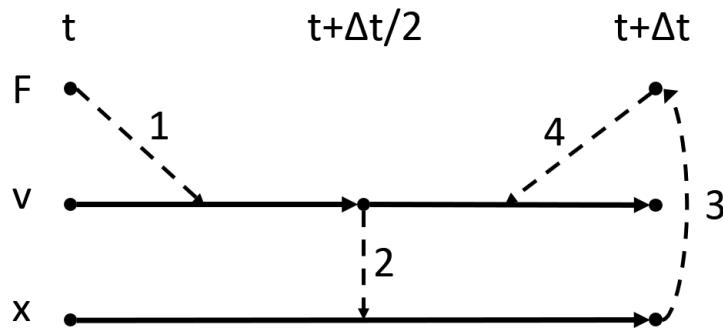


Figure 2: Basic velocity Verlet scheme.

3.2.3 Monte Carlo Methods

Monte Carlo (MC) simulation methods are based on random changes in atom geometry. In contrast to MD, they do not consider time or atomic velocities. New geometries are generated by randomly moving one randomly chosen particle or molecule. The energy change ΔE caused by this change is evaluated and compared to the previous configuration - if the energy has decreased, the new configuration is accepted. If it has increased, a random number between 0 and 1 is compared to the corresponding Boltzmann factor $e^{-\Delta E/kT}$ according to the method proposed by Metropolis *et al.* in 1953.⁵⁶ If the Boltzmann factor is larger than the random number, the new configuration is not accepted and the old configuration is taken as base for the next MC step. Otherwise, the new configuration is accepted. A target acceptance ratio is set to usually around 30-50%.^{46,48,56}

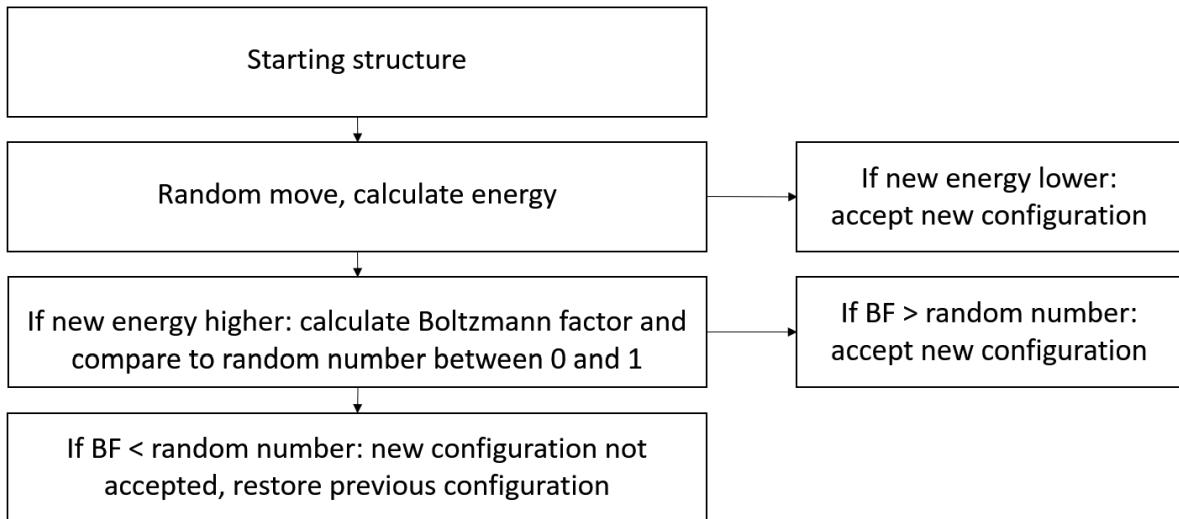


Figure 3: Scheme of the Metropolis Monte Carlo method.

3.2.4 Non-natural Ensembles: Temperature and Pressure Control

A crucial point in MD simulations is the comparability to experimental results - since a regular MD simulation produces an NVE (microcanonical) ensemble in which temperature and pressure vary over time, techniques to control these state properties are necessary. An NVT (canonical) ensemble where the temperature is constant or an NpT (isothermal-isobaric) ensemble with constant temperature and pressure are often desired and achievable with various techniques. Control of temperature is achieved either by hard scaling or a thermostat, while pressure is controlled by a barostat, also referred to as manostat. While there are more options available for thermostats as well as manostats, only the ones used in this work will be described in greater detail. Examples for other thermostats include the Nosé-Hoover and more advanced Nosé-Hoover chain algorithms and Langevin dynamics; for pressure control, the Berendsen manostat or a Nosé-Hoover manostat are also available.⁴⁶⁻⁴⁸

Hard Scaling

Hard scaling is usually the simplest way to adjust the temperature during the equilibration period, since it works quickly. The basic principle is scaling the velocities \mathbf{v}_i of the particles, making use of the fact that the temperature is connected to the kinetic energy E_{kin} according to equation 28.

A scaling factor λ is calculated according to equation 36 and the particle velocities are multiplied by it following equation 37. λ will be greater than 1 if the system temperature T is smaller than the target temperature T_0 .

$$\lambda = \sqrt{\frac{T_0}{T}} \quad (36)$$

$$\mathbf{v}'_i = \mathbf{v}_i \cdot \lambda \quad (37)$$

Strictly speaking, hard scaling will cause the simulation to no longer be in the canonical ensemble since the kinetic energy will no longer change, thus violating statistical mechanics for the canonical ensemble.⁴⁸ However, due to the fast temperature adjustment, it is often used to quickly cool or heat a system before sampling runs with a more sophisticated thermostat.

Berendsen Thermostat

The Berendsen thermostat introduces a relaxation time τ . It may be understood as the system being coupled to an external heat bath with the fixed target temperature T_0 , removing or adding kinetic energy with the scaling factor λ , calculated by equation 38. T is the current system temperature and Δt the time step. The new velocities are again computed according to equation 37.⁵⁷

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau} \cdot \left(\frac{T_0}{T} - 1 \right)} \quad (38)$$

Monte-Carlo Manostat

With a Monte-Carlo manostat, the volume is changed by a random amount ΔV within bounds of $-\Delta V_{max}$ and $+\Delta V_{max}$, which is uniformly randomly distributed. The maximum trial step size $\Delta V_{max}/V$ is arbitrary. The acceptance ratio N_{acc}/N is checked against a set target acceptance ratio every N number of steps and, if necessary, maximum trial step size $\Delta V_{max}/V$ is adjusted.⁵⁸

$$\Delta W = E'_{Pot} - E_{Pot} + p_0 \Delta V - N \cdot k_b \cdot T \cdot \ln \left(\frac{V'}{V} \right) \quad (39)$$

with

$$V' = V + \Delta V \quad (40)$$

E'_{pot} and V' being the potential energy and the new volume after the volume change; p_0 corresponds to the target pressure.

3.2.5 Constraints - SHAKE/RATTLE Algorithm

Constraints of bonds are often useful to increase the MD time step in the simulation - this is because the maximum possible size of the time step depends on the fastest oscillations in the system. Often, bonds hydrogen-containing bonds (i.e., X-H, with X being C, N, O) are constrained to a certain length because of the much higher vibrational frequency resulting from the small mass of hydrogen atoms.^{45,46,48} The SHAKE algorithm was developed by Ryckaert *et al.* in 1977.⁵⁹ The basic principle is first letting atoms move unconstrained and subsequently correcting their positions according to the constraints.

Constraints of different geometry features are possible: bond lengths, angles and/or torsion angles. Constraining the bond length with \mathbf{r}_i and \mathbf{r}_j as the positions of atoms i and j respectively and d_{ij} as the target bond length yields equation 41.

$$\sigma_l = ||\mathbf{x}_i - \mathbf{x}_j||^2 - d_{ij}^2 = 0 \quad (41)$$

Integration of the equations of motion must follow this constraint, therefore equation 34 is extended, yielding equation 42. λ_l is a Lagrangian multiplier dependent on time only, denoting the interference of the constraint σ_l in order to arrive at the desired position of the atoms. l is the number of bonds.

$$\mathbf{F}_i = -\frac{\partial}{\partial \mathbf{x}_i}(E_{pot} + \sum_{l=1}^{N_c} \lambda_l \sigma_l) \quad (42)$$

With added constraints, the total force acting on a particle \mathbf{F}_{tot} is composed of two parts: potential energy E_{pot} and the force due to the constraint \mathbf{F}_{con} , the latter of which can be described using equation 43, which is derived by differentiating the last term of equation 42.⁵⁹

$$\mathbf{F}_{con} = -\sum_{l=1}^{N_c} \lambda_l(t) \nabla_i \sigma_l \quad (43)$$

In SHAKE, two approximations are made in order to simplify, and an iterative execution of the algorithm becomes necessary:

1. Quadratic terms in the determination of λ_l are ignored, thus linearizing the equation.
2. Each constraint is viewed as independent and decoupled from all other constraints.

