Simulation moléculaire multi-échelles de l'adsorption de fluides dans les matériaux nanoporeux flexibles

Guillaume Fraux

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT

DE L'UNIVERSITÉ PSL

Préparée à Chimie ParisTech

Simulation moleculaire multi-échelles de l'adsorption de fluides dans les matériaux nanoporeux flexibles

Molecular simulation of fluid adsorption in flexible nanoporous materials at multiple scales

Soutenue par

Guillaume FRAUX

Le 25 juin 2019

École doctorale n°388

Chimie Physique et Chimie Analytique de Paris Centre

Spécialité

Chimie Physique

Composition du jury :

Caroline MELLOT-DRAZNIEKS Directrice de Recherche, Collège de France	*Présidente*
Sofía CALERO Professeure, Universidad Pablo de Olavide	*Rapportrice*
Paul FLEURAT-LESSARD Professeur, Université de Bourgogne	*Rapporteur*
Renaud DENOYEL Directeur de Recherche, Aix-Marseille Université	*Examinateur*
Alain FUCHS Professeur, PSL Université	*Examinateur*
François-Xavier COUDERT Chargé de Recherche, Chimie ParisTech	*Directeur de thèse*

Chimie Paris | PSL

ParisTech

REMERCIEMENTS

En premier lieu, je voudrais adresser ici mes plus vifs remerciements à Sofia Calero et Paul Fleurat-Lessart pour avoir accepté d'être les rapporteurs de ce manuscrit, et pour l'intérêt qu'ils ont porté à ce travail. Je remercie aussi Caroline Mellot-Draznieks, Renaud Denoyel, et Alain Fuchs qui m'ont fait l'honneur de faire partie de mon jury de thèse.

J'ai eu la chance d'effectuer ma thèse sous la direction de François-Xavier Coudert, qui m'a beaucoup apporté tant sur le plan scientifique qu'humain. Je souhaite lui exprimer ma gratitude pour tout ce que j'ai pu apprendre à son contact, et pour son immense disponibilité. Il m'a non seulement aidé à acquérir des compétences scientifique variée, allant de la physique statistique à la chimie en passant par la simulation moléculaire et la programmation ; mais aussi une compréhension du monde de la recherche et du travail de chercheur. Je veux aussi remercier Anne Boutin et Alain Fuchs pour le temps qu'ils ont consacré à mon travail, leurs conseils et nos discussions. Je leur dois un intérêt fort pour la thermodynamique et les méthodes Monte Carlo.

Je remercie aussi tous ceux avec qui j'ai eu l'occasion de travailler à Chimie ParisTech, qui ont fait du labo un lieu de travail agréable et chaleureux. Merci à Adrian, Clarisse, Emmanuel, Fabien, Félix, Miguel, Siwar, Srinidhi, Thibaud et Wenke pour toutes nos discussions, toujours enrichissantes. Merci en particulier à Jack et Jean-Mathieu pour leur bonne humeur ; Romain pour son ouverture sur l'industrie qui m'a beaucoup apporté ; Daniela pour ses conseils et ses qualités humaines, au labo comme en conférences ; Johannes et Laura pour nos collaborations et discussions scientifiques.

La thèse ça n'avance pas toujours comme on le voudrait, et il est important d'avoir un lieux où se vider la thèse pour mieux y revenir. Le groupe SGDF de Montreuil m'a permis de passer du temps au vert avec des jeunes formidables. Je voudrais remerciés particulièrement Patricia, Florence, les deux Laurent et Anne-Cécile pour leur soutien et leurs encouragements. Merci aussi à Nathanaël, Julia, Ophélie, Antonin et Elsa : vous avez été de super co-chefs ! De même, de grands mercis au club de musique Folk de l'ENS, et en particulier à Martin, Chloé, Jeremy, Héloïse et Olivier.

Je souhaite aussi remercier ma famille : mes frères et sœurs Béatrice, Paul, David et Matthieu ; ma grand-mère qui m'a apporté une éthique de vie et des racines ; et surtout mes parents. Sans leur soutien et leurs encouragements constants, je ne serais pas là aujourd'hui.

Enfin, je dois énormément à Margot, qui m'a soutenue pendant cette thèse, et a su supporter mon humour douteux et mes couchers tardifs. Merci d'apporter dans ma vie joie et bonne humeur, discussions insensée et profondes, expériences culinaires et voyages au bout du monde !

TABLE OF CONTENTS

GENERAL INTRODUCTION

Nanoporous materials are material presenting nanometric cavities in their structure. These cavities (called *pores*) and the associated high specific surface area are the foundation for crucial industrial applications, in particular in the domains of fluid separation and storage, water purification, and heterogeneous catalysis. Two classes of crystalline nanoporous materials that particularly stand out today are zeolites and metal–organic frameworks. Zeolites are natural and artificial porous aluminosilicate known since 1756 and artificially synthesized since the 1940s. They are currently highly used at the industrial level, in diverse applications including as catalysts in the oil industry and water softeners in laundry detergents. Since the 2000s, a new family of hybrid organic–inorganic crystalline nanoporous materials called metal–organic frameworks have emerged and sparked interest in the scientific community. These new materials are based on metallic clusters, linked together by organic linkers. This composition offers them immense structural and chemical diversity makes them tuneable: it is possible to engineer new materials with specific pore sizes, shapes, and chemistry by using different combinations of linkers and metals. The relatively weak coordination bonds between the metals cations and organic linkers create intrinsic structural flexibility in metal–organic frameworks, which can be either local or extended to the whole material. Some of these metal–organic frameworks — grouped together under the name *"soft porous crystal"* — exhibit large scale structural transformations under external stimuli such as temperature, pressure, adsorption or even exposure to light.

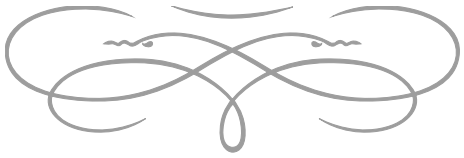
Most if not all applications of nanoporous materials are related to the entry of other chemical species (from the liquid or gas phase) inside the material's pores. When the external fluid is in the gaseous state, the process is called adsorption, while for liquids it is called intrusion. Both adsorption and intrusion have an effect on the physical and chemical properties of the fluid confined in the nanopores. Confined fluids are usually organized more regularly, taking some aspect of a solid phase while remaining mobile. The opposite is also true: the presence of the fluid inside the pores can modify the properties and behavior of the surrounding material. Flexible materials are especially impacted and can undergo adsorption-induced phase transitions, resulting in macroscopic phenomena like gate opening, breathing or negative gas adsorption. The coupling between adsorption or intrusion and changes in the structure of flexible nanoporous materials is difficult to study, because it involves the equilibrium between confined and bulk fluids, as well as equilibrium between different phases of the material.

I have been interested during my PhD in the molecular simulation of fluid adsorption and intrusion in flexible nanoporous materials. Molecular simulation tools can speed up the development of new materials tailored to specific applications by predicting the properties of materials before they are synthesized, lowering the cost of screening thousands of materials. Which property can be studied, and the reliability of the corresponding prediction depend on the techniques used to model the systems of interest. Over the course of my PhD, I used multiple techniques at different time and length scales to look at adsorption and intrusion in flexible porous materials, and the effects on both the confined fluid and the materials. I used macroscopic models based on classical thermodynamics for the study of co-adsorption of gases, umbrella sampling simulations to extract free energy profiles during water intrusion, classical molecular dynamics and Monte Carlo for the intrusion of water and aqueous solutions in both rigid and flexible materials, and *ab initio* molecular dynamics simulations to investigate the molecular organization of gases inside pores.

Unfortunately, current molecular simulations methods only allow looking at the coupling between adsorption and deformations from a single point of view. Molecular dynamics simulations can be used to describe deformations of a supramolecular framework, but are unable to simulate open systems, in particular the adsorption of particles in chemical equilibrium with a reservoir. Monte Carlo simulation, on the other hand, can describe such open systems and thus the adsorption phenomenon, but are very inefficient for the study of collective deformations. I have thus taken interest in (and worked on) the Hybrid Monte Carlo simulation method, which can give us the best of both Monte Carlo and molecular dynamics simulations. I also approached molecular simulation in general and Hybrid Monte Carlo in particular from the point of view of their implementation in molecular simulation software, studying various simulations techniques, their links to statistical thermodynamics and efficient implementation strategy.

This thesis presents my work on the molecular simulation of adsorption and intrusion in flexible nanoporous materials, and is divided into six chapters. I start by presenting nanoporous materials, in particular zeolites, metal–organic frameworks and zeolitic imidazolate frameworks. I introduce the main characteristics at the origin of their huge structural diversity. I also describe the phenomena of adsorption and intrusion, how they relate to one another, and what effects they can have both on nanoporous solids and on confined fluids. In a second chapter, I review the classical thermodynamic models of adsorption and co-adsorption, and then demonstrate how to use classical thermodynamics with the OFAST theory to study co-adsorption in flexible nanoporous materials. My third chapter establishes statistical thermodynamics and the molecular simulations methods we can use to study chemical systems: molecular dynamics and Metropolis Monte Carlo. I also briefly discuss different techniques used to model the interactions between molecules, such as *ab initio* methods and classical force-fields. In chapter four, I present the basis for density functional theory calculations, which I have used to study the adsorption of nitrogen in ZIF-8, showing that the confined fluid undergoes a reorganization linked to framework changes as the pressure increases. The fifth chapter contains classical molecular dynamics studies of the adsorption and intrusion of water in nanoporous solids. I studied intrusion of electrolyte aqueous solutions in ZIF-8, and its impact on

the fluid structure, dynamics, and thermodynamics. I also looked at the confinement of water in aluminosilicate nanotubes, and the corresponding impacts on structure and dynamics. The last chapter describes my work on the implementation of molecular simulation software, and in particular the Hybrid Monte Carlo simulation method, which is particularly suited to the study of adsorption in flexible nanoporous materials. Finally, I present some conclusion from my PhD project, and identify some perspectives and challenges for future work.



1

ADSORPTION AND INTRUSION IN NANOPOROUS MATERIALS

1.1 Nanoporous materials

Porous materials are materials which present a structural porosity, such that their tridimensional structure shows cavities called *pores*. This network of pores can vary in homogeneity and regularity, creating a wide variety of porous materials. They all have in common a high specific surface area, which is the accessible internal surface by grams of the material — up to thousands of square meters by gram of material[1] in the most extreme cases. This very high specific surface area of porous materials is exploited in a number of important industrial applications, especially in the domains of adsorption and catalysis. For example, porous materials are used to separate gases in mixtures as molecular sieves; to filter and remove heavy metals from water; or in heterogeneous catalysis in oil refineries by the cracking process.

The International Union for Pure and Applied Chemistry (IUPAC) recommends to classify porous materials in three groups, depending on the size of the pores[2]. First are the *microporous* solids, where the pores are less than 2 nm in diameter. Then we find the *mesoporous* solids with pore diameter between 2 and 50 nm. Porous solids with pores larger than 50 nm are called *macroporous* solids. Microporous and mesoporous solids are often grouped together as *nanoporous* solids, where the size of pores does not exceed 50 nm.

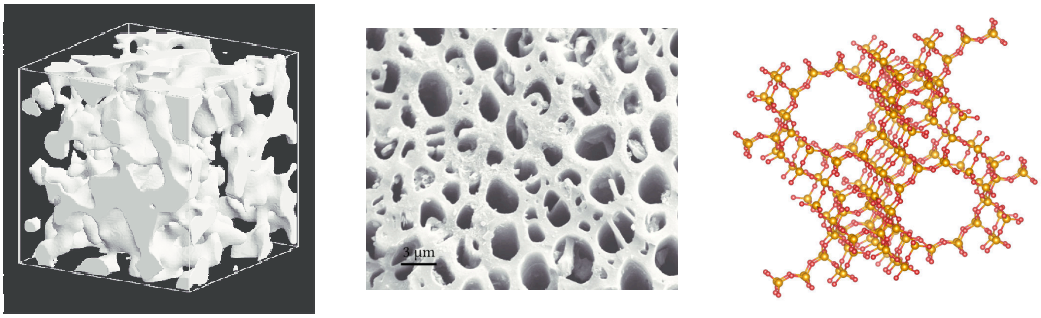


Figure 1.1 – Three examples of porous materials. The left image is a representation of a vycor glass (disordered) (Reprinted with permission from reference [3], copyright (2003) Elsevier); the middle is a scanning electron microscope image of activated carbon (ordered but non crystalline)[4]; and right is the crystalline structure of the zeolite faujasite.

Porous materials are also classified depending on their structural regularity, from highly regular crystalline materials such as zeolites and *Metal–Organic Frameworks* (MOFs) showing a periodic organization of their atoms; to regular porous materials such as clays and carbon nanotubes, where the porosity is well defined but does not present long-distance ordering. Finally, there also exist amorphous porous materials, which have a wide distribution of pore sizes and shapes, and no periodicity. Examples in this latter class are vycor glasses, silica glasses, or aerogels. Three examples of porous solids with different pore size and regularity are presented in figure 1.1.

A final distinction we can make among porous materials is that of their chemical nature. They are usually classified as either organic or inorganic materials. The former contain materials built around carbon atoms, such as carbon nanotubes, microporous carbons, or porous organic polymers. The latter class has historically been the widest one, containing materials such as porous oxides, alumino-silicates, sulfur compounds or alumino-phosphates. In the last few decades, we have seen blooming a new class of materials, with a hybrid inorganic-organic composition. This new family contains materials such as organo-silicic crystals, and metal–organic frameworks, which have been the main subject of study in my PhD.

During my PhD, I studied hybrid inorganic-organic crystalline nanoporous materials, and in particular flexible ones. In the next sections, I will describe zeolites as the conventional example of crystalline porous materials, and MOFs as the relatively new class of hybrid inorganic-organic porous materials. I will also present the *Zeolitic Imidazolate Framework* (ZIF) family of MOFs, which are MOFs with a zeolitic topology.

1.1.1 Zeolites

Zeolites were named by the Swedish mineralogist Axel Fredrik Cronsted in 1756[5] from the Greek ζέω (zéō), meaning “to boil” and λίθος (líthos), meaning “stone”. He observed that when heating a fragment of stilbite mineral to 150°C, the stone started to cover itself with small bubbles, as if the stone was starting to boil — we now understand it as the desorption of water from inside the zeolite’s pores. In the following decades, around twenty additional natural zeolites were discovered. In 1862, French chemist Henri Sainte-Claire Deville, created the first artificial analog to zeolites, but it was only in 1930

that Linus Pauling resolved the structure of zeolites using X-ray diffraction. Today, 234 unique zeolite frameworks have been identified, and we know over 65 naturally occurring zeolite frameworks according to the International Zeolite Association[6].

Structure and composition

Zeolites are porous crystalline alumino-silicates, with a structure built around a regular arrangement of SiO_4 or AlO_4 tetrahedra linked together by their vertices as represented in figure 1.2. The porous network created by these tetrahedra is very different from one zeolite to another, resulting in a wide variety of materials and properties. For examples, the pores can be linear, spherical or in a zigzag disposition; connected together through windows or independents. The International Zeolite Association[6] gives a three letter code such as LTA, FAU or SOD to each experimental crystalline structure. It is however mathematically possible to generate an infinite number of crystalline structures using tetrahedra as a building block, and there are databases of hypothetical zeolites with more than 2 million different frameworks[7].

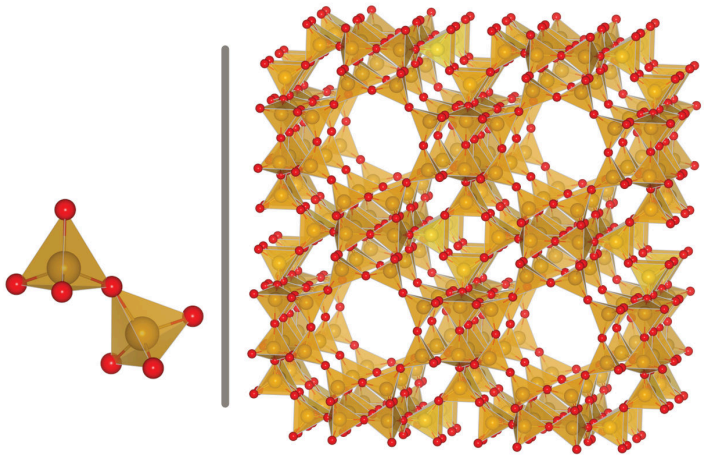


Figure 1.2 – Two SiO_4 tetrahedra on the right, and the structure of zeolite LTA on the left. Si atoms are in yellow and oxygen atoms in red.

While pure-silica zeolites (SiO_2 polymorphs) are neutral, each aluminum atom in the structure introduces a negative charge as silicon is usually in the +IV oxidation state while aluminum is in the +III state. These charges are compensated in the material by a counter ion such as sodium Na^+, potassium K^+, calcium Ca^{2+} or barium Ba^{2+}. Each zeolite topology thus defines a family of materials with varying chemical composition. The general chemical formula for a zeolite is $M_{x/m}\,Al_x\,Si_{1-x}\,O_2$, possibly with several different metal cation M. The Si/Al ratio can vary from 1 for structures with the most aluminum atoms to infinity for pure silica zeolites, also called *zeosil*. Each new aluminum atom replacing a silicon atom is accompanied by a cation to ensure the overall charge neutrality. The presence of these cations contributes to the remarkable adsorption and catalysis properties of zeolites. Zeolites are used in industrial processes for ions exchange through their extra cations; catalysis through their high specific surface area and acido-basic properties; molecular sieves through the multiples pores sizes and tunable properties.

1.1.2 Metal–Organic Frameworks

Metal–organic frameworks are a more recent class of crystalline nanoporous materials. The first works on these materials were made in the 90s by Richard Robson and collaborators[8, 9], but the systematic study of these materials only started in the 2000s. In 1999, the group of Omar Yaghi created MOF-5[10], a MOF with large permanent porosity (15 Å in diameter for the largest cavity) and chemical tunability, sparking interest in the international research community for these materials. Since then, the field of MOF has been developing exponentially, both in terms of the number of publications and in terms of number of structures reported in the literature. Figures 1.3 shows the growth of the number of structures in the Cambridge Structural Database, with a doubling time of 3.9 years for three-dimensional MOFs.