A modified version of the SHAKE procedure is the RATTLE algorithm; it works together with the SHAKE algorithm and is adapted for use with the velocity Verlet integrator. The RATTLE routines eliminate velocity components parallel to the constraint bonds at the end of each velocity Verlet cycle.⁶⁰

3.2.6 Periodic Boundary Conditions

In simulations, the system under investigation is set up in a simulation cell. This simulation cell can be regarded as being surrounded by identical copies along all directions up to infinity. A 2D version of this is shown in figure 4, with the cell in the center (shaded) being surrounded by eight periodic replicas. In 3D, this would amount to 26 neighboring cells. A particle moving outside the simulation box reenters it

again on the opposite side as a copy (image) of itself. It is essential to avoid self-interactions of particles with their own image and double-interactions of particles with both the original and an image of another particle. In order to solve this issue, the *minimum-image convention* is introduced: forces or energies between particles are only calculated for the closest version, whether that be the original or its image. A *non-bonded cutoff* is employed in order to avoid having to calculate interactions between particles more than a specified cutoff distance apart, since these terms are usually negligibly small.^{46,48}

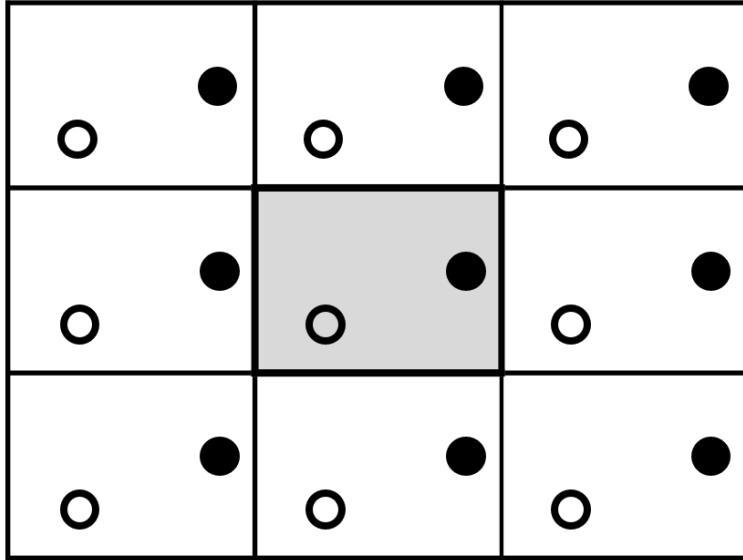


Figure 4: Periodic boundary conditions in 2D.

3.2.7 Diffusion

The self-diffusion coefficient is defined by the IUPAC as the "*diffusion coefficient D_i^* of species i in the absence of a chemical potential gradient*".⁶¹ In this work it will be referred to as D . In order to calculate D from simulation trajectories, the Einstein relation (equation 44) can be used. It is an example of a time correlation function, where the change over time of a system property is analyzed. d denotes the dimensionality of the system, t the time, and \mathbf{x}_0 and \mathbf{x}_t the positions of particles at the start time and at time t , respectively.

$$D = \frac{1}{2d} \lim_{t \rightarrow \infty} \frac{\langle ||\mathbf{x}_t - \mathbf{x}_0||^2 \rangle}{t} \quad (44)$$

For the calculation, the trajectory is divided into overlapping segments called windows. The following steps are then taken:

1. For all windows: compute

$$||\mathbf{x}_t - \mathbf{x}_0||^2$$

2. Estimate the window average:

$$\langle ||\mathbf{x}_t - \mathbf{x}_0||^2 \rangle$$

3. Perform linear regression, making sure to only use the linear part of the curve (see below):

$$\langle ||\mathbf{x}_t - \mathbf{x}_0||^2 \rangle = 2dDt; D = \frac{k}{2d}$$

According to Einstein, it is important to differentiate between the ballistic and the diffusive regime in particle transport, since only the diffusive regime should be used to determine the diffusion coefficient.⁶² In order to distinguish between the two, a graph of $\langle ||\mathbf{x}_t - \mathbf{x}_0||^2 \rangle$ plotted against the time t can be made. The diffusive regime begins when the curve becomes linear.

The activation energy of diffusion can be calculated from the diffusion coefficient D using equation 45, in which E_a refers to the average activation energy, R to the molar gas constant, T to the temperature, and D_0 to the pre-exponential factor. Plotting $\ln(D)$ against the reciprocal of T yields an Arrhenius plot, with $-E_a/R$ as the slope; for an example, see figure 10.

$$\ln(D) = -\frac{E_a}{R} \frac{1}{T} + \ln(D_0) \quad (45)$$

3.3 Neural Networks

Artificial neural networks (ANNs) have become an increasingly active field of research in many areas, not only in science but also everyday life. With the capability to boost efficiency compared to traditional methods, they find applications in such diverse fields as cybersecurity, language processing, medicine, infrastructure management, theoretical chemistry and many more.⁶⁻⁸

ANNs are based on artificial neurons, which are modelled after neurons found in living organisms. Each neuron is given an input which it reacts to using an activation function, producing an output which is then forwarded to the next neuron, which repeats the process. The connection between the neurons is called an edge and is assigned a real number value, a so-called weight. The weights determine the "strength" of the signal at the receiving end. In addition, each neuron is associated to an additive constant referred to as bias. Different kinds of activation functions processing the neuronal input based on the weights and biases have been developed for various applications. Typically, neurons are grouped into layers, with one input layer, several hidden layers and an output layer. The entirety of this conglomerate is then termed an artificial neural network.

Every neural network model has a certain pattern of connection between the neurons, with one model only allowing forward connections between neighboring layers (as in figure 5), while others may allow backwards propagation or connections within a layer.⁶

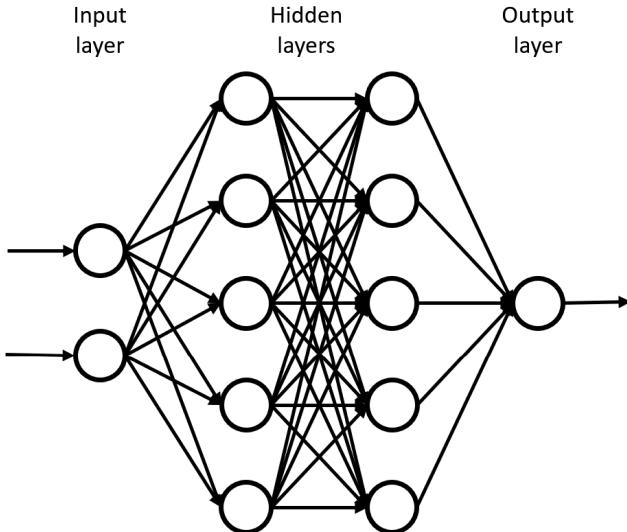


Figure 5: Schematic drawing of the layers of an artificial neural network. Each circle represents a neuron, the arrows correspond to edges with associated weights w_{ij} .

Neural Network Potentials

Neural network potentials (NNPs) are ANNs capable of computing the potential energy surface (PES) of chemical systems. Previously, a choice between accurate but computationally costly quantum mechanic (QM) methods such as density functional theory (DFT) or less accurate but also computationally less intense methods like force fields or empirical potentials had to be made. The advantage of NNPs is that they can be trained on highly accurate QM data and are, thereafter, capable of reproducing the PES at QM level but reduced computational cost. Dedicated NNPs are based on a certain system or a few similar systems, while transferable ML potentials enable simulation of an entire class of molecules. The former always require at least some reference data from QM calculations, while transferable NNPs avoid this need once trained.^{10,63}

ANI Deep Learning Molecular Potential

The ANAKIN-ME (Accurate NeurAl networK engINe for Molecular Energies) or ANI method "for building neural network-based molecular potentials"⁶⁴ was developed in 2017 by Smith *et al.* The first version of the associated neural network potential, ANI-1, was trained on 57,951 small molecules with the elements H, C, N, and O in about 17.2 million individual conformations. In 2019, ANI-1x and ANI-1cxx were released, trained on data sets ANI-1x (several QM properties based on DFT calculations) and ANI-1cxx (data points from an approximation of CCSD(T)/CBS) respectively. In 2020, the group published the next version ANI-2x, with the new elements S, F, and Cl added. These elements were chosen since they, together with H, C, N and O from previous ANI versions, account for about 90% of drug-like molecules. The authors claim a speed-up of factor 10^6 for ANI-2x compared to DFT while maintaining subchemical accuracy, i.e. $\leq 1 \text{ kcal/mol}$.¹⁰

The ANI method uses an adapted version of symmetry functions as described by Behler and Parinello in 2007⁶⁵ in order to compute atomic environment vectors (AEVs) $\mathbf{G}_i^X = \{G_1, G_2, G_3, \dots, G_M\}$, with each contribution describing an aspect of a single atom's local environment (radial and angular). This is in line with the atomistic perspective, which states that an extensive property (such as the energy) may be seen as the total of single-atom contributions, as described in equation 46.^{64,65}

$$E_{total} = \sum_i^{all atoms} E_i \quad (46)$$

A piecewise cutoff function is introduced, with R_{ij} as the interatomic distance and R_C as the cutoff radius.^{10,64,65}

$$f_C(R_{ij}) = \begin{cases} 0.5 \cdot \cos\left(\frac{\pi R_{ij}}{R_C}\right) + 0.5 & \text{for } R_{ij} \leq R_C \\ 0.0 & \text{for } R_{ij} > R_C \end{cases} \quad (47)$$

The radial contributions G_m^R to the AEV G_i^R are calculated with the original Behler-Parinello symmetry function, see equation 48, where m is an index over a set of η and R_s parameters. η has an influence over the width of the Gaussian curve, R_s changes the position of the peak.^{64,65}

$$G_m^R = \sum_{i \neq j}^{all atoms} e^{-\eta(R_{ij}-R_s)^2} f_C(R_{ij}) \quad (48)$$

The angular contributions $G_m^{A_{mod}}$ to G_i^X use a modified version of the original Behler-Parinello symmetry function. Sets of three atoms i , j , and k are considered, with the angle θ_{ijk} centered on i and the two distances R_{ij} and R_{ik} between the respective atoms. m now covers four parameters: ζ , changing the peak widths, θ_s , making possible a shift in the angular environment, and η and R_s working like in the radial environment.⁶⁴

$$G_m^{A_{mod}} = 2^{1-\zeta} \sum_{j, k \neq i}^{all atoms} (1 + \cos(\theta_{ijk} - \theta_s))^{\zeta} e^{-\eta\left(\frac{R_{ij}+R_{ik}}{2}-R_s\right)^2} f_C(R_{ij}) f_C(R_{ik}) \quad (49)$$

The dataset with the molecules for the training of ANI-1 was developed based on the GDB-11 database, with up to 8 atoms of C, N and O, resulting in a total of 57,951 molecules. For each of these molecules with 8 or less heavy atoms, a set of non-equilibrium conformations was generated using normal mode sampling $K = S(3N - 6)$ where S depends on the number of heavy atoms in the molecule and N is the molecule's number of atoms. S ranges from 500 for molecules with only one heavy atom to 5 for ones with 8 heavy atoms. During testing, ANI-1 was presented with molecules of 10 to 24 heavy atoms, up to a total atom count of 53 in order to show how transferable it truly is.

Normal mode sampling (NMS)

The data points for the ANI method need to contain information about the molecular coordinates (whether energy minimized or not does not play a role), denoted by \mathbf{q} , as well as the single point energy

E_T computed at the desired level of QM theory. The point of normal mode sampling (NMS) is to obtain relevant reference data within a certain range around the minimum energy conformation. The following steps are taken in NMS for N_a atoms forming a molecule that has undergone energy minimization.

1. Obtain a set of N_f normal mode coordinates $Q = \{q_1, q_2, q_3, \dots, q_{N_f}\}$ at the necessary *ab initio* level of theory. $N_f = 3N_a - 6$ for all molecules except linear ones, where $N_f = 3N_a - 5$. Also get the respective force constants $K = \{K_1, K_2, K_3, \dots, K_{N_f}\}$.
2. Generate a set of uniformly distributed pseudo-random numbers c_i so that the sum over all c_i is between 0 and 1.
3. Calculate displacements R_i for each coordinate according to equation 50, with N_a as Avogadro's constant and k_b as Boltzmann's constant. In order to ensure that the sampling is equal for both sides of the harmonic potential, a Bernoulli distribution with $p = 0.5$ is used to determine R_i 's sign. The temperature T is chosen appropriate to the purpose of the trained neural network.

$$R_i = \pm \sqrt{\frac{3c_i N_a k_b T}{K_i}} \quad (50)$$

4. Scale normalized normal mode coordinate by R_i : $q_i^R = R_i q_i$.
5. Use the superposition of all q_i^R , Q^R , to displace energy minimized coordinates.
6. With the displaced coordinates from the last step, calculate a new single point energy E_T at the desired level of QM theory.

Network weights are trained by back-propagation from the exponential cost function (equation 51), obtaining a cost derivative for each weight. The vector \mathbf{E}^{ANI} contains all energy outputs, E_i^{ANI} for coordinates i and E_i^{DFT} the same for DFT comparison. τ is given the value 0.5 to achieve optimum performance. The advantage of the exponential cost function lies in its robustness when confronted with outliers. Optimization of the weights is done using the ADAM update method.^{64,66}

$$C(\mathbf{E}^{\text{ANI}}) = \tau \exp \left(\frac{1}{\tau} \sum_i (E_i^{\text{ANI}} - E_i^{\text{DFT}})^2 \right) \quad (51)$$

Active learning

Active learning for ANI was introduced in ANI-1x and is a procedure that enables the use of an ML model to automate data selection for the next training generation and thus support the model's capability for prediction. The mechanism ANI uses is 'query by committee', wherein several models ("students" forming a "committee") are trained simultaneously on comparable data. The committee is then presented with new data and the disagreement is determined; the data with the highest disagreement is then accepted into the data set. For ANI, the disagreement is measured by the uncertainty estimate ρ

which is calculated according to equation 52; σ_i denotes an ensemble's standard deviation of predictions, N_i the number of atoms.^{67,68}

$$\rho = \frac{\sigma_i}{\sqrt{N_i}} \quad (52)$$

ρ is then compared to the inclusion criterion $\hat{\rho}$, which is determined empirically to include 98% of data.⁶⁷

This is done by computing ε_i , which measures the largest error of per atom prediction compared between models of the ensemble using equation 53.

$$\varepsilon_i = \frac{|MAX(\{E_T^{ANI}\}_i^{ens} - E_{T,i}^{REF})|}{\sqrt{N_i}} \quad (53)$$

Smith *et al.* were thus able to select the value of $\hat{\rho}$, which includes 98% of all errors with $\varepsilon > 1.5$ kcal/mol by plotting ε against ρ . For ANI-1x, the sampling procedures used were diverse normal mode sampling (a variation of NMS, the latter of which is outlined above), molecular dynamics sampling (MDS), molecular dynamics dimer sampling (MDDS) and N -trajectory MD sampling. In addition in case of ANI-1cxx, torsion sampling (TS) was used.^{67,68}

Molecular dynamics sampling works by running an MD simulation of a random molecule at a random temperature and regularly computing the uncertainty estimate ρ for the structure x at that time step. In case ρ is bigger than a previously set threshold, the MD run is terminated and the current molecular structure x is added to a set \bar{X} . This set contains structures with a high ρ value. Finally, DFT calculations for \bar{X} are carried out and the thus obtained data is added to the training data set.

Dimer sampling: based on the conformer sampling set, dimers are randomly generated by filling a simulation box under periodic boundary conditions with molecules. An MD simulation is then carried out and, at regular intervals, the cell is deconstructed and dimer pairs within a specified cutoff radius are tested using the 'query by committee' procedure. If ρ of the dimer exceeds $\hat{\rho}$, it is added to the set \bar{X} and DFT data is generated for all structures in \bar{X} .

N-random trajectory sampling, as it was called in Devereux *et al.* 2020 for ANI-2x,¹⁰ also referred to as *K-random trajectory sampling* by Smith *et al.* in 2018,⁶⁷ is also based on short MD simulations with an ANI model ensemble; the particles are given random initial velocities corresponding to 300 K and the system is heated to 1000 K over the run. 'Query by committee' is again used to check each step for $\rho > \hat{\rho}$ – in case the criterion is met, the simulation is stopped and the structure added to \bar{X} , for which DFT data is generated and added to the training data set.

Torsion sampling: after obtaining the ANI-1x data set, torsion sampling active learning is performed additionally for the ANI-1cxx potential. A bond capable of rotation is chosen and optimized using the current ML potential generation. Every 10°, a scan is performed and ρ is determined. If $\rho > \hat{\rho}$, the scan is stopped and the structure is saved. Normal modes and harmonic force constants are computed for

the structure with the current ANI generation. Four slightly different structures are generated by perturbing the structure along the normal modes and are added to the set \bar{X} , for which DFT data is again generated and added to the training set.^{10,68} In ANI-2x, normal mode sampling, MD dimer sampling, N -random trajectory sampling and torsion sampling active learning cycles were performed to train the neural network. Additionally, ANI-2x was trained on nonbonded interactions to achieve an improved representation of these complex contributions. It was also trained for bulk water.¹⁰

Force training in ANI: while the first versions of ANI-1 didn't include force calculation,⁶⁴ the parametrization procedure of ANI-1x and ANI-1cxx started incorporating force predictions without force training being necessary by using a "*component-wise mean of the forces from the L individual ANI models*".⁶⁷ ANI-2x was the first of the ANI models to have explicitly been trained to energies as well as forces, with the predicted forces being the analytical derivatives of the potential energy with respect to the atomic positions.⁶⁸ For this purpose, ANI-2x was trained with the loss function in equation 54, where N denotes the number of systems and M the number of atoms per system, \hat{E}_i and \hat{f}_{ij} are the molecules' energies and forces (predicted), while E_i and f_{ij} correspond to the respective QM reference. l_0 is a factor that keeps force and energy terms in balance while training. It is important in order to keep the network in check to avoid focusing on learning force data while neglecting the energy contributions. The forces do not provide information about the molecule's absolute energy.¹⁰ For a more detailed discussion of the force training algorithm used for ANI, see Smith *et al.* 2020 b.⁶⁹

$$L = \frac{1}{N} \sum_{i=1}^N \left[(\hat{E}_i - E_i)^2 + \frac{l_0}{M_i} \sum_{j=1}^{M_i} (\hat{f}_{ij} - f_{ij})^2 \right] \quad (54)$$

3.4 Simulation Protocol and Data Analysis

In the following the applied simulation settings and analysis methods are outlined. All systems considered in this work were treated in an orthorhombic simulation cell subject to periodic boundary conditions. The simulation time step Δt was set to 2.0 fs, and time integration was performed via the velocity Verlet algorithm. C-H bond lengths were constrained using the SHAKE/RATTLE algorithm^{59,60} to an ensemble average obtained from trial simulation trajectories of 50 ps at 298.15 K and 1.013 bar, the exact value depending on whether it was an oleofinic or an aromatic bond (i.e. 1.103 027 7(8) Å or 1.090 839(4) Å, respectively).

The ANI-2x neural network potential was applied in the description of energy and forces. Temperature control was achieved with a Berendsen thermostat with the relaxation time $\tau = 0.1ps$. In addition, the pressure was controlled to a value of 1.013 bar using a Monte-Carlo (MC) manostat,⁵⁸ with a trial volume change every 25 MD steps (which is equal to 50 fs). The maximum trial step size was given a start value of 1e-4 and as described in section 3.2.4 was checked every 10 MD steps against the target acceptance ratio $\frac{N_{acc}}{N}$ of 30%.

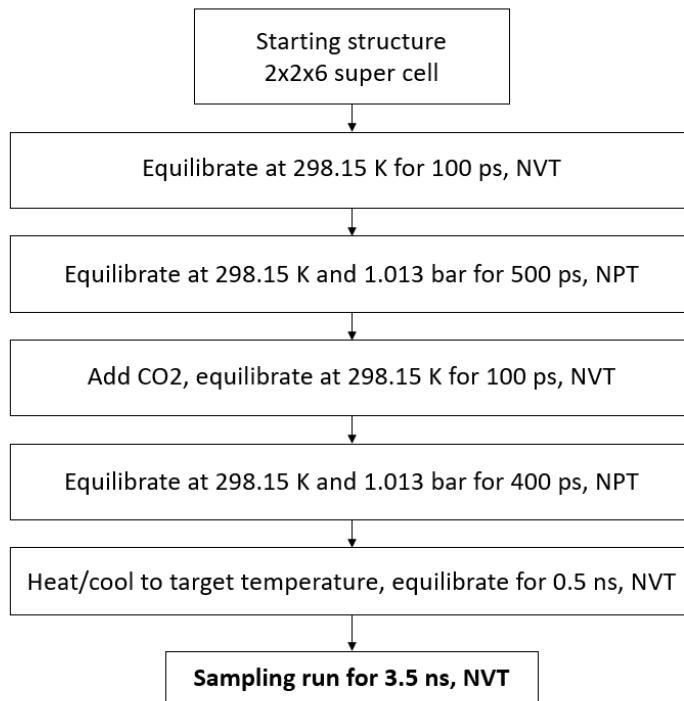


Figure 6: Simulation protocol.