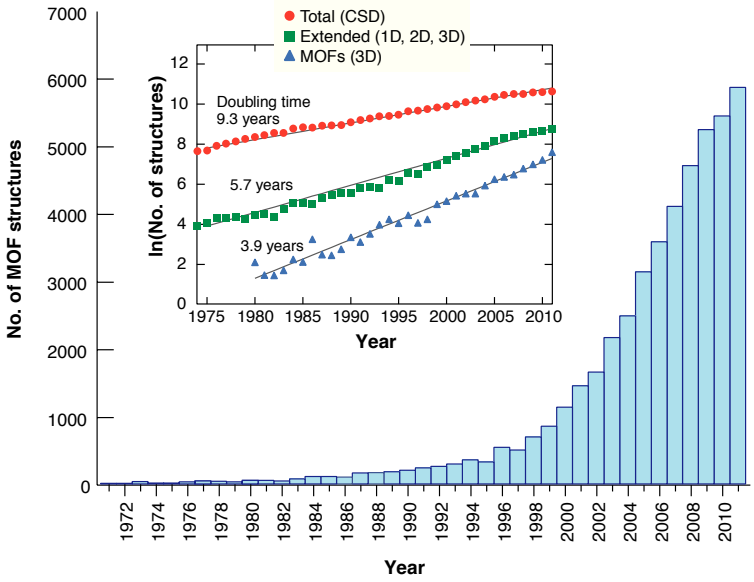


Figure 1.3 – Number of MOF structures in the Cambridge Structural Database showing an exponential growth. Reprinted with permission from AAAS, reference [11].

Materials in the MOF family are built around metallic centers, connected by organic linkers, assembled into nano- or mesoporous crystalline structures (see figure 1.4). Compared to zeolites, they can be synthesized in a solvent at lower temperatures (from ambient temperature to 200°C), by mixing metallic salts and organic linkers. Due to the presence of these organic linkers, MOFs have smaller thermal stability (up to 400°C) compared to the inorganic zeolites that can withstand up to 1000°C. Nevertheless, this reduced stability is made up for by the incredible adaptability of the MOFs. By combining the variety of linker-metal coordination chemistry with the adaptability of organic chemistry, they offer a huge number of different structures, and can be tuned for specific applications by changing linkers and/or metal centers. Figure 1.5 presents an example of how the shape and size of the pores can be adapted by changing the linker used. Furthermore, compared to zeolites which present a rigid crystalline structure, some MOFs grouped together under the name *Soft Porous Crystals*[12] have extraordinary structural flexibility in response to external stimuli[13–15].

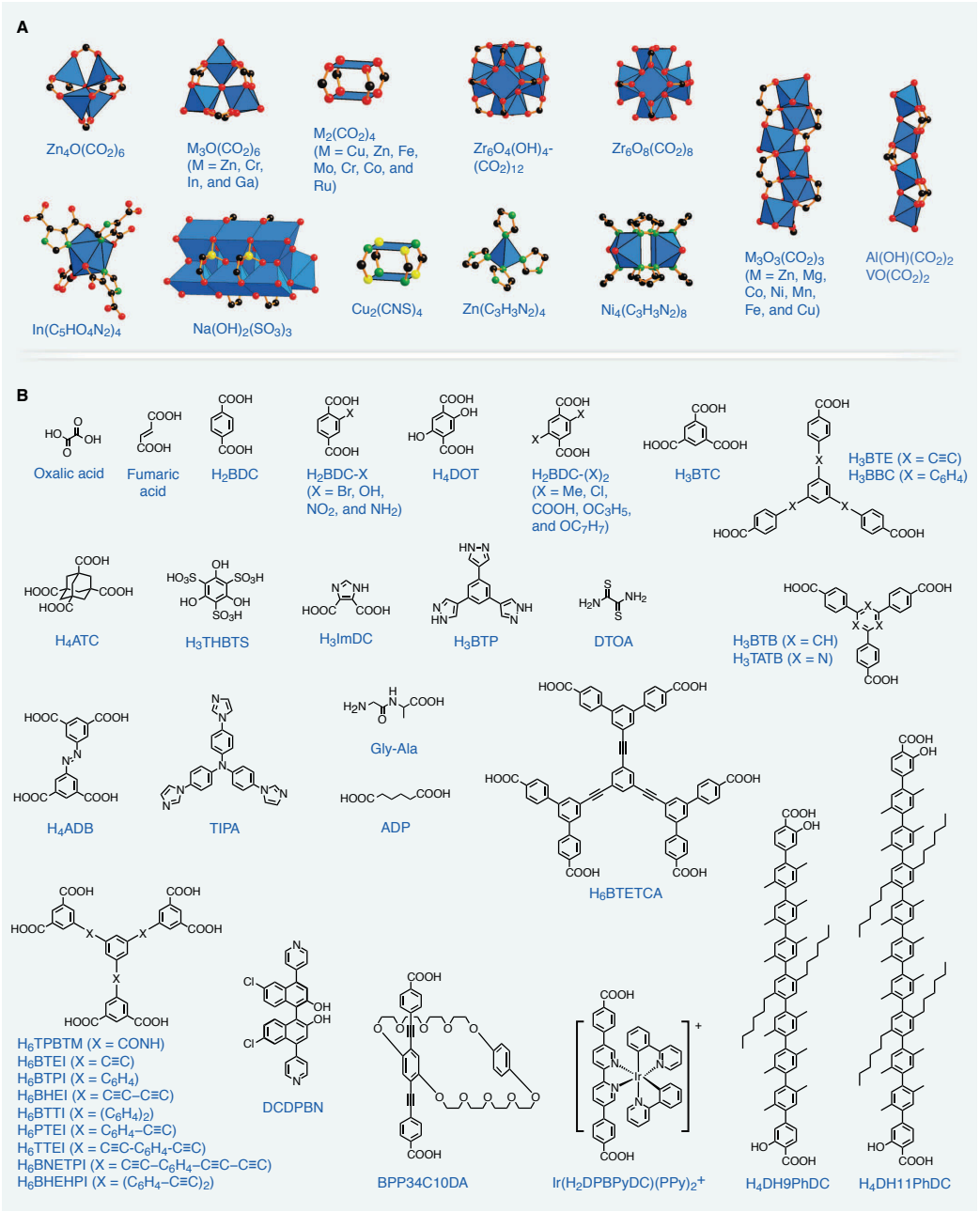


Figure 1.4 – Examples of inorganic clusters (A) and organic linkers (B) used in MOF synthesis. Reprinted with permission from AAAS, reference [11].

High diversity of structures

One of the main advantage of MOFs over other porous materials such as zeolites or activated carbons is the diversity of structures that can be built from the different available inorganic clusters and organic linkers (see figure 1.4). With all the versatility provided by the coordination and organic chemistry, one is only limited by the thermodynamic and chemical stability of the structures obtained, and by the existence of a synthetic route to generate the porous polymorphs of interest, when another phase might be more stable. This diversity of structure gave rise to a *design to applications* approach, where one aims at producing the best structure for a given functionality. For example, researchers have reported MOFs with the highest methane[16] or hydrogen[17] uptakes until now.

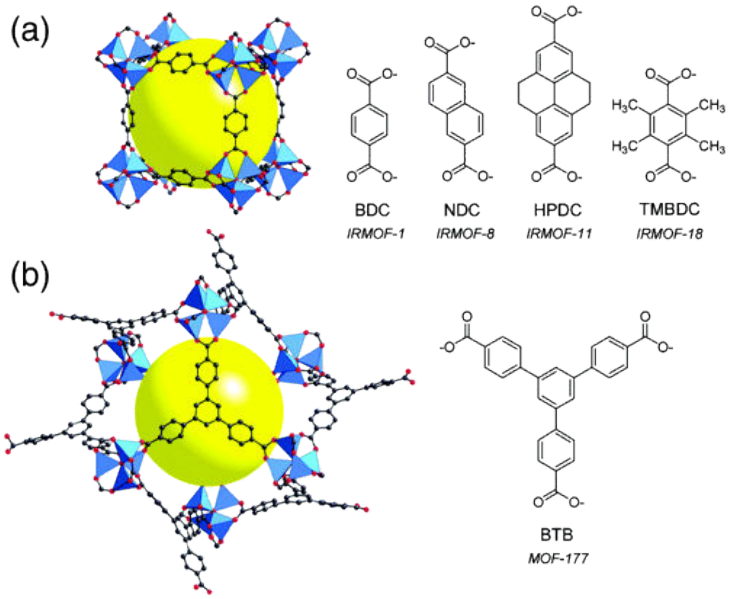


Figure 1.5 – Examples of two MOFs built with a zinc oxide cluster with the same coordination geometry. (a) Structure of MOF-5, made with a linear linker. (b) Structure of MOF-177, made with a trigonal linker. Reprinted with permission from reference [18], copyright (2004) American Chemical Society.

They are two different approaches used to create new MOF structures. The first one is to use linkers and metallic centers with different connectivity, such as the ones shown in figure 1.5. Most of the linkers used in MOF present either carboxylate or nitrogen organic functions that bind the metals. Using a linker with a different number of these functions or a metal center with different oxidation degree will create a new structure. The second way to create a new structure is by changing the chemical structure of the linker, while keeping the number and positions of binding functions intact. The resulting structures are said to be *isoreticular*, *i.e.* they share the same topology or net.

1.1.3 Zeolitic Imidazolate Frameworks

Zeolitic Imidazolate Framework or ZIFs are a family of MOFs built around tetravalent M^{2+} metal centers such as Fe, Co, Zn, Cd or Cu; linked together by imidazolate or functionalized imidazolate linkers. They present the same possible topologies as zeolites, with metal(imidazolate)$_2$ building blocks taking the role of SiO_2 — as illustrated in figure 1.6. The first ZIFs (ZIF-1 to ZIF-12) were synthesized in 2006[19], and found to be more

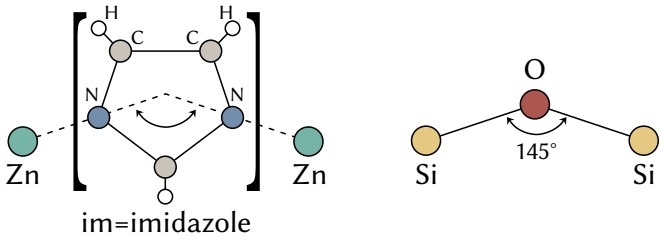


Figure 1.6 – Illustration of the analogy between ZIFs and zeolite coordination.

resistant to water and heat that typical MOFs, which made them interesting for commercial applications. This higher stability is due to the relatively strong metal–imidazolate coordination bond.

ZIF-8 is the member of the ZIF family which I worked on specifically during my PhD. It uses 2-methylimidazolate (mim) as a linker, and Zn^{2+} as its metallic centers. ZIF-8 has formula $Zn(mim)_2$ and adopts the sodalite (**sod**) topology. In this topology, large quasi-spherical pores corresponding to the sodalite cages are connected by windows formed by 6 and 4 zinc atoms. In ZIF-8, the 4 members windows are too small for any molecules to go through, and all of the connectivity of the pore space happens through the 6 members windows. I also studied materials derived from ZIF-8, where the methyl group of the linker is replaced by halogens such as chlorine or bromide. The corresponding studies are presented in sections 4.2 and 5.1.

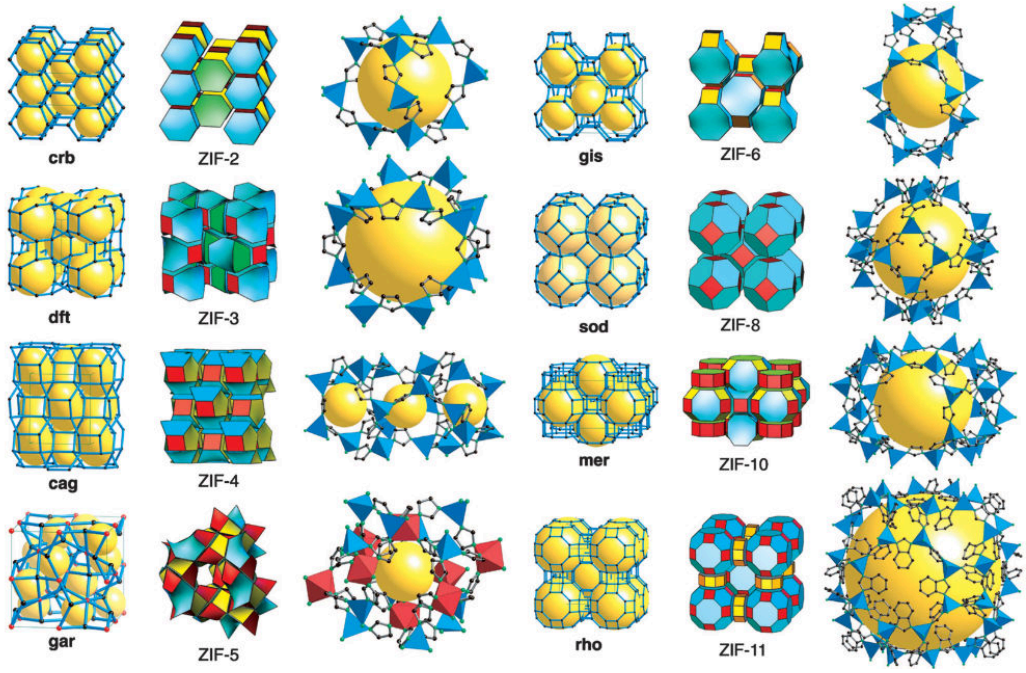


Figure 1.7 – Examples of ZIF structures. From left to right are the crystalline structure of a zeolite with the same topology, the crystalline structure of the ZIF and the biggest sphere inside the cages. Reprinted from reference [19], copyright (2006) The National Academy of Sciences of the USA.

1.1.4 Structural flexibility

In 1998, Susumu Kitagawa proposed a classification of MOFs in three categories[12] depending on their behavior with respect to molecule adsorption and desorption. The first category of materials has a non-permanent porosity, in the sense that if we remove the solvent molecules from the structure after the synthesis, the material collapses. The second category regroups materials that have a strong enough crystalline structure and keep the same porosity during adsorption and desorption. They are essentially rigid, and generally present a very good mechanical and thermal stability. The last category contains materials that retain porosity when the synthesis solvent is removed, but may deform and change shape during adsorption. Materials in this last category are called *Soft Porous Crystal*, and have a flexible, dynamic structure that can change under external stimuli such as mechanical pressure, adsorption, temperature, or even exposure to light[13, 15].

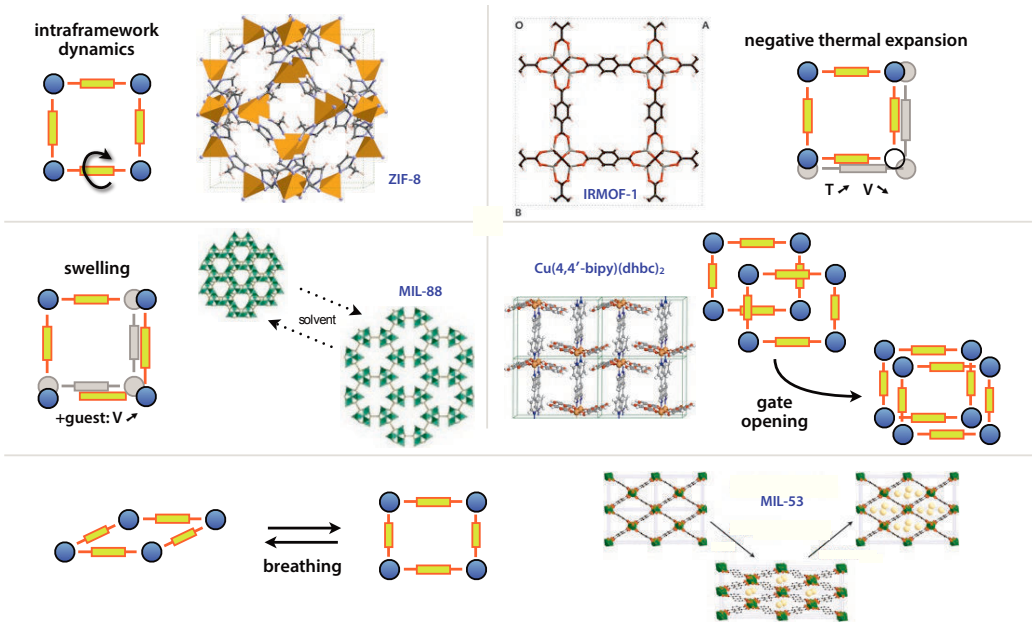


Figure 1.8 – Illustration of the main flexibility modes of MOFs: linkers rotation, thermal expansion, swelling, gate opening and breathing. Reprinted with permission from reference [20], copyright (2011) Wiley.

This flexibility is inherent to the hybrid organic-inorganic nature of MOFs. Indeed, their structure is based on both strong covalent bonds inside the organic linkers, and weaker bonds such as coordination bonds, π-stacking of linkers, hydrogen bonds, *etc.* These weaker bonds can vary locally in length or orientation, inducing large scale deformations of the materials. Possible deformation modes for MOFs are represented in figure 1.8. All MOFs are flexible to some extent through local deformations, such as linkers rotation, this type of flexibility happening without any global framework deformations. The other types of flexibility only occur in specific materials. For example, we observe in some materials a volume contraction as they heat up, *i.e.* a called *negative thermal expansion*[21].

In soft porous crystal, we additionally observe large scale deformations of the structure. For some materials, the whole structure will swell in presence of a solvent. For example, the volume of MOFs of the MIL-88 family can grow by more than 200% of the initial

volume in presence of lutidine[22]. Other MOFs present multiple stable phases, and can switch from one phase to another under external stimuli. The *gate opening* phenomenon is one of such transitions, between an initially non-porous structure to a more open and more porous structure through the movement of linkers or the displacement of a sub-network. Materials from the MIL-53 family present two transitions, from an open to a close-pore phase, and then back to an open-pore phase[23] under continuous gas loading increase, creating a *breathing*-like behavior.