A 2x2x6 supercell was created with Gaussview according to the data provided in the original publication by Alahakoon *et al.*⁴ This starting structure underwent geometry optimization using L-BFGS and was subsequently equilibrated at 10 K in an NVT ensemble. After equilibration of the system at 298.15 K under NVT conditions without CO₂ molecules for 100 ps, it was again allowed to equilibrate at 298.15 K and 1.013 bar under NPT conditions for 500 ps – during these equilibration runs, a time step of 0.5 fs was employed. As soon as equilibration under these circumstances had been achieved, the average C-H bond lengths were determined and CO₂-molecules were added. For each amount of CO₂-molecules per pore (1, 2, 4, 8, 10, 12, 16), the system was once again allowed to equilibrate at 298.15 K for 100 ps under NVT conditions and 500 ps at 1.013 bar and NPT conditions before it was heated or cooled to temperatures between 198.15 K and 348.15 K in 25 K increments, once again equilibrating at each target temperature for 0.5 ns under NVT-conditions. Once stable, 3.5 ps sampling runs at each temperature and each CO₂-concentration were performed, resulting in 49 sampling runs (7 target temperatures and 7 CO₂-concentrations). The reasoning behind this elaborate scheme is that at temperatures >298.15 K, HEX-COF 1's layers started to move in the x-y-plane relative to their neighbors, thus partially blocking the pore canals. Possible explanations will be explored in the discussion in section 5.2. The sampling run trajectories were analyzed in order to calculate the diffusion coefficient. From the diffusion coefficients, activation energies of diffusion were calculated for each temperature and concentration of CO₂, as discussed in section 3.2.7; for results see section 5.3.

4 Space Groups and Symmetry

Symmetry is one of the most important properties in nature, science and technology. It is essential for living organisms as well as inanimate objects such as crystals. Symmetry encompasses all scales: from the macroscopically observable symmetry in many living organisms down to molecular symmetry.⁷⁰ For humans, symmetry is intimately enmeshed with our concept of beauty – symmetric objects are often seen as more beautiful than asymmetric ones. An obvious example of this is the use of symmetry in art and architecture – throughout humankind's history, symmetry can be found in paintings, ornaments, sculptures, buildings and more. Arguably, music may also be seen as an example of symmetry - however, in contrast to symmetry in paintings, it is not spatial but temporal – rhythmic repetition of patterns is a sort of symmetry.^{71,72} However, the symmetry that is of importance for this thesis is that of crystals and molecules.

Lattices and Crystal Structures

A crystal is defined in Borchardt-Ott (2011) as an "*anisotropic, homogeneous body consisting of a three-dimensional periodic ordering of atoms, ions or molecules*".⁷⁰ Further explaining the ordering, he continues: "*a point or space lattice is a three-dimensional periodic arrangement of points, and is a pure mathematical concept*".⁷⁰ Occupying a lattice with an arrangement of atoms, ions or molecules leads to a crystal structure: lattice + basis (the arrangement of atoms/ions/molecules) = crystal structure. A lattice's unit cell is defined by the vectors **a**, **b** and **c** which are along the crystallographic axes a, b and c, respectively. Another option is to define it using six lattice parameters or constants: the lengths of the vectors **a**, **b** and **c** and the angles α , β and γ between them. The unit cell is repeated in all directions of space through translation (only translation!).^{70,73}

Symmetry may also arise from other symmetry operations: in plane lattices, rotation, reflection and inversion are valid symmetry operations. The point, line or plane which remains the same during each symmetry operation is known as the symmetry element. The respective symmetry elements are the rotation axis, the mirror plane, and the inversion center.⁷⁰

For rotations, different orders of rotational symmetry are denoted by n , with $\phi = \frac{360^\circ}{n}$ giving the degrees by which the object has to be rotated in order to achieve congruence again. For example, a hexagon rotated by 60° is indistinguishable from the original (as long as the corners are not labelled) – it is said to have 6-fold symmetry. In space lattices, 5-fold rotational axes are impossible, only 2-, 3-, 4-, and 6-fold rotational axes exist. This is due to the fact that a 5-fold rotation does not result in lattice lines having equal distances from each other. The symbols for the rotational axes are simply n , for example, a 6-fold rotation axis is given the symbol 6. In diagrams, different symbols are available.

The mirror plane with the symbol m (or an m and a bold line in graphical depictions) reflects all points on one side of it to the other side.

The inversion center is present in all space lattices and leads to the generation of a point on the opposite side from a central point – all coordinates of the point are inverted. Since the inversion center is equal to a rotoinversion, wherein a 360° rotation (1-fold) is followed by the inversion, the symbol for the inversion center is $\bar{1}$.

As the mentioned existence of a rotoinversion axis already implies, compound symmetry operations exist, where two operations are performed together, resulting in a new symmetry operation. This is in contrast to simply combining symmetry operations, where the two operations are performed in sequence, retaining the original symmetry operations. Different types of compound symmetry operations exist: rotoinversion (rotation and inversion), rotoreflection (rotation and reflection), screw rotation (rotation and translation) and glide reflection (translation and reflection). With further observation, 2-fold rotation may also be understood as a compound symmetry operation, similar to the inversion.

Lattices can be classified into 14 Bravais lattices, distinguishing themselves by the geometry of the axes in combination with where the lattice points are. If all lattice points are at the corners of the unit cell, it is called a *primitive* unit cell or P. If, on the other hand, lattice points are present at the corners as well as in the center of the unit cell's body, it is called a body-centered unit cell (I). If there are lattice points at the corners as well as in the center of the unit cell's six faces, it is called a face-centered unit cell (F); if the lattice points on the faces are only at two opposite faces, it is called a side-centered unit cell or A, B, C, depending on which face the lattice points are present on. For trigonal primitive unit cells, the letter R is also used.

A set of symmetry operations defines a point group. There exist 32 crystallographic point groups belonging to 7 crystal systems, as shown in table 2. Combining a point group with lattice translation vectors leads to space groups, of which there are 230. 73 of these combine the Bravais lattices with symmetry operations without translation and thus are *symmorphic*, which means that all symmetry operations have one point in common which they leave unchanged.

Table 2: The 32 crystallographic point groups and their corresponding crystal system.

Crystal system	Crystallographic point groups in Hermann-Mauguin notation						
Triclinic	1	$\bar{1}$					
Monoclinic	2	m	2/m				
Orthorhombic	222	mm2	mmm				
Tetragonal	4	$\bar{4}$	4/m	422	4mm	$\bar{4}2m$	4/mmm
Trigonal	3	$\bar{3}$	32	3m	$\bar{3}m$		
Hexagonal	6	$\bar{6}$	6/m	622	6mm	$\bar{6}2m$	6/mmm
Cubic	23	$m\bar{3}$	432	$\bar{4}3m$	$m\bar{3}m$		

Unit cells are classed into six distinct crystal systems based on the crystallographic axes and their orientation to each other. In crystallography, the axis system of $a = b \neq c$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$ is usually separated into two crystal systems: hexagonal and trigonal or rhombohedral. This distinction is made based on the presence of rotational symmetry elements: a 6-fold rotational or rotoinversion axis in the hexagonal and a 3-fold rotational axis in the trigonal system (see table 2). A list of the crystal systems and their respective axes is shown in table 3.⁷⁰

Table 3: Crystal systems and their respective axes.

Lattice System	Unit cell edges	Unit cell angles
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ \neq \beta$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$

5 Results

The HEX-COF 1 system was modelled according to Alahakoon *et al.* as an orthorhombised 2x2x6 supercell with the optimized unit cell parameters $a = 34.12\text{\AA}$, $b = 29.80 \text{\AA}$ and $c = 31.64 \text{\AA}$ and eight pore channels.⁴ The unit cell contained 84 atoms, the supercell 2016. After geometry optimization and equilibration at 10.15 K, the system was heated to room temperature using a Berendsen thermostat and again equilibrated.⁵⁷ The system after equilibration is depicted in figure 7. Following that, the system was filled with varying amounts of CO₂. Equilibration was once again performed for 100 ps at 298.15 K under NVT conditions, followed by 400 ps under NPT conditions with an MC-manostat at 1.013 bar and 298.15 K. The resulting structures were taken as a starting structures for simulations at different temperatures: 198.15 K, 223.15 K, 248.15 K, 273.15 K, 298.15 K, 323.15 K and 348.15 K, where they were once again equilibrated under NVT conditions, this time for 0.5 ns. Each composition was then sampled under NVT conditions for 3.5 ns, resulting in 49 samplings (7 temperatures and 7 compositions).

5.1 Interaction Energies

The interaction energy between the layers of HEX-COF 1 without CO₂-molecules in the pores after geometry optimization was calculated according to equation 55 in order to obtain an indicator of how well the layers adhered to each other. E_i refers to the interaction energy between the layers per cell, E_l denotes the energy of the combined layer system, l is the number of layers investigated and E_1 is the energy of the isolated monolayer system. The division by four is necessary to account for the 2x2x1 supercell.

Results can be seen in table 4. The interaction energy between layers is negative, indicating good adherence between layers.

$$E_i = \frac{E_l - l \cdot E_1}{4} \quad (55)$$

Table 4: Layer interaction energy.

Layers	Energy in keV	Difference to monolayer in eV	E_i in eV	Energy quotient in eV/mol
1	-5459.43	-	-	1
2	-10918.94	-20.49	-10.25	2.00
3	-16378.44	-38.47	-12.82	3.00
4	-21837.94	-56.02	-14.01	4.00
5	-27297.44	-73.48	-14.70	5.00
6	-32756.94	-90.95	-15.16	6.00

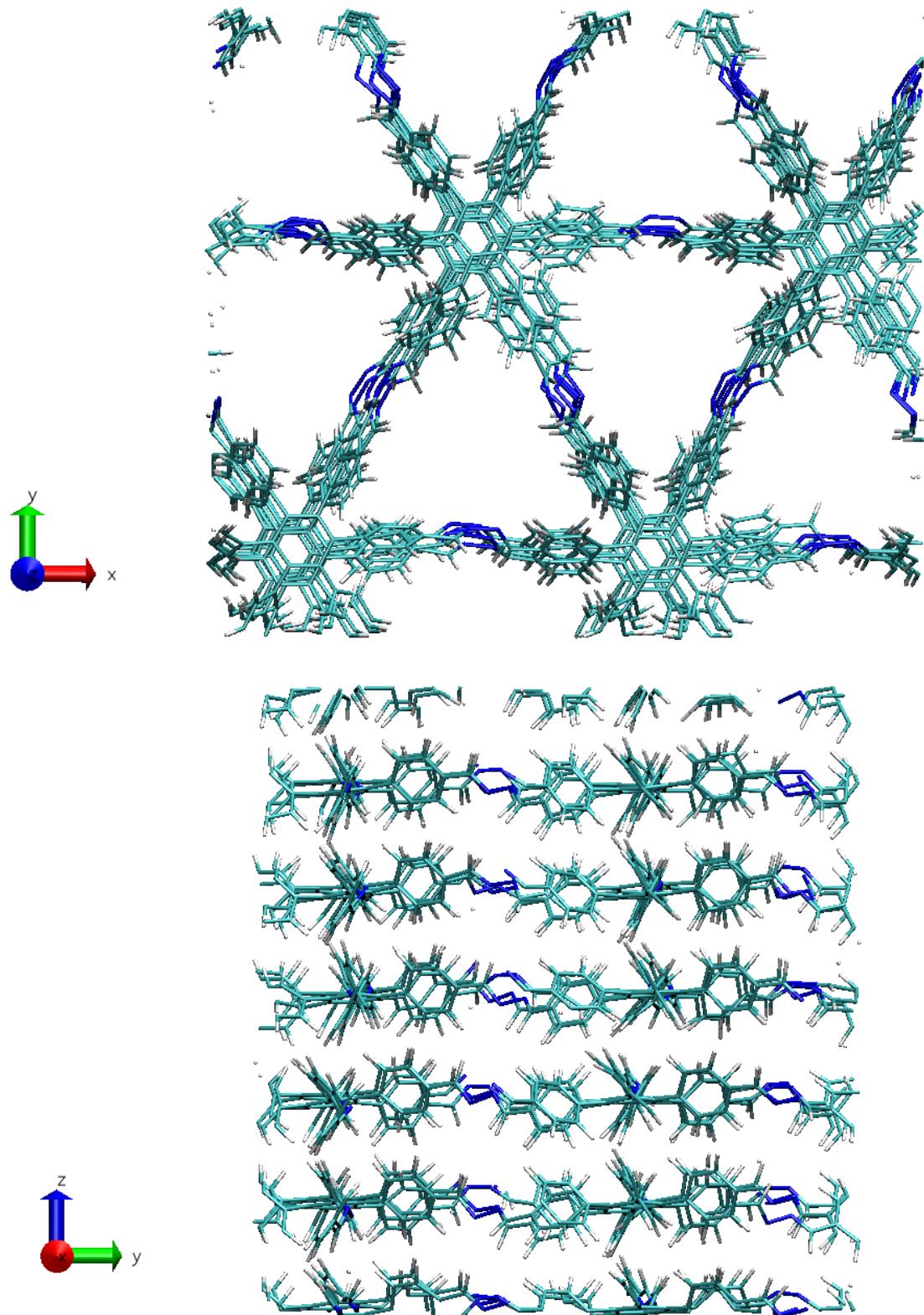


Figure 7: Top: HEX-COF 1 viewed along the z-axis. The different pore channels are clearly visible. Bottom: HEX-COF 1 seen along the x-axis. The 6 individual layers of the simulation system can be directly discerned.

5.2 Structure of HEX-COF 1

The reason why the NPT ensemble was only used during equilibration lies in the fact that the layers of HEX-COF 1 moved relative to each other in the x-y-axis at temperatures higher than 298.15 K, thereby blocking the pore channels and hindering CO₂-movement, see also image 8. By only allowing volume changes during equilibration, this issue was avoided. This may be due to the absence of face-to-face parallel stacking interaction of the phenyl groups and a smaller effective interlayer adhesion.⁴

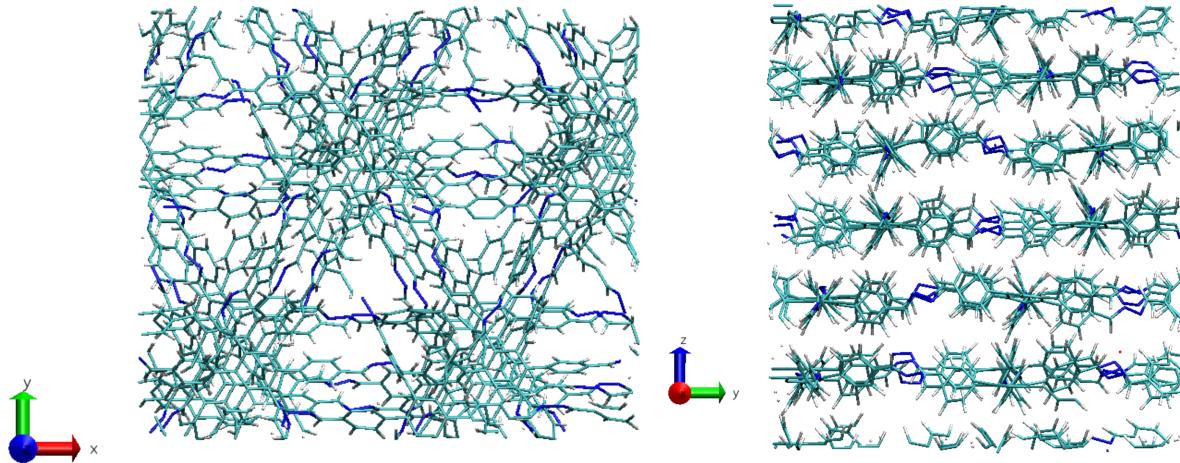


Figure 8: HEX-COF 1 with layers displaced relative to each other at suboptimal simulation parameters. Also compare to figure 7. Left: View along the z-axis. Right: View along the x-axis.

5.3 Diffusion Analysis

The resulting trajectories were used to calculate diffusion coefficients for the CO₂ molecules according to Einstein's relation (see section 3.2.7 for the theoretical background). For HEX-COF 1 and its pore structure, only the diffusion coefficient D_z along the z-axis was calculated, since this geometry significantly impedes diffusion along the x- and y-axis, making it irrelevant compared to D_z , as can be seen in figure 9.

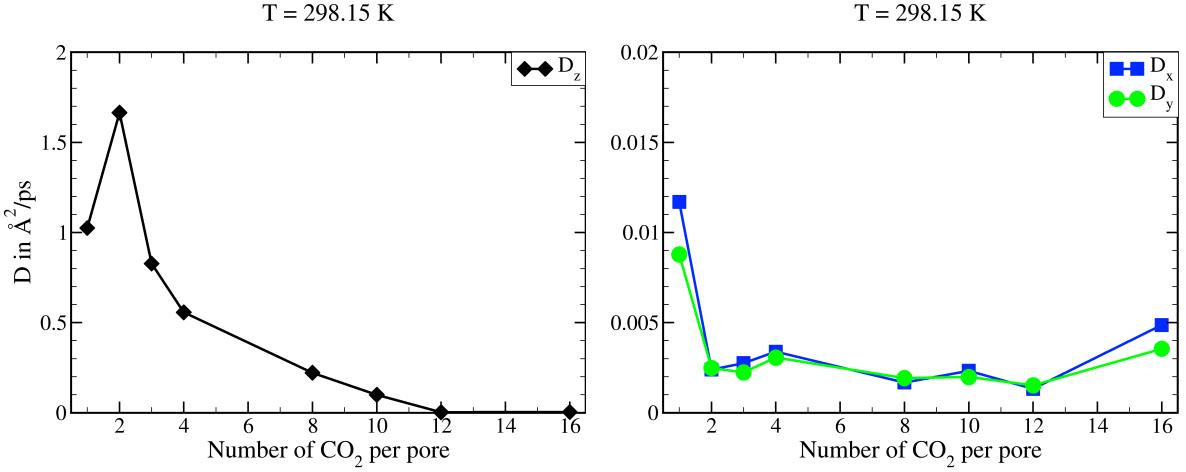


Figure 9: Comparison of diffusion coefficients along x-, y-, and z-axis – note the different scales on the ordinates. Left: Diffusion coefficient D_z along z-axis. Right: Diffusion coefficients D_x and D_y along x- and y-axis.

The Arrhenius plot of $\ln(D_z)$ against the reciprocal of the temperature is shown in figure 10 and follows the expectations of an exponential decrease of diffusivity with sinking temperature and a decrease of diffusivity with increased CO₂-concentration.