1.1.5 Industrial applications

Zeolites enjoy a wide range of applications, from air separation using pressure swing adsorption[24], catalysis in oil refining[25], wastewater cleaning and heavy metal removal [26], capture of radioactive particles[27], molecular sieve[28], and even laundry detergent [29]. Each year, almost 3 million tonnes of zeolites are produced or extracted for these applications. MOFs are still relatively new, and don't enjoy as many commercial and industrial applications as zeolites yet. One of the obstacles to wider usage of MOFs is their price, as they can be a hundred times more expensive than zeolites. In the following, I will give examples of potential applications of MOFs in various fields.

Gas separation, purification, and storage

Gas separation and purification processes are at the root of the chemical industry, either generating small carbon chain molecules from oil and natural gas, separating oxygen from nitrogen in the air, or CO_2 from H_2 during ammonia production. Current best estimates indicate that gas separation accounts for 10-15% of the energy consumed globally[30]. Furthermore, when faced with climate change resulting from continually-increasing anthropogenic CO_2 emissions, optimization of storage and separation processes is increasingly critical. There are nowadays multiple techniques used for gas separation using cryogenic separation, liquid phase adsorption or gas phase adsorption. Few techniques based on gas phase adsorption are illustrated in figure 1.9.

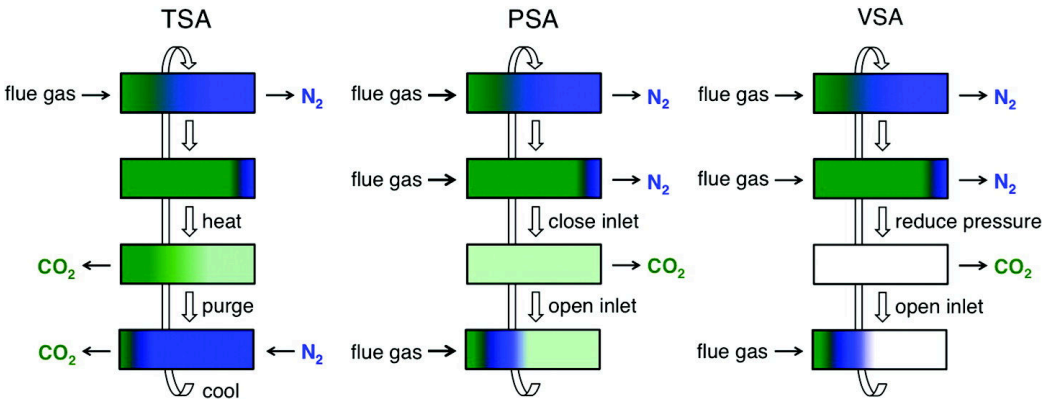


Figure 1.9 – Schematic diagrams of temperature swing adsorption (TSA), pressure swing adsorption (PSA), and vacuum swing adsorption (VSA) processes for regenerating solid adsorbent in a fixed-bed column. Reprinted with permission from reference [31], copyright (2011) American Chemical Society.

As porous materials with huge specific surface area, MOFs look like ideal candidates for any application involving gas adsorption. Separating gases by adsorption requires different interactions between the porous material and the different adsorbates. These

interactions can be tuned in MOFs by changing the linkers and modifying the pores size, shape and chemical properties. During gas purification, one chemical species is present in majority (the gas to purify) and we try to remove the remaining impurities. For example, HKUST-1 (a MOF with formula $Cu_3(btc)_2$) turned out to be very efficient to remove sulfur compounds such as tetrahydrothiphene and thiophene from natural gas, thanks to the formation of a Cu – S bond[32]. It was able to adsorb 70 g of tetrahydrothiphene by liter of material, which is noticeably better than the standard materials used in the industry (less than 10 g/L for activated carbons).

Considering gas storage, an emerging application is the storage of hydrogen gas for energy generation and automobile transportation. Historically, hydrogen has been stored through liquefaction or compression in bottles and tanks, but these approaches have security and cost issues. MOFs are interesting as safer media for gas storage because they are able to pack the hydrogen molecules densely at a lower pressure. For example, $Ni_2(m\text{–dobdc})$ is one of the current best candidate for H_2 storage, able to adsorb 11 g/L of H_2 at ambient temperature, and up to 23 g/L via a temperature swing from -75°C to 25°C[33].

Heterogeneous catalysis

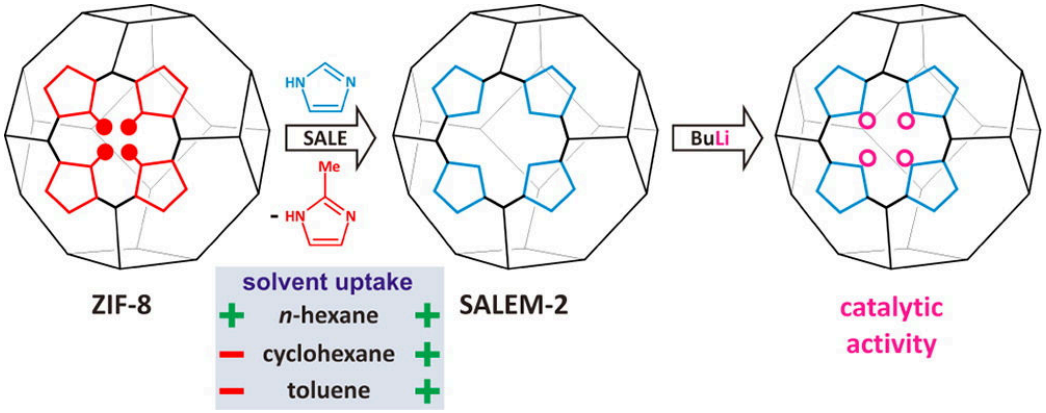


Figure 1.10 – Post-synthetic modifications of ZIF-8 to synthesize SALEM-2 and obtain catalytic activity. Reprinted with permission from reference [34], copyright (2012) American Chemical Society.

Heterogeneous catalysis is used in numerous industrial processes. The selectivity of these catalytic processes is often based on the shape and size of the reacting species, hence the interest in using catalysts with a regular and uniform porosity. Porous crystalline materials are especially interesting in this regard. Moreover, since catalysts can be re-used, and only small quantities of them are needed, the advantages of MOFs can offset their cost. The possibilities to create chiral MOFs using enantiomerically pure linkers, or conformational chirality[35] also opens the way to asymmetric catalysis.

A first strategy for using MOFs as catalysts involves the metallic centers, which can have a catalytic activity of their own. For example, MIL-100(Fe) and MIL-100(Cr) have an interesting activity for the Friedel–Craft benzylation reaction, used for the production of linear alkyl-benzene; surpassing the activity of the HBEA and HY zeolites traditionally used[36]. HKUST-1 MOF can be activated after the synthesis by removing an apical water

molecule, creating a very reactive Lewis acid site which can be used for cyanosilylation reactions[37].

Another approach for adding catalytic activity to MOFs entails adding functional groups to the linkers with the desired activity. For example, SALEM-2 can be activated with *n*-butyllithium and exhibits a strong Brønsted base catalytic activity[34] (see figure 1.10). Around the same idea, the amino functions of IRMOF-3 and amino-MIL-53 make them suitable as basic catalysts for the Knoevenagel condensation of ethyl-cyanoacetate and ethyl-acetoacetate with benzaldehyde[38].

Chemical sensors

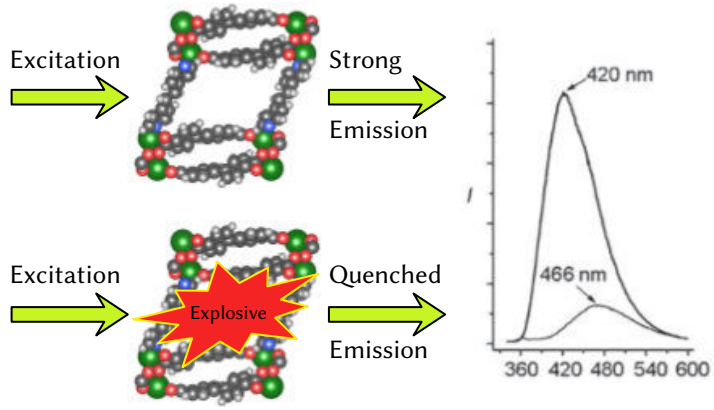


Figure 1.11 – Schematic representation of optical properties modifications of $Zn_2(bpdc)_2(bpee)$ MOF. Image adapted with permission from reference [39], copyright (2009) Wiley.

Some MOFs have luminescent properties, linked either to the presence of aromatic or functionalized linkers or a metal cation from the lanthanide group. These luminescent properties, combined with the adsorption capabilities of MOFs open the possibility of using MOFs as chemical sensors. The structural and/or electronic transition of a MOF upon adsorption will modify the light emission properties of the material, allowing to detect specific gaseous compounds. This is have been developped in particular to detect explosive materials at low concentrations. For example, the $Zn_2(bpdc)_2(bpee)$ MOF is able to detect multiple nitro-substituted molecules found in explosives such as 2,4-dinitrotoluene or 2,3-dimethyl-2,3-dinitrobutane in around 10 s[39] (see figure 1.11). Another example is the $[Zn_2(oba)_2(bpy)]_3$ MOF, also able to detect explosive and aromatic molecules through a fluorescence quenching phenomenon[40].

1.2 Adsorption and intrusion

Adsorption is an interface phenomenon, where molecules or atoms are depositing and creating a thin film at the surface of a material. Because they have a high specific surface area, the adsorption in porous materials is several orders of magnitude higher than on the external surface of bulk materials. Adsorption is used as a characterization technique for porous media, and is also the basis of their most prevalent applications.

1.2.1 Adsorption of gases

It is possible to study adsorption both with experimental and theoretical approaches. The central tool for the study of adsorption is the single component adsorption isotherm. Considering a pure gas adsorbing in a porous matrix, the adsorption isotherm records the amount of adsorbed gas (also called *loading*) at a fixed temperature as a function of the pressure P, or the more commonly, the pressure relative to the vapor pressure of the gas P/P_0. Experimentally, these adsorption isotherms are obtained with gravimetric, volumetric or chromatographic methods[41, 42].

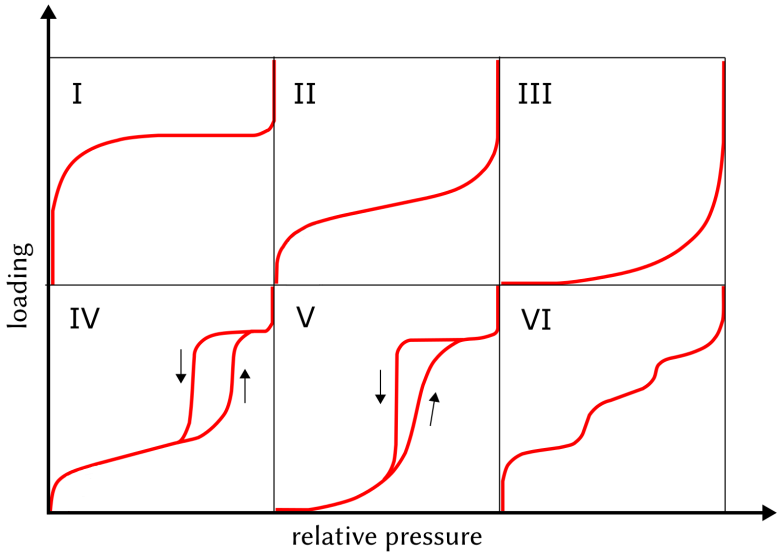


Figure 1.12 – IUPAC classification of isotherms.

The IUPAC classifies adsorption isotherms depending on their shape, as represented in figure 1.12[43]. Only the isotherms of type I, IV and V can occur in nanoporous solids. Type I isotherms are concave, reversible and the adsorbed quantity of matter goes to a finite value as the pressure increases. It is the most common isotherm type, and express the prevalence of gas–solid interactions over gas–gas interactions in a material containing a single kind of equivalent adsorption sites.

Type IV isotherms also present a maximal loading, and one or two steps with sometimes a hysteresis loop between the adsorption and desorption branches. These steps are usually attributed to a modification in the structure of the adsorbing solid[29]. Type V isotherm shows an inflection point: at low pressure the adsorption is very low, mainly due to small gas–solid interactions. For higher pressures the adsorption becomes stronger thanks to interactions between the gas molecule themselves. In flexible porous materials, type IV isotherms would be associated with *breathing* materials, and type V isotherms with *gate*

opening. Finally, isotherms types II, III and VI do not present a maximal loading, and are usually observed in solids with multiple pore sizes and a continuous transition from mono-layer adsorption to multi-layers adsorption and capillary condensation.

It is possible to study and predict adsorption isotherms using theoretical methods such as thermodynamic models and molecular simulations methods such as Grand Canonical Monte Carlo. I will present with more details the methods I used during my PhD to study adsorption in the corresponding chapters: chapter 2 for macroscopic thermodynamics modeling; and chapter 3 for molecular simulations.

1.2.2 Intrusion of liquids

The intrusion of liquids, and in particular mercury, has been used for a long time to characterize porous materials having pore widths in the macropore range of 50 nm to 500 µm[44]. Intrusion can be seen as adsorption of fluids above the vapor pressure ($P > P_0$), *i.e.* when the fluid is in its liquid state. A sketch of an intrusion–extrusion porosimeter is given in figure 1.13. In this device, a non-wetting fluid is pressurized and will enter the porous network at a pressure linked to the pore width h by the Washburn equation[45]:

$$P = \frac{2\gamma \cos\theta}{h}, \tag{1.1}$$

where γ is the surface tension of the liquid/gas interface, and θ the angle of contact along the triple line solid/liquid/gas.

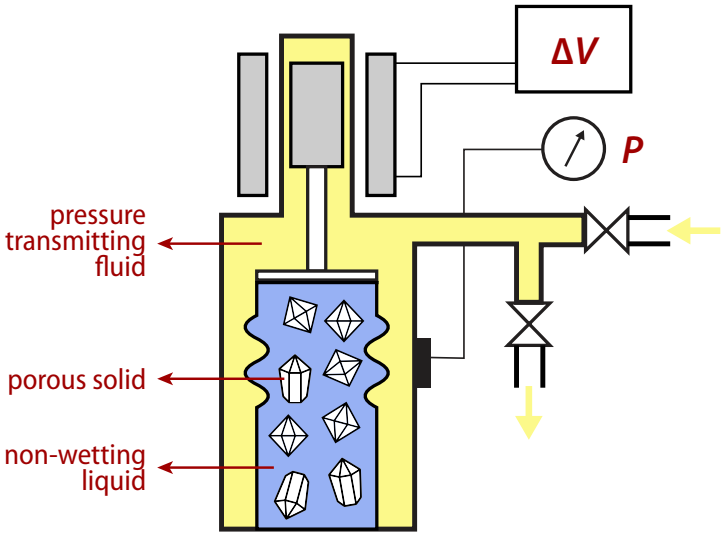


Figure 1.13 – Sketch of an intrusion–extrusion porosimeter. A sample of porous solid powder is placed in a non-wetting liquid (in blue) and pumps control the pressure of a pressure-transmitting fluid (in yellow). A manometer records the pressure P *during the experiment, and a transducer records changes in volume of the system.*

In the 2000s, water was used as non-wetting fluid in an adapted porosimetry device to characterize hydrophobic nanoporous materials such as silica gels[46] and siliceous zeolites[47, 48] or micelle-templated silica[49]. More recently, forced wetting of electrolyte solutions was studied, disclosing very interesting giant osmotic pressure effects[50, 51]. Water can only be used as a non-wetting fluid in *hydrophobic solid*, where the potential energy interaction between the water molecules and the confining solid wall is much weaker than the water molecules' mutual interaction. This use of water as a non-wetting fluids opened a new field of applications for the intrusion in nanoporous solids based on mechanical energy storage and dissipation[45, 47]. Interested readers can consult the review I helped to write on the subject, published in *Chemical Society Reviews* (2017)[45], for more details.

The general idea of storing energy by forced intrusion of a non-wetting fluids in porous media was first explored by Fadeev et al. some years ago[46]. In 2001, Eroshenko et al.[47] reported a stepwise intrusion-extrusion isotherm of water in two different hydrophobic zeolites, namely zeolite β and silicalite-1 at pressures around 60 and 80 MPa respectively, *i.e.* notably above the water saturation pressure of 3500 Pa at room temperature.

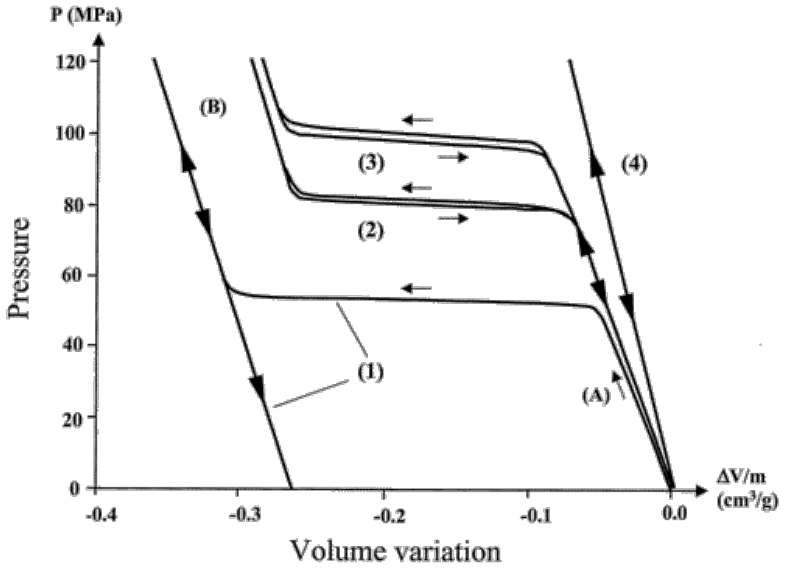


Figure 1.14 – Pressure-volume intrusion isotherms of water in (1) zeolite β; (2) silicalite-1 (OH^-); (3) silicalite-1 (F^-) and (4) Na-ZSM-5. Reprinted with permission from reference [47], copyright (2001) American Chemical Society.

The reported intrusion isotherms are reproduced in figure 1.14. From these results, the authors proposed that it was possible to use these heterogeneous systems to "accumulate, restore and dissipate mechanical energy", thus opening new routes in the field of energetics. In terms of energy storage devices, a system displaying an intrusion-extrusion cycle without hysteresis can simply be termed as a *spring*. A system with hysteresis is a *shock absorber* and an incomplete cycle in which water is retained in the porous framework upon pressure release can be called a *bumper* (see also the figure 5.1 for an illustration of these behaviors).

Patarin and coworkers[53, 54] were the first to show that the intrusion pressure in silicalite-1 was increased by a factor of 3 when using a concentrated solution of lithium

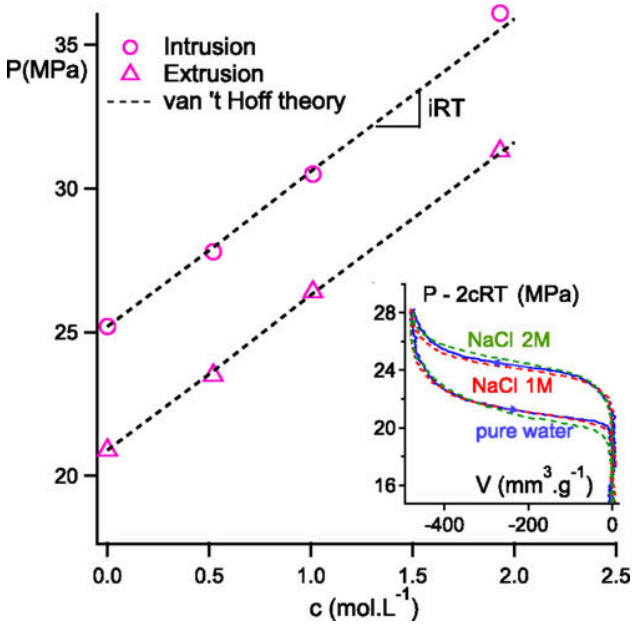


Figure 1.15 – Intrusion and extrusion pressures for NaCl solutions of various concentrations in ZIF-8 at 323 K. The inset displays intrusion–extrusion curves, plotted not in absolute pressure P *but as a function of* P − 2cRT*, showing that the effect of concentration is well explained by the van 't Hoff osmotic pressure. Reprinted with permission from reference [51], copyright (2015) American Physical Society.*

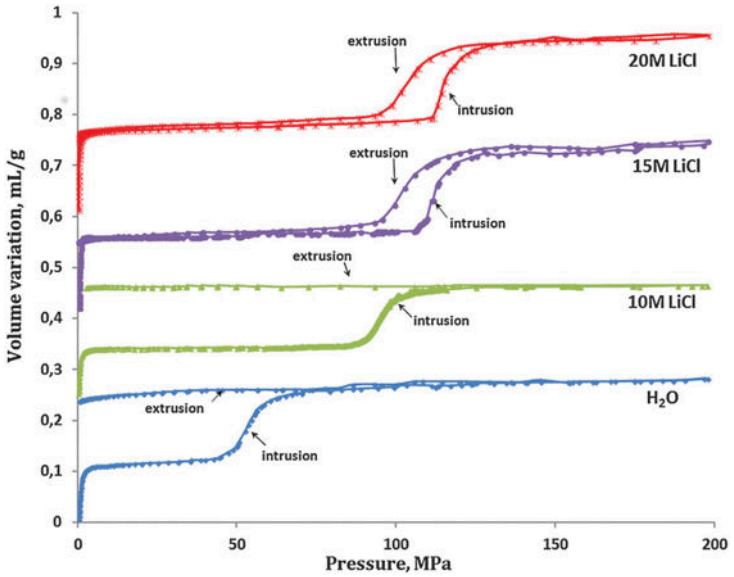


Figure 1.16 – Intrusion-extrusion process in β zeosil. With pure water, it behaves like a bumper. As the LiCl concentration increases, the intrusion pressure is enhanced and a change in behavior from bumper to shock-absorber is observed from 10 M to 15 M of LiCl. Reproduced from reference [52], with permission from the PCCP Owner Societies.

chloride, instead of pure water. Consequently, the stored energy was also increased by the same factor, which seems very promising in view of potential energy storage applications. This has lead to a number of such aqueous solutions intrusion-extrusion studies in various zeosils as well as hydrophobic ZIF materials. In addition to the effect of increasing the intrusion pressure, the use of electrolyte solutions can lead to a drastic change in the intrusion-extrusion behavior. Ryzhikov et al.[52] showed that β-zeosil changed its behavior from bumper to shock-absorber, as the LiCl concentration increased from 10 mol/L to 15 mol/L. This is shown in figure 1.16.

Intrusion of electrolyte solutions in ZIF-8 was recorded using several concentrations and different electrolytes[51, 55]. In all cases, the intrusion pressure increased as the electrolyte concentration increased and the intrusion-extrusion cycles were shifted replicas of the pure water cycles. Based on the findings of a molecular dynamics study[56], Michelin-Jamois suggested that only pure water was intruded in the material's pore. According to this hypothesis, the shift in intrusion pressure is due to the osmotic pressure, *i.e.* the difference in pressure between the pure water pressure inside the pore and the aqueous solution pressure outside the porous framework. A simple application of the van't Hoff osmotic equation ($\Pi = icRT$) confirmed this model in most of the studied cases, as illustrated in figure 1.15.

Even more recently, Arletti et al.[57] coupled intrusion-extrusion experiments of $MgCl_2$ aqueous solution to *in situ* high-pressure synchrotron X-ray diffraction analysis. This study clearly demonstrated the presence of both ions and water molecules in the high pressure intruded liquid. During my PhD, I approached the question of when and why ions in the aqueous solution enter the microporous frameworks using classical molecular simulations, the corresponding work is presented in chapter 5.

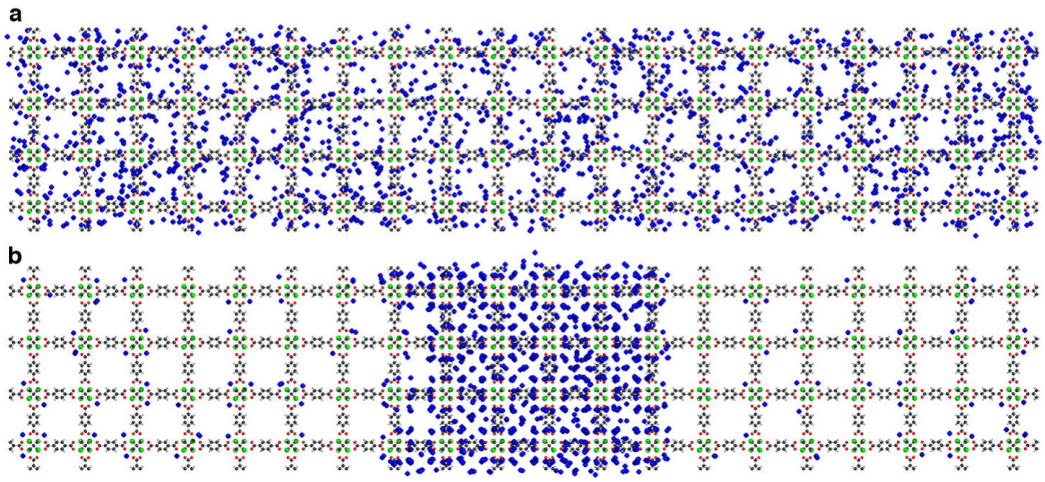


Figure 1.17 – Snapshots of molecular dynamics simulations of benzene in MOF-5 at (a) 125 K *and (b)* 50 K. *At* 50 K, *vapor–liquid coexistence takes place, the vapor and liquid phases extending over many unit cells. Adapted with permission from reference [58], copyright (2015) Wiley.*

1.2.3 Coupling adsorption and deformations

Adsorption and intrusion in flexible nanoporous materials can present very interesting behaviors. First, some materials deforms during adsorption, creating the aforementioned effects of breathing, gate opening or even negative gas adsorption[59]. At the same time, because they are constrained by the surrounding porous solid, fluids inside the pores adopt different static organization and dynamic behavior than fluids in the bulk state. These effects of the porous solids on the fluids are grouped together under the category of *confinement* effects.

Confinement effects introduce another way to look at adsorption and intrusion: two different phases of the same chemical species are at equilibrium, and transfers from one phase to the other occurs as the fluid enters or leave the porous structure. In some cases, there can even be multiple different fluid phases inside the porous volume. Recently, Braun et al.[58] demonstrated, using NMR relaxometry and molecular dynamics, that a true liquid–vapor coexistence existed in fluid benzene confined in MOF-5. The liquid and vapor phases were shown to extend over many unit cells thanks to the open 3D framework, as shown in figure 1.17.

While confinement effects also occurs in rigid nanoporous material, when working with flexible nanoporous materials we need to consider multiple phases equilibrium occurring at the same time. The fluid phase equilibrium between the bulk state and the confined state can have an influence on phase equilibrium between the porous material phases and reciprocally.

During my PhD work, I explored confinement effects mainly in ZIF-8, with different fluids: gaseous nitrogen at 77 K in chapter 4, and aqueous solutions at ambient temperature in chapter 5. I also looked at more rigid systems — namely aluminosilicate nanotubes — in chapter 5.2. I have been particularly interested in the simulation methods one can use to study the coupling between adsorption and deformations in flexible nanoporous materials.

2

MACROSCOPIC STUDIES

A first way to approach the problem of coupled deformation and adsorption in porous materials is through the use of macroscopic or thermodynamic modeling methods. Such methods are based on classical thermodynamic principles and allow us to gain high-level understanding of the processes at play. In this chapter, I will introduce the concept of thermodynamic potential and thermodynamic ensemble, that can be used together to predict the evolution of a system. In a second part, I will show how these thermodynamic potentials can be used to make predictions on the co-adsorption of multiple gases in the same porous framework, but that one needs to be careful to use the thermodynamic ensemble adapted to the problem at hand.

2.1 Classical thermodynamics

2.1.1 First law of thermodynamics

The starting point for the macroscopic methods I will describe here are the laws of thermodynamics. The first law of thermodynamics is a generalization of the conservation of energy during physical processes. It can be stated as follows:

> *For a closed system, the change in energy between two states of internal thermodynamic equilibrium is the sum of the work of external forces and the heat received by the system.*

If we divide the total energy $\mathrm{E_{tot}}$ of a system into its kinetic energy $\mathrm{E_{kin}}$, potential energy $\mathrm{E_{pot}}$ and internal energy $\mathcal{U}$, and letting W be the work received by a system during a transformation and Q the heat transfer during the same transformation, the first law of thermodynamics can be written as

$$\Delta \mathrm{E_{tot}} = \Delta \mathrm{E_{kin}} + \Delta \mathrm{E_{pot}} + \Delta \mathcal{U} = \mathrm{W} + \mathrm{Q}. \tag{2.1}$$

If now we consider a system at rest in a constant external potential field, $\mathrm{E_{kin}}$ and $\mathrm{E_{pot}}$ are constant, and we find the usual formulation of the first law of thermodynamics:

$$\Delta \mathcal{U} = \mathrm{W} + \mathrm{Q}. \tag{2.2}$$

It is often interesting to use a differential formulation of this relation, considering infinitesimal changes in $\mathcal{U}$, W, and Q:

$$\mathrm{d}\mathcal{U} = \delta\mathrm{Q} + \delta\mathrm{W}. \tag{2.3}$$

If the system composition is not constant, *i.e.* if it is undergoing chemical reactions or if the system is open and exchanges particles with an external reservoir, we need to add another term depending on the quantity of matter $\{n_i\}$ for each chemical species in the system:

$$\mathrm{d}\mathcal{U} = \delta\mathrm{Q} + \delta\mathrm{W} + \sum_i \mu_i \mathrm{d}n_i. \tag{2.4}$$

In this formulation, the μ_i are called the *chemical potentials*; they represent the relative stability of the different chemical species in the system.

An interesting special case is that of an external pressure acting on the system. In this case,

$$\delta\mathrm{W_{pressure}} = -p_{\mathrm{ext}}\mathrm{dV}\ ; \tag{2.5}$$

where p_{ext} is the external pressure. If p_{ext} is constant during the transformation, this can be written as $\delta\mathrm{W_{pressure}} = -\mathrm{d(PV)}$, using P as the constant value of the pressure. We can then define the *enthalpy* H as $\mathrm{H} = \mathcal{U} + \mathrm{PV}$; which allow us to rewrite equation (2.4) as

$$\mathrm{dH} = \delta\mathrm{Q} + \delta\mathrm{W_{other}} + \sum_i \mu_i \mathrm{d}n_i. \tag{2.6}$$

Here, $\delta\mathrm{W_{other}}$ is the work coming from all forces *excluding* pressure forces.

2.1.2 Second law of thermodynamics

The second law of thermodynamics allows us to make predictions on the "natural" evolution of a system.

> *There exist a function of state called entropy, which is an increasing function of time for any transformation of an isolated system.*

The variation of entropy, noted S, during a transformation of a closed but non isolated system can be linked to the heat transfer between the system and its surroundings by the Clausius relation:

$$\Delta S = \frac{Q}{T} + S_{created}. \tag{2.7}$$

The Q/T term is homogeneous to entropy and called the exchanged entropy, and is the entropy given to the system by its surroundings. $S_{created}$ is the entropy created during the transformation, and is a positive quantity by the second law of thermodynamics. When $S_{created}$ is null, the transformation is said to be *reversible*.

Using the relations (2.4) and (2.7); and considering that only the pressure forces act on the system, we can extract an evolution principle for any transformation:

$$TdS = \delta Q + T\delta S_{created} \tag{2.8}$$

$$TdS = (d\mathcal{U} - \delta W - \sum_i \mu_i dn_i) + T\delta S_{created} \tag{2.9}$$

$$d\mathcal{U} + PdV - \sum_i \mu_i dn_i - TdS = -T\delta S_{created}. \tag{2.10}$$

If there is a function Φ such as $d\Phi = d\mathcal{U} + PdV - \sum_i \mu_i n_i - TdS$, then for any transformation of a closed system, $d\Phi = -T\delta S_{created} \leq 0$. This means that Φ is a decreasing function of time, and it will be minimal in the equilibrium state. In this case, Φ is called a *thermodynamic potential*.

2.1.3 Thermodynamic ensembles

A thermodynamic ensemble is defined by a set of thermodynamic state variable that obey some constraints. For example, the ensemble where the composition of the system is fixed, together with the volume and total energy is called the NVE or microcanonical ensemble. Other well-known ensembles present a fixed pressure or temperature: the NVT or canonical ensemble, and the NPT or isobaric-isothermal ensemble.

Each of these ensembles has an associated thermodynamic potential. For the microcanonical ensemble, we start with equation (2.10):

$$d\Phi = d\mathcal{U} + PdV - \sum_i \mu_i dn_i - TdS. \tag{2.11}$$

$d\mathcal{U}$ and dV will be zero as the energy and volume is constant. Moreover as the system composition is fixed, $dn_i = 0$. This gives

$$d\Phi = -TdS \leq 0. \tag{2.12}$$

The thermodynamic potential for the NVE ensemble is the negentropy $\mathcal{N} \equiv -\mathrm{TS}$, which will be minimal (and the entropy will be maximal) at equilibrium.

If we now allow the energy to change but fix the kinetic energy and thus the temperature by the mean of an external thermostat, equation (2.10) becomes

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} + \mathrm{PdV} - \sum_i \mu_i \mathrm{d}n_i - \mathrm{TdS} - (\mathrm{SdT} - \mathrm{SdT}) \tag{2.13}$$

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} + \mathrm{PdV} - \sum_i \mu_i \mathrm{d}n_i - \mathrm{d(TS)} + \mathrm{SdT}. \tag{2.14}$$

As the temperature, volume and composition are fixed, $\mathrm{dT} = \mathrm{dV} = \mathrm{d}n_i = 0$, which means that

$$\mathrm{d}\Phi = \mathrm{d}(\mathcal{U} - \mathrm{TS}). \tag{2.15}$$

We can introduce the Helmholtz free energy $\mathrm{F} \equiv \mathcal{U} - \mathrm{TS}$ (sometimes simply called free energy), which is the thermodynamic potential in the NVT ensemble.

If now we allow the volume to change in the system, while fixing the pressure with a barostat and keeping the temperature fixed, equation (2.10) becomes

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} + \mathrm{PdV} + (\mathrm{VdP} - \mathrm{VdP}) - \mathrm{TdS} - (\mathrm{SdT} - \mathrm{SdT}) - \sum_i \mu_i \mathrm{d}n_i \tag{2.16}$$

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} + \mathrm{d(PV)} - \mathrm{d(TS)} - \mathrm{VdP} + \mathrm{SdT} - \sum_i \mu_i \mathrm{d}n_i \tag{2.17}$$

Again, as the pressure, temperature and composition are fixed, this reduces to

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} + \mathrm{d(PV)} - \mathrm{d(TS)}. \tag{2.18}$$

We define the Gibbs free energy G as $\mathrm{G} \equiv \mathcal{U} + \mathrm{PV} - \mathrm{TS} \equiv \mathrm{F} + \mathrm{PV} \equiv \mathrm{H} - \mathrm{TS}$, which is the thermodynamic potential in the NPT ensemble.