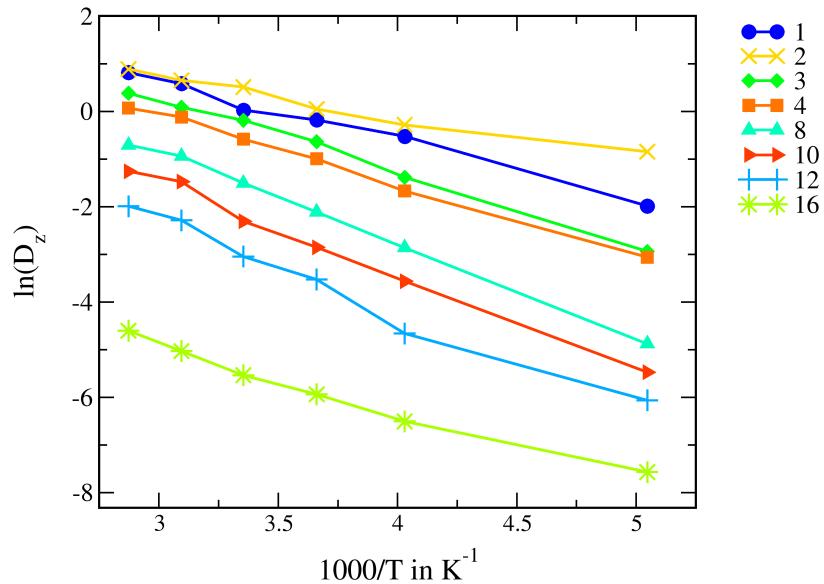


Figure 10: Arrhenius plot of CO₂ diffusion in HEX-COF 1 for different loads of CO₂ per pore.
 D_z in Å² ps⁻¹.

The activation energy of diffusion was calculated according to equation 45 for each CO₂ concentration, see figure 11. Interestingly, it is higher than expected for 1 and 3 CO₂ per pore in comparison to the other concentrations; at 2 and 4 CO₂ molecules per pore, a decrease in activation energy can be observed compared to 1 and 3 CO₂ molecules per pore, respectively. In other studies on the self-diffusivity of gas molecules in porous materials such as CH₄ and H₂ in the zeolite ZK4,^{74,75} or CH₄, CF₄, He, Ne, Ar, Xe and SF₆ in silicalite,⁷⁶ diffusivity increased at loadings greater than one guest molecule per pore.

Also of note is the fact that the activation energy grows until a maximum at 8 CO₂ per pore is reached, after which it decreases again. The increase in activation energy of diffusion implies an interaction that makes adsorption easier after a first CO₂-molecule has been adsorbed - this interaction may be intramolecular charge transfer. Since ANI-2x is trained on DFT data, it may be able to model such effects despite the fact that it did not undergo training for this specific interaction. After this initial increase, the activation energy of diffusion decreases again for loadings larger than 8 CO₂ molecules per pore, or 17.0 wt%. Comparing this value to the experimentally determined maximum uptake capacity of 20 wt%,⁴ it is in good agreement. The decrease in activation energy after this maximum suggests that once the maximum uptake capacity has been reached, all interaction sites near the azine linkers are occupied and the remaining CO₂ molecules are thus able to diffuse more easily.

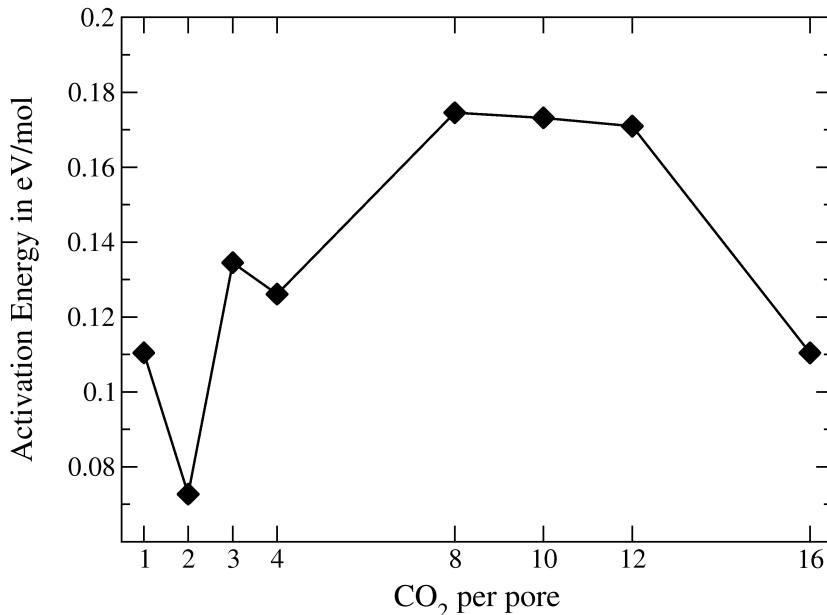


Figure 11: Activation energy of diffusion for each CO₂ concentration.

5.4 Consistency of Results

In order to determine the consistency of the results among the seven runs at each temperature, the diffusion coefficients calculated from each run were compared, as can be seen in figure 12. A pattern observable at all CO_2 loads is that the runs at lower temperatures (198.15 K to 248.15 K) were more consistent, whereas the runs at higher temperatures (273.15 K to 348.15 K) showed more variation. This is untypical and possibly due to the thermal motion of the host material. However, the results are still consistent enough to conclude that ANI-2x is suitable for the description of the diffusion behavior of CO_2 in HEX-COF 1.

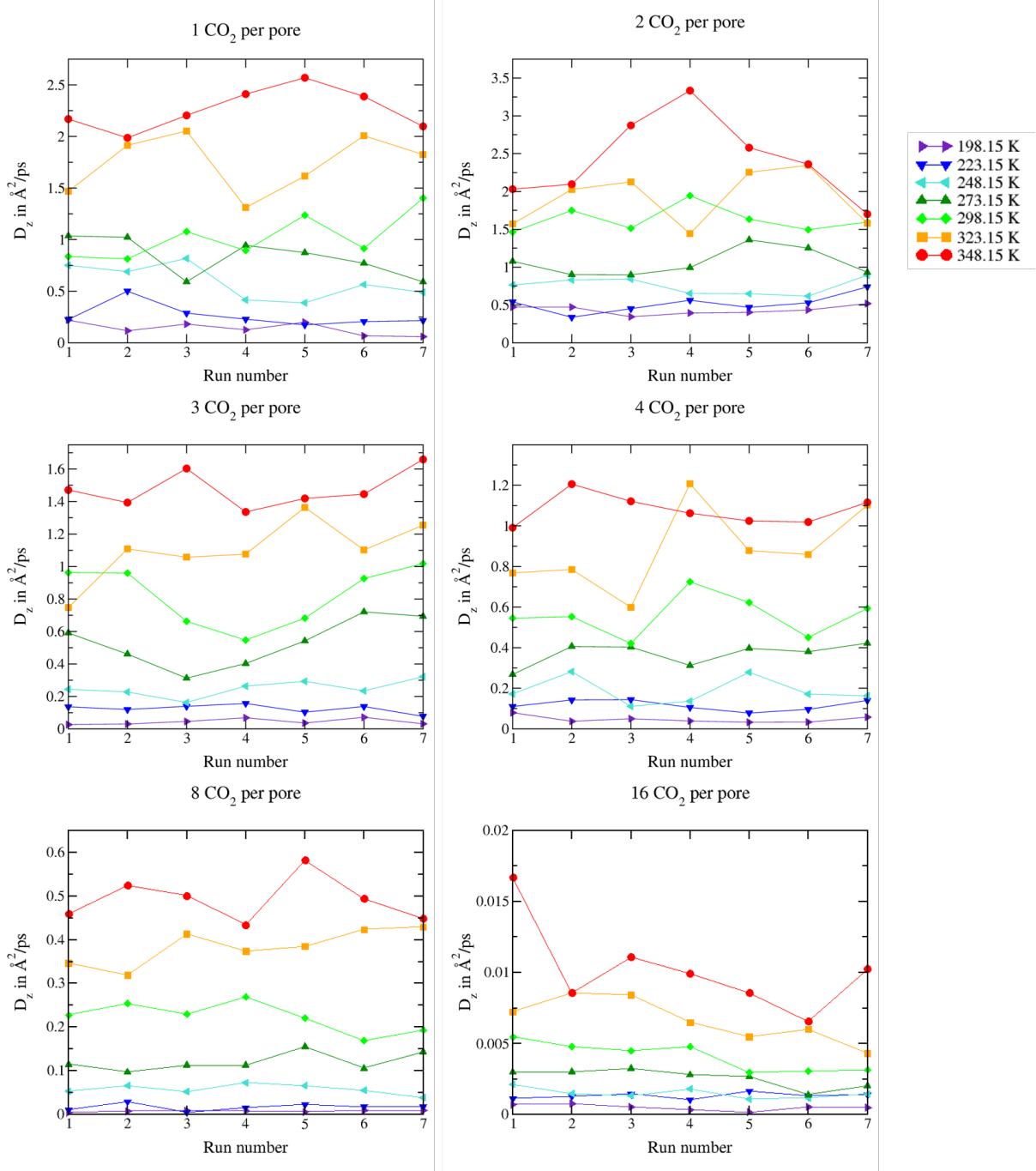


Figure 12: Consistency of results across runs.

6 Conclusion

In this thesis, the covalent organic framework HEX-COF 1 was investigated using molecular dynamics simulations in regard to its capacity as CO₂ storage material, with the ANI-2x neural network potential employed in the description of energy and forces. The research questions were whether ANI-2x is capable of describing pristine HEX-COF 1 and CO₂ adsorption in HEX-COF 1. The results concerning CO₂ uptake capacity and diffusion from the simulations were compared to experimental data in HEX-COF 1's original publication,^{4,5} showing that the results from this thesis were in good agreement with the experimental results. This points to ANI-2x being a suitable tool for such simulations, along with the fact that the results were quite consistent between simulation runs.

An interesting phenomenon in the simulations was that the layers of HEX-COF 1 moved relative to each other in the x-y-plane, thus partially blocking the pore channels. This may be because the face-to-face parallel stacking interaction of the phenyl groups is absent in HEX-COF 1, causing reduced inter-layer adhesion, as also described by Alahakoon *et al.*⁴ Keeping sampling runs in an NVT ensemble with fixed volume prevented this from happening, while still enabling the investigation of CO₂-diffusion in HEX-COF 1.