2.1.4 Ensembles for adsorption processes

Getting back to the topic of adsorption in porous material, we need to describe the thermodynamic ensemble in which the adsorption process takes place. The temperature is always fixed, as well as the composition of the adsorbing host. If the host is rigid, and does not deform under adsorption, then we can consider that the volume of the system is constant. The main difference with the previously defined ensembles is that the composition of the gas phase is not fixed: the number of gas molecules adsorbed in the system is allowed to change during adsorption. In this case, equation (2.10) becomes

$$\mathrm{d}\Phi = \mathrm{d}\mathcal{U} - \mathrm{d(TS)} + \mathrm{PdV} + \mathrm{SdT} - \sum_i \mu_i \mathrm{d}n_i \tag{2.19}$$

Removing the null terms dV and dT, we find the definition of the *grand-canonical* thermodynamic potential Ψ, associated with the μVT ensemble:

$$\Psi \equiv \mathcal{U} - \mathrm{TS} - \sum_i \mu_i n_i \equiv \mathrm{F} - \sum_i \mu_i n_i. \tag{2.20}$$

The Gibbs-Duhen relation ($\mathrm{VdP} - \mathrm{SdT} = \sum_i n_i \mathrm{d}\mu_i$) also gives us:

$$\Psi = -\mathrm{PV} \tag{2.21}$$

If we now consider adsorption in a flexible host, the volume is no longer fixed. Instead, in most experimental setups, external pressure is fixed and set equal to the pressure of the adsorbed gases outside of the host. The thermodynamic ensemble suited for the study of adsorption in flexible materials is called the *osmotic ensemble*, first introduced in 1994[60] for the study of fluid mixtures, and adapted to multi-components phase equilibrium in 1998[61]. In this ensemble, the pressure, temperature, number of atoms of the host phase N_{host} are fixed, as well as the chemical potential of the adsorbed species. Again, starting from equation (2.10)

$$d\Phi = d\mathcal{U} - d(PV) - d(TS) - VdP + SdT - \sum_i \mu_i dn_i. \tag{2.22}$$

As usual, dT and dP are null as the corresponding variable is fixed. In the sum over the $\mu_i n_i$ terms, only the adsorbed species quantities of matter are allowed to change. The osmotic potential is thus defined by

$$\Omega \equiv \mathcal{U} - TS + PV - \sum_i \mu_i n_i \equiv F + PV - \sum_i \mu_i n_i \equiv G - \sum_i \mu_i n_i \tag{2.23}$$

This potential will be the basis we use to predict co-adsorption in flexible porous media in the next section. The table 2.1 presents a summary of all the thermodynamic ensembles discussed so far, as well as the associated fixed quantities and thermodynamic potentials.

Ensemble	Fixed quantities	Thermodynamic potential		
microcanonical	N, V, E	$\mathcal{N}$	$\equiv$	$-TS$
Canonical	N, V, T	F	$\equiv$	$\mathcal{U} - TS$
Isobaric-Isothermal	N, P, T	G	$\equiv$	$H - TS$
Grand-Canonical	μ_i, V, T	Ψ	$\equiv$	$F - \sum_i \mu_i n_i$
Osmotic	N_{host}, μ_i, P, T	Ω	$\equiv$	$G - \sum_i \mu_i n_i$

Table 2.1 – Thermodynamic ensembles and associated thermodynamic potentials

2.2 Macroscopic calculations for gas separation in flexible materials

Gas separation is an important step in multiple industrial processes, from separation of hydrocarbons in oil chemistry, to CO_2 separation and storage or oxygen extraction in the air. The two main methods used for gas separation are cryogenic distillation, mainly used for air separation, and differential adsorption. Adsorption-based processes for gas separation, which rely on microporous materials in the form of an adsorber bed, are very versatile because of the large choice of materials available — and the possibility to tune them for a specific gas system. To choose an adsorber and the size of a production plant for the separation of a gas mixture, good knowledge on the co-adsorption of these gases in the candidate porous adsorbent is required.

Experimental characterization of this co-adsorption is typically done through multi-component gas adsorption studies. This problem is inherently high-dimensional, e.g., for a ternary mixture there are four variables to vary (temperature, total pressure, and two independent variables for the mixture composition). Because such experimental studies of co-adsorption equilibrium thermodynamics are typically long and expensive, there has been a great expense of literature devoted to theoretical models for the prediction of mixture co-adsorption based on single-component adsorption data. The most commonly used method in the field is the Ideal Adsorbed Solution Theory (IAST)[62], which is relatively simple to implement and robust, and allows the prediction of multi-component adsorption behavior from individual single-component isotherms. Other theories are used when the ideality of the system can no longer be assumed: non-ideal adsorbed solution models[42, 63], the vacancy solution theory (VST)[64], the Real Adsorbed Solution Theory [65] (RAST), *etc.*

Materials which undergo large-scale reversible structural transitions impacting their total volume or internal pore volume appear to be particularly common among metal-organic frameworks based on relatively weaker bonds (coordination bonds, π–π stacking, hydrogen bonds, or some covalent bonds) compared to inorganic dense nanoporous materials (such as zeolites). In particular, some of these materials show transitions between an *open* phase with large pore volume, and a *condensed* or *narrow pore* phase with smaller pore volume — or, in some cases, no microporosity at all. Such transitions, known as *gate-opening*[66–68] or *breathing*[23] depending on the order in which the phases occur upon adsorption, can lead to stepped adsorption isotherms for pure components.

In recent literature, many authors have relied on IAST predictions to predict that several such flexible MOFs would present very good selectivity for gas separation. In some cases, the authors explicitly used IAST to derive such predictions on flexible materials[69–72]. In other cases, IAST was not used explicitly, but the assumptions made for the behavior of mixtures stem from the *classical* understanding of selectivity rules in rigid materials, and would not necessarily be valid in flexible materials[73–78].

IAST is one of the available methods for direct macroscopic calculations and predictions, taking macroscopic data in the form of pure component adsorption isotherms and predicting macroscopic co-adsorption isotherms. In practice, using IAST is akin to working in the Grand-Canonical ensemble, but we showed in the previous section that when the

adsorption host is flexible, one should use the Osmotic ensemble. Instead, an alternative method called the Osmotic Framework Adsorbed Solution Theory (OFAST)[79], based on the Osmotic ensemble, should be used when structural transitions occur during adsorption. This theory has been developed in the group[80, 81], and takes into account the coupling between adsorption and deformations in flexible porous materials.

In this section, I will compare the results of IAST and OFAST on two sets of adsorption data from the published literature on gate-opening materials, and show that the IAST method gives unrealistic results: it does not reproduce the gate-opening behavior upon mixture adsorption, and overestimates the selectivity by up to two orders of magnitude. All of this work is published in *Adsorption* (2018)[82].

2.2.1 Ideal Adsorbed Solution Theory

The Ideal Adsorbed Solution Theory (IAST) starts by assuming that for a given adsorbent and at fixed temperature T, the pure-component isotherms $n_i(\mathrm{P})$ for each gas i of interest is known. Then, given a mixture of ideal gases adsorbing at total pressure P in a host framework and the composition of the gas phase y_i — such that the partial pressures follow the ideal mixing law $\mathrm{P}_i = y_i\mathrm{P}$ — the goal of the method is to predict the total adsorbed quantity n_{tot} and the molar fractions x_i in the adsorbed phase.

In order to do so, Myers et al.[62] introduced for each mixture component a quantity homogeneous to a pressure, P_i^*. The IAST method links this pressure to the compositions of the gas and adsorbed phases by defining a link between P_i^* and the known variables:

$$\mathrm{P}y_i = \mathrm{P}_i^* x_i; \tag{2.24}$$

and by imposing the equality of chemical potentials at thermodynamic equilibrium:

$$\forall i,j \quad \int_0^{\mathrm{P}_i^*} \frac{n_i(p)}{p}\mathrm{d}p = \int_0^{\mathrm{P}_j^*} \frac{n_j(p)}{p}\mathrm{d}p. \tag{2.25}$$

In the simpler case of two-component gas mixture containing gases B and C, these two equations and the conservation of matter can be rewritten as a set of four equations:

$$\begin{aligned} \mathrm{P}y_\mathrm{B} &= \mathrm{P}_\mathrm{B}^* x_\mathrm{B} & x_\mathrm{B} &= \frac{\mathrm{P}_\mathrm{C}^* - \mathrm{P}}{\mathrm{P}_\mathrm{C}^* - \mathrm{P}_\mathrm{B}^*} \\ \frac{1}{n_{\mathrm{tot}}} &= \frac{x_\mathrm{B}}{n_\mathrm{B}(\mathrm{P}_\mathrm{B}^*)} + \frac{1-x_\mathrm{B}}{n_\mathrm{C}(\mathrm{P}_\mathrm{C}^*)} \\ \int_0^{\mathrm{P}_\mathrm{B}^*} \frac{n_\mathrm{B}(p)}{p}\mathrm{d}p &= \int_0^{\mathrm{P}_\mathrm{C}^*} \frac{n_\mathrm{C}(p)}{p}\mathrm{d}p \end{aligned} \tag{2.26}$$

Solving these equations for P_B^* and P_C^* will give all the information on the system composition. It can be done with either numerical integration of the isotherms, or by fitting the isotherms to a model, and then integrating the model analytically.

2.2.2 IAST and flexible frameworks

The original derivation of the IAST equations[62] by Myers et al. highlights three hypotheses on the co-adsorption process on which the model is built:

(H1) The adsorbing framework is inert from a thermodynamic point of view;

(H2) The adsorbing framework specific area is constant with respect to temperature and the same for all adsorbed species;

(H3) The Gibbs definition of adsorption applies.

While the meaning of the last assumption **(H3)** has been diversely interpreted by different authors, Myers originally meant[62] and later confirmed[83] it to qualify the method by which the adsorption isotherms are measured. There is consensus on the fact that absolute adsorption should be used in IAST calculations — as opposed to excess or net adsorption[84]. This assumption thus applies equally to both rigid and flexible adsorbents. However, the first two hypotheses are not valid for flexible nanoporous materials. **(H2)** is clearly invalid, as modifications in both the host's volume and internal structure lead to variations of pore size and specific area upon structural transitions. We note here, in passing, that **(H2)** should already be ruled out for systems of pore size close to the adsorbate diameter, as well as gas mixtures of widely different size or shape. It should, for example, not apply to molecular sieves systems, yet those can often be described reasonably well by IAST in practice. Finally, **(H1)** is violated by all the systems that feature adsorption-induced deformation, and in particular by systems presenting gate-opening or breathing behavior. As a conclusion, IAST has no theoretical foundation for those systems and should not be used for co-adsorption prediction in flexible nanoporous frameworks.

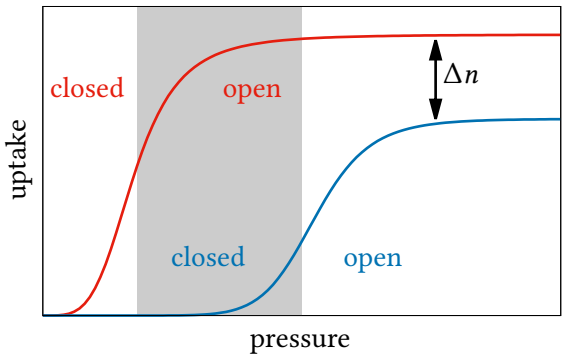


Figure 2.1 – Typical single-component isotherms for adsorption of two gases (red and blue) in a material with gate-opening. The gate-opening pressure is not the same for the two adsorbates, creating a pressure range with a high difference in the adsorption capacity for single-component isotherms (gray zone in the figure). Contrary to intuition, selectivity will not necessarily be high in this pressure range, but will depend on the difference in saturation uptake Δn.

Aside from the mathematical treatment and thermodynamic hypotheses, we can show in a qualitative way why it is not possible, in flexible host frameworks, to use the single-component isotherm directly to predict multi-components adsorption. We address here

a common misconception, due to an invalid graphical interpretation of the isotherms. Figure 2.1 depicts the equilibrium adsorption isotherms for two different guests in a material presenting a gate-opening behavior. Gate-opening is an adsorption-induced structural transition from a non-porous to a porous phase of the host, leading to a step in the single-component adsorption isotherm. Gate-opening occurs at two different pressures for the two adsorbates, due to the specific host–guest interactions of the two gases (characterized notably by the enthalpy of adsorption and saturation uptake). In the pressure range in-between the transition pressures (in gray in figure 2.1), the uptake of one species is close to zero in the single-component isotherm and the uptake of the other species is close to its maximum value. If these isotherms were encountered for a rigid host material, the adsorption selectivity would be extremely high in this range, with one guest adsorbing but not the other.

Yet, the step in the isotherms here is not simply linked to host–guest interactions but instead to a change in the host structure. In particular, upon adsorption of a gas mixture in this gate-opening framework, a phase transition of the host structure will occur at a given pressure. Before this transition, the structure will be contracted and show no (or little) adsorption for either guest, and thus no usable selectivity. After the transition, *both species* will adsorb into the open pore framework. The selectivity is then governed — at least qualitatively — by the respective saturation uptakes of the two fluids (Δn in figure 2.1). While the difference in adsorbed quantities in the intermediate pressure range visually suggests great selectivity, it is not possible for one component to adsorb inside the close phase framework while at the same time the other component adsorbs inside the open phase of the framework. The framework is either in one phase or in the other, at any given time.

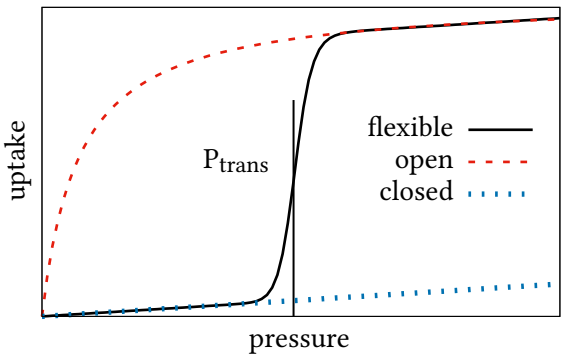


Figure 2.2 – Generation of the total isotherm in gate-opening materials by the combination of two single-phase isotherms: an open *pores isotherm, and an* closed *pores isotherm. The transition between the two host phases occurs at* P_{trans}.

The whole issue with using single-component isotherms to predict multi-component adsorption in frameworks with phase transition boils down to the origin of the stepped isotherms. The single-component isotherm (represented in figure 2.2) is a combination of two isotherms: one in the first phase (the contracted pore phase), and one in the second phase (the open pore phase). Both phases — and the thermodynamic equilibrium between them — need to be taken into account to predict the multi-component adsorption.

2.2.3 Osmotic Framework Adsorbed Solution Theory

As established in section 2.1.4, the thermodynamic ensemble suited for the study of adsorption in flexible materials is the osmotic ensemble. Let's recall that in this ensemble, the thermodynamic potential Ω is a function of the mechanical pressure P, the temperature T, the number of atoms in a given host phase α and the adsorbed species chemical potentials μ_i:

$$\Omega(\mathrm{T}, \mathrm{P}, \mu_i) = \mathrm{F}_\alpha + \mathrm{PV}_\alpha - \sum_i \mu_i n_{\alpha,i}, \tag{2.27}$$

where F_α is the Helmholtz free energy of the empty host in phase α, V_α the volume of the host in this phase, and $n_{\alpha,i}$ the molar uptake of guest i in phase α. This expression can be reworked and expressed as a function not of chemical potentials, but of fluid pressure (taken equal to mechanical pressure P) and adsorption isotherms:[85]

$$\Omega(\mathrm{T}, \mathrm{P}, \mu_i) = \mathrm{F}_\alpha + \mathrm{PV}_\alpha - \sum_i \int_0^{\mathrm{P}} n_{\alpha,i}(\mathrm{T}, p)\mathrm{V}_i^m(\mathrm{T}, p)\,\mathrm{d}p \tag{2.28}$$

Here, $n_{\alpha,i}(\mathrm{T}, \mathrm{P})$ are the co-adsorption isotherms for each component and $\mathrm{V}_i^m(\mathrm{T}, \mathrm{P})$ the molar volume for the species i in the bulk phase. Supposing that the gases are ideal, the molar volume is given by RT/P, with R the ideal gas constant.

I have shown previously that IAST cannot be used for the study of co-adsorption in frameworks with adsorption-induced phases transition, because the framework is not inert during adsorption. However, the IAST assumptions are still valid for each individual phase of the host matrix, if they are considered in the absence of a transition. As a consequence, it means that the IAST model can be used, for each possible host phase α, to calculate the co-adsorption isotherms $n_{\alpha,i}(\mathrm{P}, \mathrm{T})$ in this given phase. Then, the thermodynamic potential of each phase Ω_α can be calculated from these isotherms through Equation (2.28), allowing to predict which phase is the more stable at a given gas phase pressure and composition — and where the structural transition(s) occur. This method, extending the IAST theory in the osmotic ensemble to account for host flexibility, has been called Osmotic Framework Adsorbed Solution Theory (OFAST)[79, 80].

Although the amount of published data from direct experimental measurements of co-adsorption of gas mixtures in flexible MOFs is very limited, the OFAST method has been well validated in the past against experimental data.[81, 86, 87]. For example, Ortiz et al.[81] compared the method against experimental data for adsorption of CO_2/CH_4 mixtures in the MIL-53(Al) MOF. MIL-53(Al) is the seminal example of material with a *breathing* behavior: at low loadings, its most stable phase is the high porous volume *open-pore* phases, at intermediate loadings it transitions to a *closed-pore* phase, and at high loading it goes back to the open-pore structure. One of the very interesting things that Ortiz et al. predicted using OFAST is the increase of the stability domain of the closed pore phase when using mixtures with respect to pure component adsorption. The predicted phase diagrams of MIL-53(Al) is reproduced in figure 2.3.

In practice, the use of OFAST takes the following steps. First, the host phases of interest are identified and the single-component adsorption isotherms $n_{\alpha,i}(\mathrm{T}, p)$ for these are obtained: this can be achieved from a fit of experimental isotherms (see figure 2.2) or from molecular simulation.