Diffusion analysis of the trajectories showed that motion along the x- and y-axes was insignificant compared to diffusion along the columnar pore channels of the z-axis. Interestingly, the respective energies in case of 1 and 3 CO₂ molecules per pore were found to be higher than in case of 2 and 4 CO₂ molecules. This is in line with other studies by Tunca and Ford^{74,75} as well as Skouidas and Sholl⁷⁶ who observed a similar effect upon loading of more than 1 CO₂ molecule per pore.

The activation energy of diffusion is increasing until a maximum at 8 CO₂ molecules per pore is reached, after which E_a decreases again. The conclusion this points to is that adsorption may be facilitated after a first guest molecule is adsorbed. This might be caused by intramolecular charge transfer, which ANI-2x might be able to model thanks to its training on DFT data. The lower activation energy of diffusion after 8 CO₂ molecules per pore implies that after all adsorption sites are filled, diffusion becomes easier again for the "surplus" CO₂ molecules. Thus, this maximum in activation energy might point towards the maximum uptake capacity of the CO₂@HEX-COF 1 system.

To summarize, it was shown that molecular dynamics simulations of a covalent organic framework with a neural network potential are perfectly feasible. The CO₂ uptake capacity of the investigated covalent organic framework is in agreement with experimental results. Diffusion analysis was also successful and showed interesting results. The results obtained in this work have been included in a recent publication published in the journal Frontiers in Chemistry.⁷⁷

List of Figures

1	Schematic drawing of the simulation loop. The <i>integration</i> step refers to the integration of the equations of motion.	19
2	Basic velocity Verlet scheme.	20
3	Scheme of the Metropolis Monte Carlo method.	21
4	Periodic boundary conditions in 2D.	24
5	Schematic drawing of the layers of an artificial neural network. Each circle represents a neuron, the arrows correspond to edges with associated weights w_{ij}	26
6	Simulation protocol.	31
7	Top: HEX-COF 1 viewed along the z-axis. The different pore channels are clearly visible. Bottom: HEX-COF 1 seen along the x-axis. The 6 individual layers of the simulation system can be directly discerned.	36
8	HEX-COF 1 with layers displaced relative to each other at suboptimal simulation parameters. Also compare to figure 7. Left: View along the z-axis. Right: View along the x-axis.	37
9	Comparison of diffusion coefficients along x-, y-, and z-axis – note the different scales on the ordinates. Left: Diffusion coefficient D_z along z-axis. Right: Diffusion coefficients D_x and D_y along x- and y-axis.	38
10	Arrhenius plot of CO_2 diffusion in HEX-COF 1 for different loads of CO_2 per pore. D_z in $\text{\AA}^2 \text{ ps}^{-1}$	38
11	Activation energy of diffusion for each CO_2 concentration.	39
12	Consistency of results across runs.	40

List of Tables

1	Comparison between PETM and current anthropogenic climate change.	9
2	The 32 crystallographic point groups and their corresponding crystal system.	34
3	Crystal systems and their respective axes.	34
4	Layer interaction energy.	35

References

- (1) Côté, A. P.; Benin, A. I.; Ockwig, N. W.; O'Keeffe, M.; Matzger, A. J.; Yaghi, O. M. *Science* **2005**, *310*, 1166–1170.
- (2) Geng, K.; He, T.; Liu, R.; Dalapati, S.; Tan, K. T.; Li, Z.; Tao, S.; Gong, Y.; Jiang, Q.; Jiang, D. *Chemical Reviews* **2020**, *120*, 8814–8933.
- (3) IPCC, *Carbon dioxide capture and storage*; Metz, B., Davidson, O., de Coninck, H., Loos, M., Meyer, L., Eds.; Cambridge University Press: Cambridge, 2005, p 442.
- (4) Alahakoon, S. B.; Thompson, C. M.; Nguyen, A. X.; Occhialini, G.; McCandless, G. T.; Smaldone, R. A. *Chemical Communications* **2016**, *52*, 2843–2845.
- (5) Alahakoon, S. B.; Thompson, C. M.; Nguyen, A. X.; Occhialini, G.; McCandless, G. T.; Smaldone, R. A. *Chemical Communications* **2016**, *52*, 1–12.
- (6) Bishop, C. M. *Review of Scientific Instruments* **1994**, *65*, 1803–1832.
- (7) Abiodun, O. I.; Jantan, A.; Omolara, A. E.; Dada, K. V.; Mohamed, N. A. E.; Arshad, H. *Heliyon* **2018**, *4*, DOI: 10.1016/j.heliyon.2018.e00938.
- (8) Sarker, I. H. *Springer Nature Computer Science* **2021**, *2*, 1–21.
- (9) Kocer, E.; Ko, T. W.; Behler, J. *Annual Review of Physical Chemistry* **2022**, *73*, 163–186.
- (10) Devereux, C.; Smith, J. S.; Davis, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2020**, *16*, 4192–4202.
- (11) *Paleoclimate, global change and the future*; Alverson, K. D., Bradley, R. S., Pedersen, T. F., Eds.; Global change - the IGBP series; Springer: Berlin, Heidelberg, New York, 2003.
- (12) *Past climate variability through Europe and Africa*; Battarbee, R. W., Gasse, F., Stickley, C. E., Eds.; Developments in paleoenvironmental research; Springer: Dordrecht, 2004.
- (13) Bender, M. L., *Paleoclimate*; Princeton primers in climate; Princeton University Press: Princeton, New Jersey, 2013.
- (14) Ruddiman, W. F., *Earth's climate: past and future*, 3rd ed.; W.H. Freeman and Company: New York, 2014.
- (15) Arrhenius, S. *Philosophical Magazine and Journal of Science* **1896**, *41*, 237–276.
- (16) Myhre, G.; Highwood, E. J.; Shine, K. P.; Stordal, F. *Geophysical Research Letters* **1998**, *25*, 2715–2718.

- (17) Canadell, J. G.; Monteiro, P. M. S.; Costa, M. H.; da Cunha, L.; Cox, P. M.; Eliseev, A. V.; Henson, S.; Ishii, M.; Jaccard, S.; Koven, C.; Lohila, A.; Patra, P. K.; Piao, S.; Rogelj, J.; Syampungani, S.; Zaehle, S.; Zickfeld, K. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021; Chapter 5, pp 673–816.
- (18) World Meteorological Organisation, *State of the Global Climate 2022*; 1316, 2022, p 57.
- (19) Salby, M. L., *Physics of the Atmosphere and Climate*, 2nd ed.; Cambridge University Press: 2012, pp 1–666.
- (20) Möller, D., *Chemistry of the climate system*; De Gruyter: Berlin, New York, 2010, p 722.
- (21) McIlveen, R., *Fundamentals of weather and climate*, 2nd ed.; Oxford University Press: Oxford, 2010, p 632.
- (22) Forster, P.; Storelvmo, T.; Armour, K.; Collins, W.; Dufresne, J.-L.; Frame, D.; Lunt, D. J.; Mauritsen, T.; Palmer, M. D.; Watanabe, M.; Wild, M.; Zhang, H. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021, pp 923–1054.
- (23) Sherwood, S. C.; Dixit, V.; Salomez, C. *Environmental Research Letters* **2018**, 13, 1–7.
- (24) Siedler, G.; Griffies, S.; Gould, W., *Ocean Circulation and Climate - A 21st Century Perspective*, 2013.
- (25) Williams, R. G.; Follows, M. J., *Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms*; Cambridge University Press: Cambridge, 2011.
- (26) Hurd, C. L.; Beardall, J.; Comeau, S.; Cornwall, C. E.; Havenhand, J. N.; Munday, P. L.; Parker, L. M.; Raven, J. A.; McGraw, C. M. *Marine and Freshwater Research* **2019**, 71, 263–274.
- (27) Zeebe, R. E.; Wolf-Gladrow, D. A., *CO₂ in Seawater: Equilibrium, Kinetics, Isotopes*; Elsevier: Amsterdam, 2001, p 346.
- (28) Hiscock, K. M.; Bense, V. F., *Hydrogeology: principles and practice*, 2nd ed.; Wiley-Blackwell: Chichester, 2014, p 519.
- (29) Speight, J. G., *Lange's Handbook of Chemistry, Seventeenth Edition*, 17th ed.; McGraw Hill: New York, 2016, p 1312.
- (30) Stott, D.; Kennett, P. *Nature* **1991**, 353, 225–229.