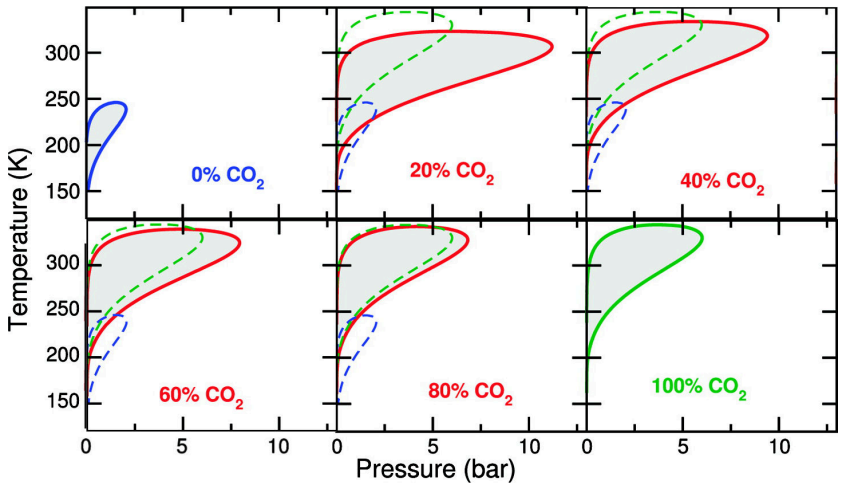


Figure 2.3 – Temperature–pressure phase diagram of MIL-53(Al) upon adsorption of a CO_2/CH_4 mixture, with increasing CO_2 molar fraction. Dashed lines correspond to pure component diagrams: blue for pure CH_4, green for pure CO_2, plain lines correspond to different mixtures. Reprinted with permission from reference [81], copyright (2011) American Chemical Society.

Secondly, the relative free energies of the host phases (which reduces to a single ΔF_{host} in our case of two host phases) can be computed from equation (2.28) and the experimental single-component stepped isotherm. For example, with two phases α and β, and considering ideal gas, we can express equation (2.28) for each phase:

$$\Omega_\alpha(T, P, \mu_i) = F_\alpha + PV_\alpha - RT \sum_i \int_0^P \frac{n_{\alpha,i}(p)}{p} \, dp \tag{2.29}$$

$$\Omega_\beta(T, P, \mu_i) = F_\beta + PV_\beta - RT \sum_i \int_0^P \frac{n_{\beta,i}(p)}{p} \, dp \tag{2.30}$$

At the transition ($P = P_{trans}$ in figure 2.2, which is typically known experimentally) the two thermodynamic potentials will be equal, which gives us a way to evaluate the free energy difference between the phases:

$$\Delta F_{host} = RT \sum_i \int_0^{P_{trans}} \frac{\Delta n_i(T, p)}{p} dp - P_{trans} \Delta V_{host} \tag{2.31}$$

Then, for all values of thermodynamic parameters of interest (pressure and gas mixture composition) the osmotic potential of the host phases is computed, enabling the identification of the most stable phase: the phase with the lowest osmotic potential is the most stable at this pressure and composition. The pressures at which the osmotic potential in two phases are equal are the phase transition pressures for a given composition. Finally, we can compute adsorption properties (guest uptake and selectivity) using IAST in each phase, within its domain of stability.

2.2.4 Comparing IAST and OFAST

I present here two examples of co-adsorption of gas mixtures in metal-organic frameworks with gate-opening behavior, based on experimental data from the published literature, comparing the predictions of IAST with those of OFAST. The first example deals with the adsorption of CO_2, CH_4, and O_2 in the Cu(dhbc)$_2$(4,4′-bpy) MOF[66] (see figure 2.4; dhbc = 2,5-dihydroxybenzoate; bpy = bipyridine). These isotherms correspond very closely to the archetypal *gate-opening* scenario described above. The second example deals with linear alkanes (ethane, propane, and butane) adsorption in RPM3-Zn MOF[75]; figure 2.7 presents the framework structure of RPM3-Zn and relevant experimental adsorption and desorption isotherms.

Simple isotherms in Cu(dhbc)$_2$(4,4′-bpy)

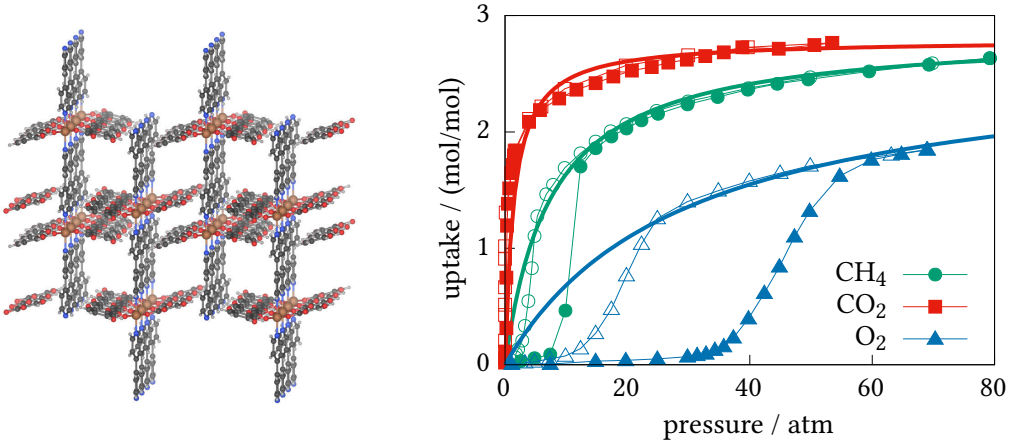


Figure 2.4 – (left) Cu(dhbc)$_2$(4,4′-bpy) structure (from reference. [66]). (right) Sorption isotherms and model isotherms fit at 298 K *in Cu(dhbc)$_2$(4,4′-bpy) for various gas compounds. Adsorption data are presented using filled symbols, and desorption data using empty symbols. Thick lines are Langmuir isotherms fitted at high loading. Experimental data published by Kitaura et al.[66]*

Cu(dhbc)$_2$(4,4′-bpy) is a textbook example of gate-opening upon adsorption, with single-component adsorption isotherms (reproduced in figure 2.4) that clearly shows the transition from a nonporous (at low gas pressure) to a microporous (at higher pressure) host phase. From the experimental data[66] I fitted the isotherms at high loading using a Langmuir model (equation (2.33)) for the isotherm in the open-pore structure; and at low loading using a Henry isotherm model (equation (2.32)) for the closed-pore structure following the model proposed in figure 2.2. Langmuir model is able to reproduce type I isotherms, and Henry model is usually valid for low loadings.

$$N(p) = K_H\, p \tag{2.32}$$

$$N(p) = N_L\, \frac{K_L\, p}{1 + K_L\, p} \tag{2.33}$$

The resulting fit parameters are in table 2.2. In the closed-pore phase, I assumed that no adsorption takes place in the whole pressure range, this is why the K_H coefficient is always zero. Using these parameters and equation (2.31), I computed the free energy

Gas	K_H / (mol/atm)	N_L / mol	K_L / atm	P_{trans} / atm	ΔF / (kJ/mol)
CH_4	0.0	2.86	0.134	5	-3.56
CO_2	0.0	2.79	0.699	1	-3.59
O_2	0.0	2.68	0.034	20	-3.39

Table 2.2 – Fitted coefficients for the sorption isotherms and free energy difference between open and closed structures in Cu(dhbc)$_2$(4,4′-bpy). See equations (2.32) and (2.33) for the definitions of K_H, N_L and K_L.

difference for all the isotherms. I took the value of -3.5 ± 0.1 kJ/mol as the free energy difference between the phases.

I performed OFAST calculations using Wolfram Mathematica, the corresponding code is available in the lab's GitHub repository[88]. I used the PyIAST Python package for the pure IAST calculations[89]. For these IAST calculations, I did not fit the isotherms to a specific model, but rather solved the IAST equations by numerical integration and interpolation between experimental data points. At partial pressures higher than the last point in the experimental isotherm, that last point was used as saturation uptake. I computed both partial loading and selectivity between the different gas in a mixture.

Figure 2.5 presents the selectivity obtained with IAST and OFAST; and figure 2.6 shows the partial and total loadings for all the gas combinations. The adsorption selectivity calculated with OFAST follows what one would expect: at low pressure, the pores are closed and no gas enters the structure, making the selectivity ill-defined — the isotherms at low pressure cannot be fitted and exploited for calculation of separation. Then, at a pressure depending on the composition of the gas phase, the gate-opening transition occurs. At pressure higher than gate-opening pressure, the framework is in its open pore form, and the value of selectivity depends on the relative saturation uptake of the two phases. The selectivities observed are almost independent of the fluid mixture composition, they are ≈ 20 for CO_2/O_2 and ≈ 4 for CH_4/O_2 mixtures.

In stark contrast with this picture, the selectivities calculated by IAST are clearly non-physical. All selectivity curves present a maximum in the pressure range where gate-opening occurs, with selectivities that can be several orders of magnitude too high, with for example 2 000 instead of 20 for CO_2/O_2. Even at higher pressure — above the gate-opening pressure range — the behavior is not identical to the OFAST calculations, because the incorrect behavior at low pressure affects IAST directly in the integration of the isotherms (equation (2.26)).

Moreover, the IAST selectivity for CO_2/O_2 presents a big jump around 40 atm when $y_{CO_2} = 0.1$. Looking at the partial loading in figure 2.6 top right panel, we can attribute this jump to an equilibrium displacement: O_2 is replacing CO_2 in the structure. This shows again the fact that IAST behaves as if the structure was closed for O_2 while at the same time being open for CO_2 at pressures lower than 40 atm. I thus confirmed by a quantitative study that IAST is not adapted for adsorption in flexible nanoporous materials.

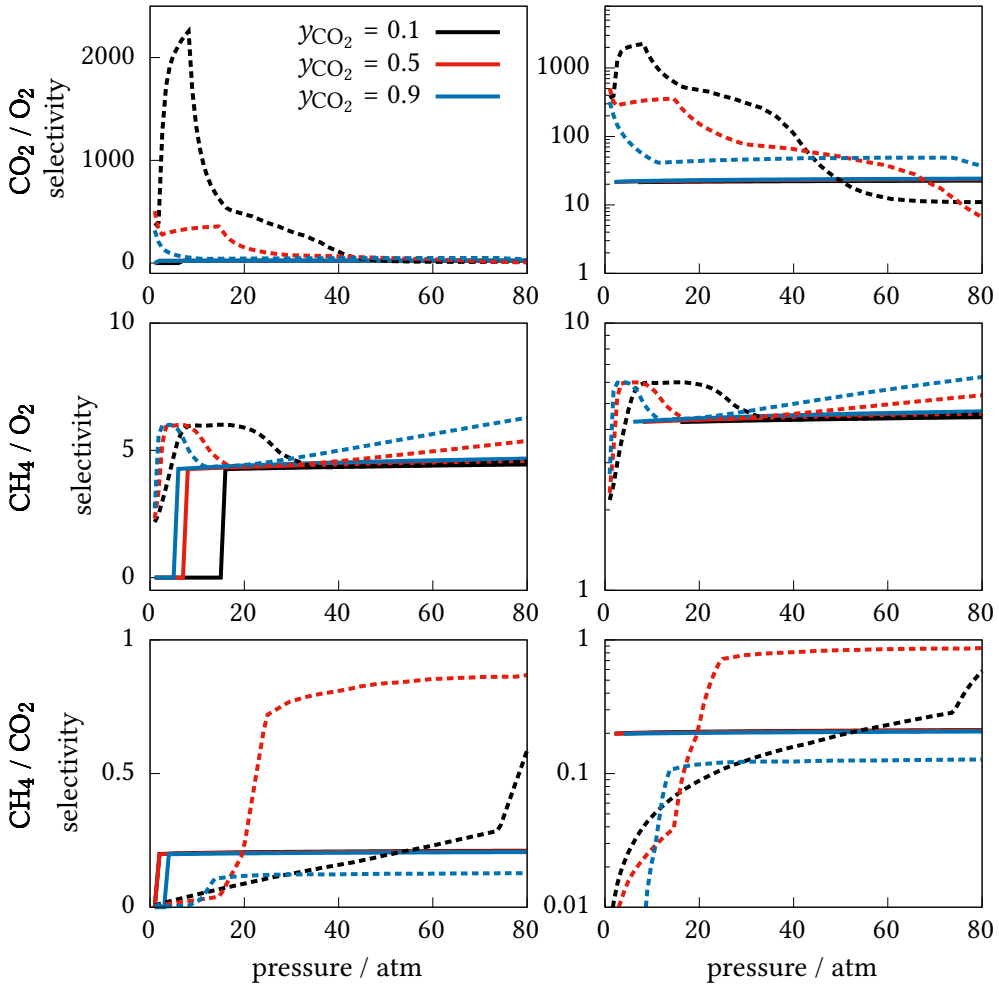


Figure 2.5 – Comparison of IAST (dashed lines) and OFAST (plain lines) adsorption selectivity for CO_2/O_2 (top); CH_4/O_2 (middle) and CH_4/CO_2 (bottom) mixtures in Cu(dhbc)$_2$(4,4′-bpy). The same curves are presented twice, using linear scale for the y axis on the left panels, and logarithmic scale on the right panels.

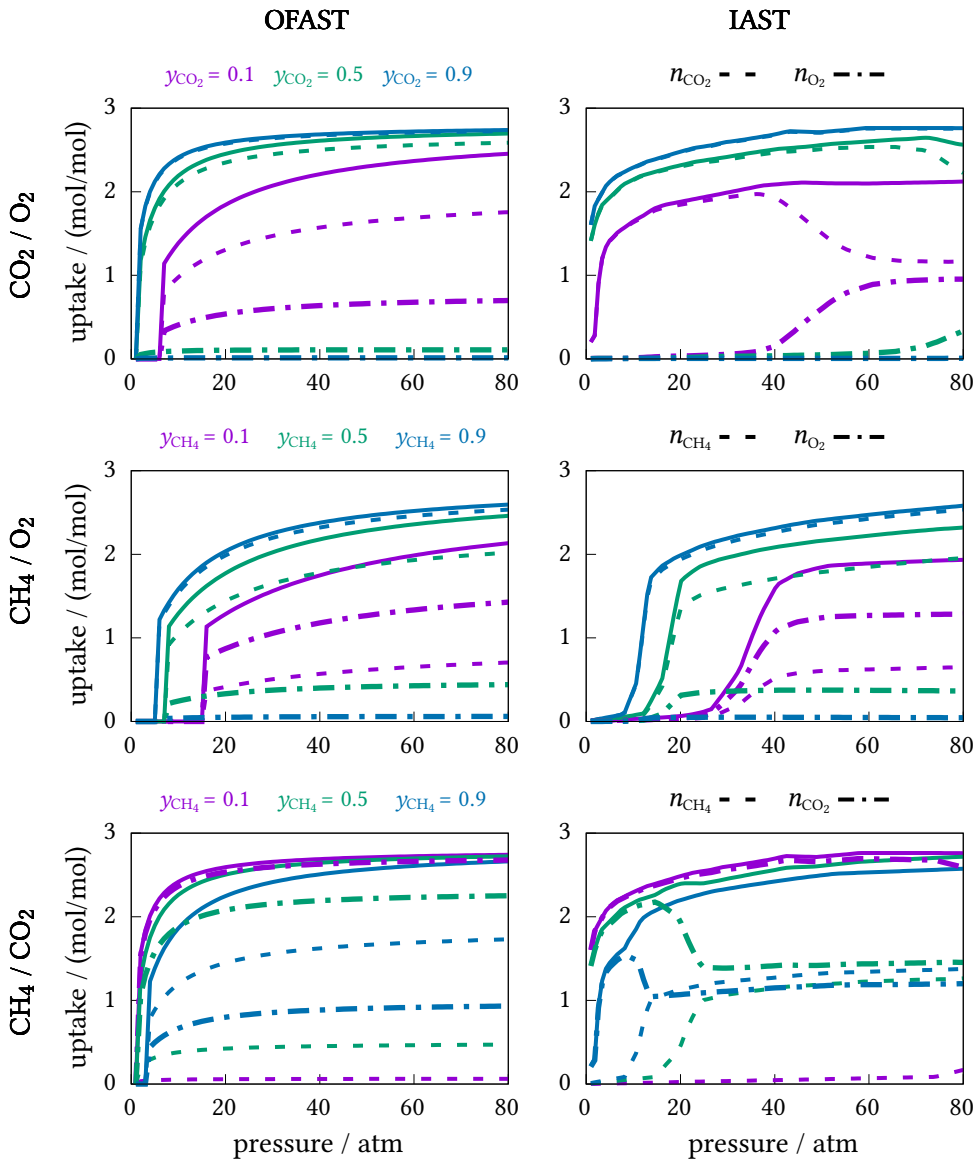


Figure 2.6 – Total (full lines) and partial (dashed lines) loading as function of pressure in Cu(dhbc)$_2$(4,4′-bpy) for all the gas pairs: from top to bottom CO_2 / O_2; CH_4 / O_2; and CH_4 / CO_2. OFAST results are presented on the left, and IAST results on the right.

More complex isotherms: the case of RPM3-Zn

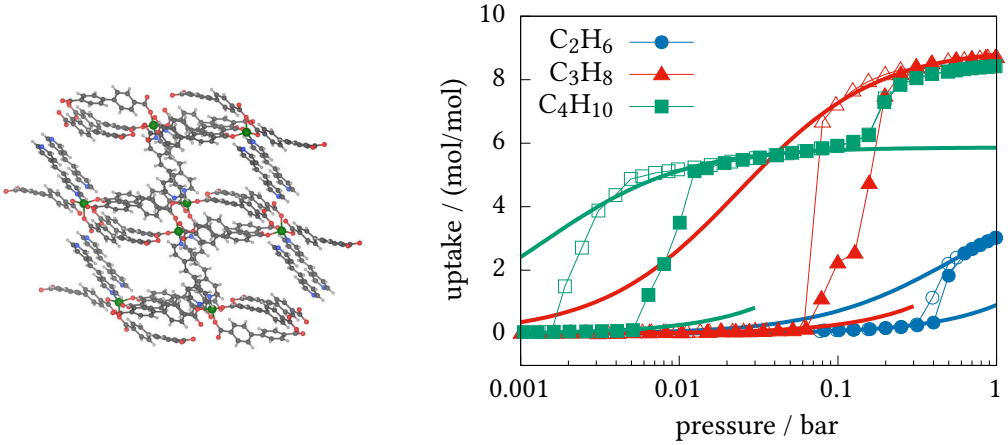


Figure 2.7 – (left) RPM3-Zn structure (from reference [90]). (right) Sorption isotherms at 298 K *for short alkanes in RPM3-Zn. Blue circles are for C_2H_6, red triangles for C_3H_8, and green squares for C_4H_{10}. Filled symbols for adsorption, empty symbols for desorption. Thick lines are the open and closed phases fit of the isotherms. Experimental data published by Nijem et al.[75]*

We now turn to a second example of gate-opening material, RPM3-Zn [90], which presents more complex adsorption–desorption isotherms for short alkanes (ethane, propane, butane) — depicted on the right panel of figure 2.7. While adsorption of C_2H_6, and C_3H_8 in this material displays a typical gate-opening behavior, with a well-marked single transition from a nonporous to a microporous phase, the adsorption of C_4H_{10} presents two steps at 0.01 atm and 0.2 atm. There, the first transition can be attributed to the structural transition (gate-opening), but the second one is of a different nature. Because there is no hysteresis loop for the second step, and because it occurs for the larger and more anisotropic guest molecule, it can be attributed to a fluid reorganization (or fluid packing) transition inside the pores. Because experimental *in situ* characterization (such as single X-ray diffraction) would be necessary to definitely affirm the character of this second step, I chose to avoid the issue and work in a reduced pressure range — although the OFAST method itself works with host materials with more than two phases. I thus fitted the C_4H_{10} isotherm using a Langmuir isotherm for pressures below 0.2 atm. The OFAST selectivity after this pressure will thus not be quantitatively accurate, but will be sufficient for the needed physical insight. I also performed tests by computing the selectivity under the assumption that the second jump is due to fluid reorganization by using Langmuir-Freundlich isotherms instead of single site Langmuir isotherm in the open phase, and the selectivity only differs at pressures higher than 0.2 atm.

From the C_3H_8 and C_4H_{10} isotherms, I computed the free energy difference between the nonporous and microporous phases, which I find to be $\Delta F = -30.0 \pm 0.1$ kJ/mol. The details are in table 2.3. I did not use the C_2H_6 isotherms for this purpose, as it has only limited data at high loading (for pressures above 1 bar), which somewhat increases the uncertainty of the fit. I was still able to fit the C_2H_6 isotherm with a Langmuir model and use it to compute co-adsorption data, as the free energy difference between the two host phases do not depend on the gas.

Gas	K_H / (mol/bar)	N_L / mol	K_L / bar	P_{trans} / bar	ΔF / (kJ/mol)
C_2H_6	0.905	4.82	1.74	/	/
C_3H_8	2.88	9.00	42.0	0.07	-30.1
C_4H_{10}	27.3	5.87	699	0.01	-29.8

Table 2.3 – Fitted coefficients for the sorption isotherms and free energy difference between open and closed structures in RPM3-Zn. See equations (2.32) and (2.33) for the definitions of K_H, N_L and K_L.

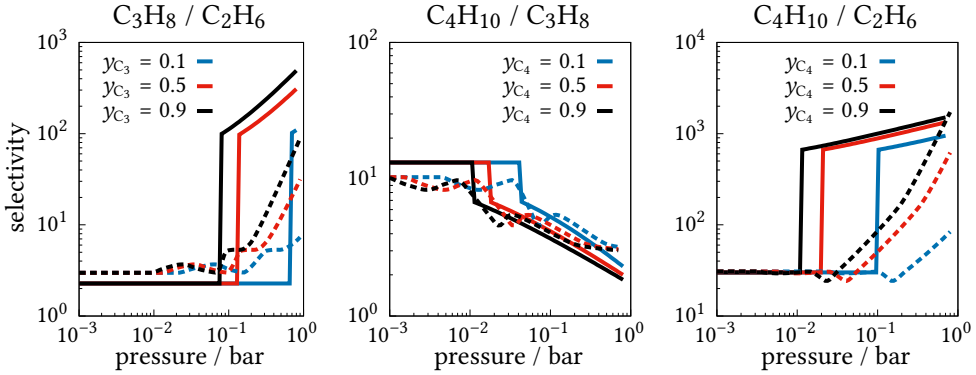


Figure 2.8 – IAST (dashed lines) vs OFAST (plain lines) adsorption selectivity for C_3H_8/C_2H_6 (left) and C_4H_{10}/C_3H_8 (right) mixtures in RPM3-Zn at different compositions.

Figure 2.8 displays the selectivity curves obtained with IAST and OFAST for various gas mixtures and compositions in RPM3-Zn; and figure 2.9 shows the partial and total loading predicted by both methods. Again, the OFAST selectivity curve follows the expected behavior: it is constant at low loading, where single-component isotherms follow the Henry model. In this low-pressure region, adsorption is negligible and the selectivity cannot be exploited in adsorption-based processes. However, we can see that because IAST is using numerical integration, it is much more sensitive to details in the single-component isotherms than the OFAST method, which is based on fits.

OFAST correctly describes the occurrence of gate-opening, at a pressure which depends on mixture composition but is in the range of the pure component gating pressures. After gate-opening, the selectivity jumps to its value in the open pore framework. C_3H_8/C_2H_6 mixtures have a behavior similar to that observed in Cu(dhbc)$_2$(4,4′-bpy), with a slowly growing (in logarithmic scale) selectivity at high loading. On the other hand, OFAST selectivity for C_4H_{10}/C_3H_8 mixture displays a different behavior. The selectivity is lower after the transition than before, and further decreases as the pressure and loading increases. This is due to the fact that the single-component isotherms in the open pore structure cross, with C_3H_8 adsorbing more than C_4H_{10} for pressure bigger than 0.03 bar. Thus, the low-pressure selectivity is reversed at high pressure.

In contrast, the IAST fails to describe the gate-opening phenomenon, with selectivity showing a continuous evolution. Even the trends displayed by this evolution are in poor agreement and make no physical sense, featuring non-monotonic evolution as a function of pressure and composition.

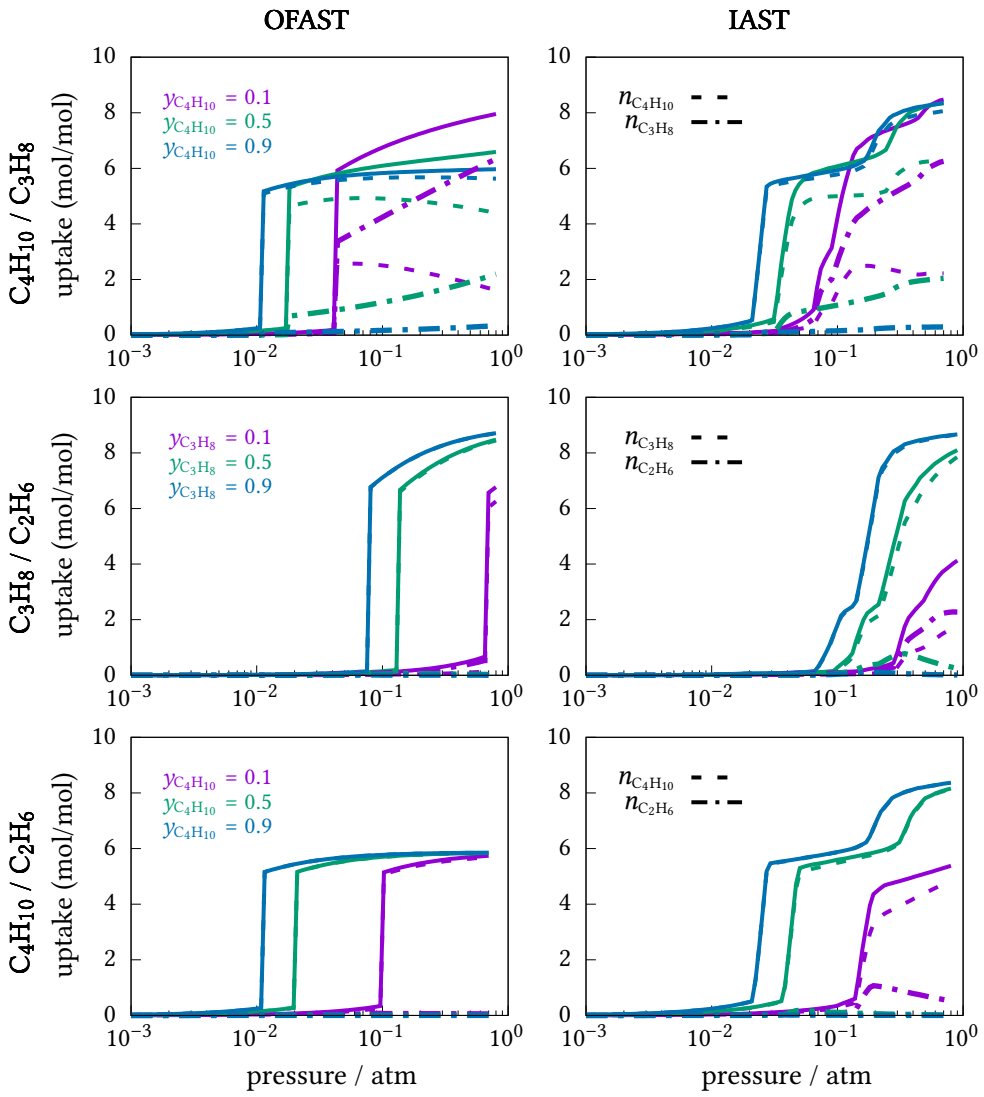


Figure 2.9 – Total (full lines) and partial (dashed lines) loading as function of pressure in RPM3-Zn for all the gas pairs: from top to bottom C_4H_{10} / C_3H_8; C_3H_8 / C_2H_6; and C_4H_{10} / C_2H_6. OFAST results are presented on the left, and IAST results on the right.

Conclusions

In this chapter, I presented and compared the existing IAST and OFAST macroscopic models for the prediction of co-adsorption of fluid mixtures in two different frameworks presenting a gate-opening behavior. In both cases, the selectivities derived by the IAST method are nonphysical and differ widely from the OFAST results, over- or under-estimate the selectivity, sometimes by up to two orders of magnitude. Moreover, this shows that even without explicitly using IAST for calculations of selectivity in flexible frameworks, one has to be cautious in comparing single-component isotherms of different guests. Differences in step pressure of stepped isotherms can lead to claims of strong selectivity using flexibility, when applying — without noticing it — concepts that are valid only for rigid host matrices.

Macroscopic methods such as the ones used here are only applicable under restrictive hypotheses; for example both OFAST and IAST assume an ideal mixture of perfect gases when deriving and solving the equations. Going past these hypotheses is possible but requires empirical models for the chemical potentials such as the ones used by non-ideal adsorption models such as VST or RAST. These theories could be coupled to the osmotic ensemble to create extensions to OFAST able to take into account the non-ideality of the system.

Despite these limitations, macroscopic modeling methods are very useful in the study of adsorption in flexible materials as they are orders of magnitude faster than fully atomistic methods and allow for faster screening and estimations of the performance of different materials. If we want to overcome the limitations of such macroscopic methods, we can turn to atomistic simulations. Such simulations allow exploring non-ideal systems containing real gases without restricting ourselves to a specific expression for the chemical potentials. In the next chapter, I am going to present the framework of statistical thermodynamics that is used to link together atomistic and macroscopic description of a system.

3

MOLECULAR SIMULATIONS

Macroscopic modeling methods such as the ones I discussed in the previous chapter are not always sufficiently precise to gain a complete understanding of the phenomenon at play. In particular, as these methods describe the systems at the macroscopic level, they don't take into account the individual atoms, and the interactions between them: they don't describe the *chemistry* of the system. Statistical thermodynamics is a tool that we can use to bring together the microscopic description of matter and the macroscopic behavior and characteristics of the system (pressure, temperature, …).

In this chapter, I will derive and recall some concepts from statistical thermodynamics I used during my PhD. For a more in-depth description of statistical mechanics, I recommend the book of Tuckerman[91]. I will then present the molecular simulation methods used to sample thermodynamic ensembles in practice. On this section, I recommend the book of Frenkel and Smit[92].

3.1 Statistical thermodynamics

3.1.1 Maxwell-Boltzmann statistics

We will consider an atomistic system containing N individual atoms behaving as classical particles with individual positions $\boldsymbol{r}_i$, identical masses m and momentum $\boldsymbol{p}_i$. Supposing that these particles are in some container of fixed volume V, and at thermal equilibrium with a thermostat at temperature T, they evolve in the canonical or NVT ensemble. Finally, we will also assume that the atoms in the system follow the Maxwell-Boltzmann statistics, *i.e.* that the probability density of finding the system in a state of internal energy E_i is given by:

$$\mathcal{P}_i = \frac{1}{\mathrm{N}!\ h^{3\mathrm{N}}\ \mathrm{Z}} e^{-\beta\,\mathrm{E}_i}, \tag{3.1}$$

where $\beta = 1/k_\mathrm{B}\mathrm{T}$ with k_B the Boltzmann constant and $\mathrm{N}!\ h^{3\mathrm{N}}\ \mathrm{Z}$ is a normalization constant. To be more precise, particles would either follow Bose–Einstein statistics for bosons (particles with a full integer spin, such as photons) or the Fermi–Dirac statistics for fermions (particles with half-integer spin, such as electrons or protons). But as both Bose-Einstein and Fermi-Dirac statistics reduce to the Maxwell-Boltzmann distribution when the temperature is high enough, we will use this distribution instead.

We define the state of a system by the values taken by all the positions $\boldsymbol{r}_i$ and all the momentum $\boldsymbol{p}_i$ of all the N atoms in the system. The state of the system is then defined by 6N variables, or a point in a vector space of 6N dimensions called the *phase space*. In order to compute the energy of a state, we will describe the interactions between the atoms by a potential energy $\mathrm{U}(\boldsymbol{r}^\mathrm{N})$, with no explicit dependency on time. Then, we can compute the total energy of a state using the classical Hamiltonian of the system:

$$\mathrm{H}(\boldsymbol{r}^\mathrm{N}, \boldsymbol{p}^\mathrm{N}) = \sum_i^\mathrm{N} \frac{\boldsymbol{p}_i^2}{2m} + \mathrm{U}(\boldsymbol{r}^\mathrm{N}); \tag{3.2}$$

where I use $\boldsymbol{r}^\mathrm{N}$ and $\boldsymbol{p}^\mathrm{N}$ as shorthand for the set of all positions $\{\boldsymbol{r}_i\}$ and momentum $\{\boldsymbol{p}_i\}$ respectively.

The last element in equation (3.1) we need to compute is the so-called *partition function* Z. We note that the probability for the system to be anywhere in the phase space Φ should be 1, which gives us:

$$\iint_\Phi \mathcal{P}_i\ \mathrm{d}\boldsymbol{r}^\mathrm{N}\mathrm{d}\boldsymbol{p}^\mathrm{N} = 1. \tag{3.3}$$

And finally:

$$\mathrm{Z} = \frac{1}{\mathrm{N}!\ h^{3\mathrm{N}}} \iint_\Phi e^{-\beta\,\mathrm{H}(\boldsymbol{r}^\mathrm{N}, \boldsymbol{p}^\mathrm{N})}\ \mathrm{d}\boldsymbol{r}^\mathrm{N}\mathrm{d}\boldsymbol{p}^\mathrm{N} \tag{3.4}$$

where the Planck constant h is used as a normalization factor used to make sure that Z has the right dimension, and the N! factor comes from the fact that particles are not distinguishable one from another.

We can already compute at least a part of this integral by separating the kinetic and potential energy terms in the Hamiltonian:

$$\mathrm{Z} = \frac{1}{\mathrm{N}!\ h^{3\mathrm{N}}} \left(\prod_i^{3\mathrm{N}} \int e^{-\beta\,\boldsymbol{p}_i^2/2m}\ \mathrm{d}\boldsymbol{p}_i \right) \left(\iiint_\mathrm{V} e^{-\beta\,\mathrm{U}(\boldsymbol{r}^\mathrm{N})}\ \mathrm{d}\boldsymbol{r}^\mathrm{N} \right) \tag{3.5}$$

where the potential energy integral is over all the accessible volume. The kinetic energy term is a product of Gaussian integrals, and gives us the following expression for the partition function:

$$\mathrm{Z} = \frac{1}{\mathrm{N}!} \prod_{i}^{3\mathrm{N}} \sqrt{\frac{2\pi m}{\beta h^2}} \iiint_{\mathrm{V}} e^{-\beta\,\mathrm{U}(\boldsymbol{r}^{\mathrm{N}})}\, \mathrm{d}\boldsymbol{r}^{\mathrm{N}} \tag{3.6}$$

$\lambda = \sqrt{\beta h^2 / 2\pi m}$ is the de Broglie thermal wavelength for a particle with mass m, and is homogeneous to a distance. $\Lambda = \lambda^3$ is thus homogeneous to a volume. This gives the final expression for the partition function:

$$\mathrm{Z} = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}} \iiint_{\mathrm{V}} e^{-\beta\,\mathrm{U}(\boldsymbol{r}^{\mathrm{N}})}\, \mathrm{d}\boldsymbol{r}^{\mathrm{N}} \tag{3.7}$$

And the corresponding probability for the system to be in a given conformation without constrains on the kinetic energy:

$$\mathcal{P}_i = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}\,\mathrm{Z}}\, e^{-\beta\,\mathrm{U}(\boldsymbol{r}^{\mathrm{N}})} \tag{3.8}$$

Thermodynamic quantities from the partition function

It is possible to use the knowledge of the partition function to compute some of the macroscopic properties of our system. For examples, the internal energy is the average value of the Hamiltonian:

$$\mathcal{U} = \frac{1}{\mathrm{Z}} \iiint_{\Phi} \mathrm{H}\, e^{-\beta\,\mathrm{H}} \tag{3.9}$$

If we express $\mathrm{H}e^{-\beta\,\mathrm{H}}$ as the partial derivative of $e^{-\beta\,\mathrm{H}}$ with respect to β we get

$$\mathcal{U} = -\frac{1}{\mathrm{Z}} \frac{\partial}{\partial \beta} \iiint_{\Phi} e^{-\beta\,\mathrm{H}} \tag{3.10}$$

$$\mathcal{U} = -\frac{1}{\mathrm{Z}} \frac{\partial \mathrm{Z}}{\partial \beta} \tag{3.11}$$