- (31) McInerney, F. A.; Wing, S. L. *Annual Review of Earth and Planetary Sciences* **2011**, *39*, 489–516.
- (32) Zachos, J. C.; Dickens, G. R.; Zeebe, R. E. *Nature* **2008**, *451*, 279–283.
- (33) Friedlingstein, P.; O’Sullivan, M.; Jones, M. W.; Andrew, R. M.; Gregor, L.; Hauck, J.; Le Quéré, C.; Luijkx, I. T.; Olsen, A.; Peters, G. P.; Peters, W.; Pongratz, J.; Schwingshackl, C.; Sitch, S.; Canadell, J. G.; Ciais, P.; Jackson, R. B.; Alin, S. R.; Alkama, R.; Arneth, A.; Arora, V. K.; Bates, N. R.; Becker, M.; Bellouin, N.; Bittig, H. C.; Bopp, L.; Chevallier, F.; Chini, L. P.; Cronin, M.; Evans, W.; Falk, S.; Feely, R. A.; Gasser, T.; Gehlen, M.; Gkrizalis, T.; Gloege, L.; Grassi, G.; Gruber, N.; Gürses, Ö.; Harris, I.; Hefner, M.; Houghton, R. A.; Hurt, G. C.; Iida, Y.; Ilyina, T.; Jain, A. K.; Jersild, A.; Kadono, K.; Kato, E.; Kennedy, D.; Klein Goldewijk, K.; Knauer, J.; Korsbakken, J. I.; Landschützer, P.; Lefèvre, N.; Lindsay, K.; Liu, J.; Liu, Z.; Marland, G.; Mayot, N.; McGrath, M. J.; Metzl, N.; Monacci, N. M.; Munro, D. R.; Nakaoka, S.-I.; Niwa, Y.; O’Brien, K.; Ono, T.; Palmer, P. I.; Pan, N.; Pierrot, D.; Pocock, K.; Poulter, B.; Resplandy, L.; Robertson, E.; Rödenbeck, C.; Rodriguez, C.; Rosan, T. M.; Schwinger, J.; Séférian, R.; Shutler, J. D.; Skjelvan, I.; Steinhoff, T.; Sun, Q.; Sutton, A. J.; Sweeney, C.; Takao, S.; Tanhua, T.; Tans, P. P.; Tian, X.; Tian, H.; Tilbrook, B.; Tsujino, H.; Tubiello, F.; van der Werf, G. R.; Walker, A. P.; Wanninkhof, R.; Whitehead, C.; Willstrand Wranne, A.; Wright, R.; Yuan, W.; Yue, C.; Yue, X.; Zaehle, S.; Zeng, J.; Zheng, B. *Earth System Science Data* **2022**, *14*, 4811–4900.
- (34) Pörtner, H.-O.; Roberts, D. C.; Adams, H.; Adelekan, I.; Adler, C.; Adrian, R.; Aldunce, P.; Ali, E.; Begum, R. A.; Friedl, B. B.; Kerr, R. B.; Biesbroek, R.; Birkmann, J.; Bowen, K.; Caretta, M. A.; Carnicer, J.; Castellanos, E.; Cheong, T. S.; Chow, W.; G. Cissé, G. C.; Ibrahim, Z. Z. In *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Technical Summary*; Cambridge University Press: Cambridge, UK and New York, USA, 2022, pp 37–118.
- (35) Parmesan, C.; Morecroft, M. D.; Trisurat, Y.; Adrian, R.; Anshari, G. Z.; Arneth, A.; Gao, Q.; Gonzalez, P.; Harris, R.; Price, J.; Stevens, N.; Talukdarr, G. H. In *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Pörtner, H. O., Roberts, D. C., Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., Rama, B., Eds.; Cambridge University Press: Cambridge, UK and New York, USA, 2022, pp 197–378.
- (36) Arias, P. A.; Bellouin, N.; Coppola, E.; Jones, R. G.; Krinner, G.; Marotzke, J.; Naik, V.; Palmer, M. D.; Plattner, G.-K.; Rogelj, J.; Rojas, M.; Sillmann, J.; Storelvmo, T.; Thorne, P. W.; Trewin, B.; Achuta Rao, K.; Adhikary, B.; Allan, R. P.; Armour, K.; Bala, G.; Barimalala, R.; Berger, S.; Canadell, J. G.; Cassou, C.; Cherchi, A.; Collins, W.; Collins, W. D.; Connors, S. L.; Corti, S.; Cruz, F.; Dentener, F. J.; Dereczynski, C.; Di Luca, A.; Diongue Niang, A.; Doblas-Reyes, F. J.; Dosio, A.; Douville, H.; Engelbrecht, F.; Eyring, V.; Fischer, E.; Forster, P.; Fox-Kemper, B.; Fuglestvedt, J. S.; Fyfe, J. C.; Gillett, N. P.; Goldfarb, L.; Gorodetskaya, I.; Gutierrez, J. M.; Hamdi, R.; Hawkins, E.; Hewitt, H. T.;

Hope, P.; Islam, A. S.; Jones, C.; Kaufman, D. S.; Kopp, R. E.; Kosaka, Y.; Kossin, J.; Krakovska, S.; Lee, J.-Y.; Li, J.; Mauritsen, T.; Maycock, T. K.; Meinshausen, M.; Min, S.-K.; Monteiro, P. M. S.; Ngo-Duc, T.; Otto, F.; Pinto, I.; Pirani, A.; Raghavan, K.; Ranasinghe, R.; Ruane, A. C.; Ruiz, L.; Sallée, J.-B.; Samset, B. H.; Sathyendranath, S.; Seneviratne, S. I.; Sörensson, A. A.; Szopa, S.; Takayabu, I.; Tréguier, A.-M.; van den Hurk, B.; Vautard, R.; von Schuckmann, K.; Zaehle, S.; Zhang, X.; Zickfeld, K. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., Eds.; Cambridge University Press: Cambridge, United Kingdom and New York, NY, USA, 2021, 33144.

- (37) Vishal, V.; Singh, T., *Geologic Carbon Sequestration*, 1st ed.; Vishal, V., Singh, T., Eds.; Springer Cham: 2016, p 338.
- (38) Snæbjörnsdóttir, S. Ó.; Sigfússon, B.; Marieni, C.; Goldberg, D.; Gíslason, S. R.; Oelkers, E. H. *Nature Reviews Earth & Environment* **2020**, *1*, 90–102.
- (39) Shell Canada Energy; Chevron Canada Limited; Canadian Natural Upgrading Limited; 1745844 Alberta Ltd. *Quest Carbon Capture and Storage Project Annual Summary Report*; tech. rep.; Alberta Department of Energy, 2022, p 52.
- (40) Rochelle, G. T. *Science* **2009**, *325*, 1652–1654.
- (41) Kaleeswaran, D.; Vishnoi, P.; Murugavel, R. *Journal of Materials Chemistry C* **2015**, *3*, 7159–7171.
- (42) Wei, H.; Chai, S.; Hu, N.; Yang, Z.; Wei, L.; Wang, L. *Chemical Communications* **2015**, *51*, 12178–12181.
- (43) Jena, H. S.; Krishnaraj, C.; Wang, G.; Leus, K.; Schmidt, J.; Chaoui, N.; Van Der Voort, P. *Chemistry of Materials* **2018**, *30*, 4102–4111.
- (44) Li, Z.; Zhi, Y.; Feng, X.; Ding, X.; Zou, Y.; Liu, X.; Mu, Y. *Chemistry – A European Journal* **2015**, *21*, 12079–12084.
- (45) Cramer, C. J., *Essentials of computational chemistry : theories and models*, 2nd ed.; John Wiley & Sons, Ltd: Chichester, 2004.
- (46) Leach, A. R., *Molecular modelling : principles and applications*, 2nd ed.; Pearson Education: Harlow, 2001.
- (47) Schlick, T., *Molecular modeling and simulation*, 2nd ed.; Antman, S., Marsden, J., Sirovich, L., Eds.; Springer: New York, Dordrecht, Heidelberg, London, 2010.
- (48) Jensen, F., *Introduction to Computational Chemistry*, 3rd; John Wiley & Sons: Chichester, 2017.
- (49) Nocedal, J.; Wright, S. J., *Numerical Optimization*; Springer: New York, 1999, p 636.

- (50) Broyden, C. G. *IMA Journal of Applied Mathematics* **1970**, *6*, 76–90.
- (51) Fletcher, R. *The Computer Journal* **1970**, *13*, 317–322.
- (52) Goldfarb, D. *Mathematics of Computation* **1970**, *24*, 23–26.
- (53) Shanno, D. F. *Mathematics of Computation* **1970**, *24*, 647–656.
- (54) Liu, D. C.; Nocedal, J. *Mathematical Programming* **1989**, *45*, 503–528.
- (55) Clausius, R. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1870**, *40*, 122–127.
- (56) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *The Journal of Chemical Physics* **1953**, *21*, 1087–1092.
- (57) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690.
- (58) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. *Chemical Physics Letters* **2004**, *384*, 288–294.
- (59) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. *Journal of Computational Physics* **1977**, *23*, 327–341.
- (60) Andersen, H. C. *Journal of Computational Physics* **1983**, *52*, 24–34.
- (61) Everett, D. H. *Pure and Applied Chemistry* **1972**, *31*, 617.
- (62) Einstein, A. *Annalen der Physik* **1905**, *322*, 549–560.
- (63) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. *Nature Communications* **2019**, *10*, 1–8.
- (64) Smith, J. S.; Isayev, O.; Roitberg, A. E. *Chemical Science* **2017**, *8*, 3192–3203.
- (65) Behler, J.; Parrinello, M. *Physical Review Letters* **2007**, *98*, 1–4.
- (66) Kingma, D. P.; Ba, J. L. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp 1–15.
- (67) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. *Journal of Chemical Physics* **2018**, *148*, DOI: 10.1063/1.5023802.
- (68) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. *Scientific Data* **2020**, *7*, 1–10.
- (69) Smith, J. S.; Lubbers, N.; Thompson, A. P.; Barros, K. *Simple and efficient algorithms for training machine learning potentials to force data*; tech. rep.; Albuquerque, Los Alamos: Sandia National Laboratory, Los Alamos National Laboratory, 2020, pp 1–7.
- (70) Borchardt-Ott, W., *Crystallography: An Introduction*, 3rd ed.; Springer Berlin Heidelberg: 2011, p 355.
- (71) Weyl, H., *Symmetry*; Princeton University Press: Princeton, New Jersey, 1952, p 168.

- (72) Speiser, A., *Die Theorie der Gruppen von endlicher Ordnung: mit Anwendungen auf algebraische Zahlen und Gleichungen sowie auf die Kristallographie*, 5th ed.; Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften : Mathematische Reihe; Birkhäuser: Basel, 1980, p 271.
- (73) Atkins, P.; de Paula, J., *Atkins' Physical chemistry*, 8th ed.; Oxford University Press: New York, 2006.
- (74) Tunca, C.; Ford, D. M. *The Journal of Chemical Physics* **1999**, *111*, 2751–2760.
- (75) Tunca, C.; Ford, D. M. *The Journal of Physical Chemistry B* **2002**, *106*, 10982–10990.
- (76) Skouidas, A. I.; Sholl, D. S. *The Journal of Physical Chemistry B* **2002**, *106*, 5058–5067.
- (77) Kriesche, B. M.; Kronenberg, L. E.; Purtscher, F. R. S.; Hofer, T. S. *Frontiers in Chemistry* **2023**, *11*, DOI: 10.3389/fchem.2023.1100210.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Magister-/Master-/Diplomarbeit/Dissertation eingereicht.

Datum

Unterschrift