The entropy of a system can be computed from the probability density $\mathcal{P}$, using the relation:

$$\mathrm{S} = -k_{\mathrm{B}} \iint_{\Phi} \mathcal{P} \ln \mathcal{P}, \tag{3.12}$$

which reduces after some calculations to

$$\mathrm{S} = k_{\mathrm{B}} \left[\ln \mathrm{Z} - \frac{\beta}{\mathrm{Z}} \frac{\partial \mathrm{Z}}{\partial \beta} \right]. \tag{3.13}$$

The free energy definition $\mathrm{F} = \mathcal{U} - \mathrm{TS}$ then gives us:

$$\mathrm{F} = -\frac{1}{\beta} \ln \mathrm{Z} \tag{3.14}$$

Knowing the free energy, we can use all the relations from classical thermodynamics to compute some other properties of the system:

$$\mathrm{P} = -\frac{\partial \mathrm{F}}{\partial \mathrm{V}} \qquad \mu = -\frac{\partial \mathrm{F}}{\partial n} \tag{3.15}$$

Observables

The probability for the system to be in a given state gives us the missing link between microscopic and macroscopic properties of the system. We can express the value of a macroscopic observable property A using Maxwell-Boltzmann statistics:

$$\mathrm{A} = \langle \mathrm{A} \rangle = \iiint_{\Phi} \mathcal{P}_i \, \mathrm{A}_i. \tag{3.16}$$

The value of A at a macroscopic level is the same value as the ensemble average $\langle \mathrm{A} \rangle$, which depends on both the value of the property in a given macroscopic state A_i, and the probability of the system to be in this state. Using equations (3.1) and (3.7) together, we can express the average value for any observable property in the canonical ensemble:

$$\langle \mathrm{A} \rangle = \frac{\iiint_{\Phi} \mathrm{d}\boldsymbol{r}^{\mathrm{N}} \mathrm{d}\boldsymbol{p}^{\mathrm{N}} \; \mathrm{A}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}}) \; e^{-\beta \, \mathrm{H}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}})}}{\iiint_{\Phi} \mathrm{d}\boldsymbol{r}^{\mathrm{N}} \mathrm{d}\boldsymbol{p}^{\mathrm{N}} \; e^{-\beta \, \mathrm{H}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}})}}. \tag{3.17}$$

Sampling

This theoretical approach to define macroscopic properties from microscopic data is useless unless we can compute the integrals over the whole phase space Φ in (3.17). But computing this integral explicitly in all but the simplest cases will prove difficult, as the phase space is a 6N dimensional vector space, and values for N range from a few hundred all the way up to Avogadro number. But in general, multiple states in the phase space are not relevant when computing the integral, mainly because their energy is too high and their probability becomes negligible. So instead of computing the whole integral, we resort to only using a finite number of samples in the phase space, which we try to pick as the most relevant. In a semi-formal manner, we try to generate a set of points φ inside the phase space, such that

$$\langle \mathrm{A} \rangle \approx \frac{\sum_{\varphi} \mathrm{A}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}}) e^{-\beta \, \mathrm{H}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}})}}{\sum_{\varphi} e^{-\beta \, \mathrm{H}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}})}}. \tag{3.18}$$

This is the underlying idea of molecular simulation, *i.e.* the numerical sampling of phase space based on the knowledge of a way to calculate the energy of each state. There are a few algorithms we can use to do this sampling and generate the set φ of points we will use to compute a given property. I will discuss two of them below: the Metropolis Monte Carlo (MC) method, and Molecular Dynamics (MD). If we can get these algorithms to generate a set of states in the phase space according to the Maxwell-Boltzmann probability, with the same state appearing possibly more than once in the set, we can simplify the calculation of ensemble average of observables even further. For a set of m physically representative states indexed by α, the average reads:

$$\langle \mathrm{A} \rangle \approx \frac{1}{m} \sum_{\alpha}^{m} \mathrm{A}(\boldsymbol{r}_{\alpha}^{\mathrm{N}}, \boldsymbol{p}_{\alpha}^{\mathrm{N}}). \tag{3.19}$$

3.1.2 Thermodynamic ensembles

Until now, all the calculations were done in the canonical or NVT ensemble, following the Maxwell-Boltzmann distribution. It is possible to show that in other thermodynamic ensembles one can write a similar probability distribution for the phase space. The partition function defined by the normalization of these probability distributions can be used to compute all of the properties of the system, and in particular the associated thermodynamic potential.

Isothermal-isobaric ensemble

In the NPT ensemble, the volume V is a free variable, and the probability for the system to be in a given configuration is:

$$\mathcal{P}_{\mathrm{NPT}} = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}\,\Delta} \quad e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) + \mathrm{PV}\right]} \tag{3.20}$$

$$\Delta = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}} \int \mathrm{dV} \iiint_{\mathrm{V}} \mathrm{d}\boldsymbol{r}^{\mathrm{N}}\; e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) + \mathrm{PV}\right]} \tag{3.21}$$

And the free energy is given by:

$$\mathrm{G} = -\frac{1}{\beta} \ln \Delta. \tag{3.22}$$

Grand canonical ensemble

In the μVT ensemble, the number of atoms n_i can vary while the associated chemical potential μ_i is fixed. The probability for the system to be in a state is given by:

$$\mathcal{P}_{\mu\mathrm{VT}} = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}\,\Theta} \quad e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) - \sum_i \mu_i n_i\right]} \tag{3.23}$$

$$\Theta = \sum_{\substack{\mathrm{N}=0 \\ \mathrm{N}=n_1+n_2+\ldots}}^{\infty} \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}} \iiint_{\mathrm{V}} \mathrm{d}\boldsymbol{r}^{\mathrm{N}}\; e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) - \sum_i \mu_i n_i\right]} \tag{3.24}$$

Osmotic ensemble

In the osmotic ($\mathrm{N}_{\mathrm{host}}$ μ PT) ensemble, the number of atoms n_i of guest molecules and the volume V can vary while the associated chemical potential and pressure are constants. The probability for the system to be in a state is given by:

$$\mathcal{P}_{\mu\mathrm{VT}} = \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}\,\zeta} \quad e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) + \mathrm{PV} - \sum_i \mu_i n_i\right]} \tag{3.25}$$

$$\zeta = \sum_{\substack{\mathrm{N}=\mathrm{N}_{\mathrm{host}} \\ \mathrm{N}=n_1+n_2+\ldots}}^{\infty} \frac{1}{\mathrm{N}!\,\Lambda^{\mathrm{N}}} \int \mathrm{dV} \iiint_{\mathrm{V}} \mathrm{d}\boldsymbol{r}^{\mathrm{N}}\; e^{-\beta\left[\mathrm{U}(\boldsymbol{r}^{\mathrm{N}}) + \mathrm{PV} - \sum_i \mu_i n_i\right]} \tag{3.26}$$

Here, N is the total number of atoms in the system, counting both the fixed host atoms and the varying atoms of the guest. The $\sum_i \mu_i n_i$ sum only run on the guest species.

3.2 Computing energy of a molecular system

Before we can compute properties of a system using equation (3.19), we need to be able to compute the energy $U(\boldsymbol{r}^N)$ associated with any configuration of the system. I will describe the two main approaches used to do so in this section.

3.2.1 Quantum calculations

The most generic way to compute the energy of a configuration of a system is to solve the Schrödinger equation for the N electrons in the system evolving in the potential created by the M nuclei; given here in atomic units:

$$\left[-\frac{1}{2}\sum_{i}^{N}\nabla_i^2 - \sum_{i}^{N}\sum_{j}^{M}\frac{Z_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} + \sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}\right]\psi(r) = E\psi(r). \tag{3.27}$$

There is no analytic solution for this equation, and the high dimensionality of the solution space ($\approx$ 3N) make numeric resolution very difficult. Instead, we use approximated methods to solve the equation, such as Quantum Monte Carlo, Hartree-Fock and post Hartree-Fock methods, or the Density Functional Theory (DFT). Of these methods I will only present DFT, as it is the most widely used today, and particularly suitable for the study of systems with hundreds of atoms and periodic boundary conditions. The central idea of DFT is to solve these equations in terms of the total electronic density $n(\boldsymbol{r})$, and then write the total energy of the system as a functional of the density $E[n]$. I will explain in more details in section 4.1 how DFT relates to the Schrödinger equation, and how minimizing the energy functional gives us the energy of the system in the ground state.

Even when using DFT, solving the Schrödinger equation is costly in term of computing time, and imposes a limit on both the time and length scale of systems we can study. As of 2019, using standard high performance computing cluster, we can use DFT to study systems containing up to 1000 atoms on a time scale of up to 100 ps. While these numbers increased a lot in recent years due to improvements in software used for DFT and in computing hardware, they are still many systems of interest that we cannot study with DFT. If the electronic density does not vary much across the subspace of phase space we are interested in — *i.e.* no bond creation or breakage; no charge transfer — then using classical potentials or *force fields* can be a good approximation of the real energy.

3.2.2 Classical force fields

A force field is an educated guess on the functional form of the energy of a system, decomposed as a sum of simple terms with physical meanings. It is a classical and empirical approximation of the real, quantum potential energy surface. The usual decomposition is the following:

$$V(\boldsymbol{r}) = \sum_i\sum_j V^{ij}_{\text{pairs}} + \sum_\alpha V^{\alpha}_{\text{molecular}} + V_{\text{coulomb}} \tag{3.28}$$

where V_{coulomb} is the Coulombic interaction between charged atoms, $V_{\text{molecular}}$ represent the internal molecular energy and V_{pairs} represent the non-bonded pairs interactions. These terms are often broken down even further. For example, the simplest form possible for V_{coulomb} is to use fixed point charges attributed to each atom during the force field parametrization. If this is not enough to reproduce the properties of the system of

interest, we can use diffuse Gaussian charges instead, or add a term describing atomic polarization. V_{pairs} is used to reproduce both the dispersion interactions and the Pauli repulsion between atoms at short distances. $V_{molecular}$ represent the intra-molecular interactions, and is usually decomposed over bonded coefficients:

$$V_{molecular} = \sum_{bonds} V_{bond}(r) + \sum_{angles} V_{angle}(\theta) + \sum_{dihedrals} V_{dihedral}(\varphi) \tag{3.29}$$

Each energy term only depends on the type of bonded atoms, and a single scalar variable: the distance r between the two atoms for bonds contributions; the 3-body angle θ for angles contributions, and the 4-body dihedral angle φ or out of plane distance d for dihedral angles contributions. These variables are illustrated in figure 3.1.

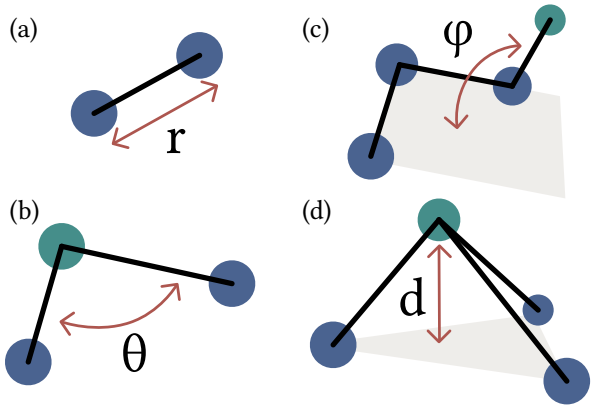


Figure 3.1 – Definition of the parameters used to compute the energy of a molecular system with classical force field. (a) bonds; (b) angles; (c) dihedral angles; (d) improper dihedrals/out of plane distance.

Typical functional forms used in force fields

There is no strict rule regarding which functional form can or cannot be used in a force field, so when creating a new force field we usually rely on chemical sense as well as few physical laws to pick them. Bellow, I describe a few well-known terms used in most force fields.

Charges

If we choose to model atomic partial charges as point charges, we can directly use the expression for Coulombic interactions, as a sum over all pairs of charged atoms in the system:

$$V(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}. \tag{3.30}$$

It is also possible to use Gaussian charges distributions on each atom, where the charge density around an atom is defined by a width parameter α_i:

$$\rho_i = q_i \left(\frac{\alpha_i}{\sqrt{\pi}}\right)^3 e^{-\alpha_i^2 r_i^2}. \tag{3.31}$$

In that case, the interaction between two atoms is given by — using erf for the error function:

$$V(r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \operatorname{erf}\left(\sqrt{\frac{\alpha_i^2\alpha_j^2}{\alpha_i^2+\alpha_j^2}}\, r_{ij}\right). \tag{3.32}$$

I will discuss the issues that arise from using these type of potentials with periodic boundary conditions, and the possible way to fix them such as Ewald summation in section 6.3.

Non-bonded pairs interactions

As we have seen, we use non-bonded pairs interactions to reproduce both the Pauli repulsion between atoms at short distances, and the dispersion attractive interaction at long distances. It has been shown[93] that the dispersion interaction can be developed in the long distances approximation as:

$$V_{\text{dispersion}} = -\frac{C_6}{r^6} + \frac{C_8}{r^8} - \frac{C_{10}}{r^{10}} + \mathcal{O}\left(\frac{1}{r^{12}}\right), \tag{3.33}$$

where the C_i coefficients have positive values. Most of the time, only the term in $1/r^6$ is used, as it will have the largest contribution to the resulting energy.

There is however no simple mathematical expression for the Pauli repulsion at short distances, so various schemes have been used to approximate it. The most prevalent one is the Lennard-Jones potential, which uses a repulsive term proportional to $1/r^{12}$. In the early day of molecular simulation, this allowed to save some computing time by squaring the already-computed $1/r^6$ term.

$$V_{\text{Lennard-Jones}}(r) = 4\,\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{3.34}$$

Here, ε is the maximal strength of the interaction, and σ the minima distance at which the interaction becomes repulsive, often representing the radius of spherical atoms.

Another commonly used form is the Buckingham potential, using an exponential function for the repulsion:

$$V_{\text{Buckingham}}(r) = A\, e^{-Br} - \frac{C}{r^6} \tag{3.35}$$

Molecular interactions

The most common strategy for describing bonds and angles is to consider only vibration of the bond length or the angle around the equilibrium, and represent the energy of the bond/angle using the harmonic approximation:

$$V_{\text{harmonic}}(r) = \frac{1}{2}k\,(r-r_0)^2 \qquad \text{or} \qquad V_{\text{harmonic}}(\theta) = \frac{1}{2}k\,(\theta-\theta_0)^2 \tag{3.36}$$

For dihedral angles, we often want to be able to reproduce the periodicity of the associated energy, which leads to the following definition of the energy:

$$V_{\text{torsion}}(\varphi) = \frac{1}{2}E_0\left[1+\cos(n\varphi+\delta)\right] \tag{3.37}$$

Force field parametrization

All these functional forms have one or more adjustable parameters, that depends on the types of the atoms participating in the pair; bond; or angle. For example, when using Coulombic interactions, the charge carried by an atom is the adjustable parameter; when using a Lennard-Jones potential both the values of σ and ε are adjustable.

The process of adjusting these parameters to make sure the force field produces the correct energy and physical properties of the system is called the parametrization of a force field. It usually involves a trade-off between accuracy — *i.e.* how well the force field can reproduce the potential energy surface — and transferability. A force field is said to be transferable if we can use the same set of parameters for different systems. I will discuss parametrization of force field in more details in the next chapter, section 4.3.

3.2.3 System size and periodic boundary conditions

Everyday chemical systems contain a huge number of atoms, of the order of Avogadro's number (10^{23} atoms). But when using molecular simulation to study a chemical system, we are limited to a much smaller number of atoms: using 1 TB of computer memory we can only store positions and velocities for around 10^{10} atoms. This is even worse if we consider the number of atoms for which we can compute properties with molecular simulation methods, of the order of 10^4 to 10^7 atoms at most from today commodity hardware to the biggest super-computers.

At these scales, the size and surface effects are not negligible: the surface energy becomes an important part of the overall energy. The typical way to remove these surface effects while keeping the number of simulated atoms low is to use *periodic boundary conditions*. The atoms are placed in a simulation box called the *unit cell*, which is replicated *ad infinitum* in all directions, as illustrated in figure 3.2. The copies of the initial box are called images of this box.

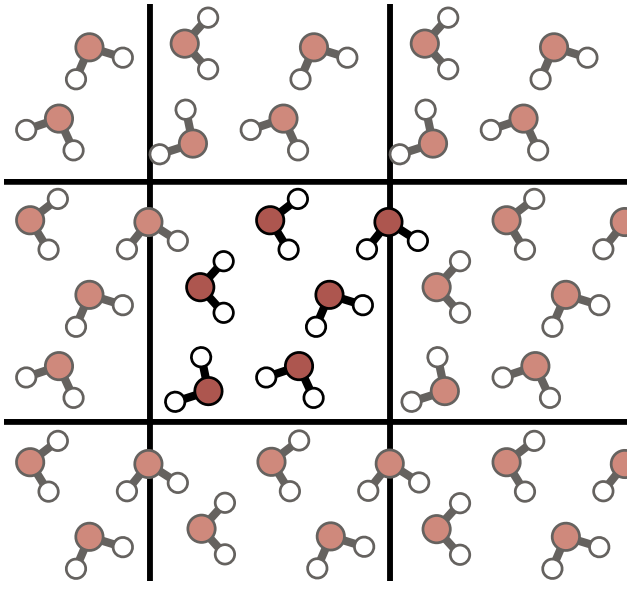


Figure 3.2 – Illustration of periodic boundary condition in a two dimensional system. Molecules from the central cell are repeated in all directions.

In this arrangement, a molecule from the central simulation box will interact with all molecules in the same box; but also with any molecule in any image box, including with molecules image of itself. The effectively infinite system created this way will not contain any surface effects, but other issues can still affect the sampling and averaging of system properties.

First of all, the use of periodic boundary conditions introduces an artificial periodicity into the system, which grows more important as the central simulation box is smaller. For the study of crystalline materials, one has to ensure that the periodicity of the simulation box matches the periodicity of the crystal primitive cell. This periodicity also makes it harder to study defects and other statistically rare features of the system. If we explicitly add a defect to the system, it will be replicated over all the images, artificially increasing the number of defects per unit of volume.

Second, because the system is now infinite, we need to compute an infinity of interactions between molecules to evaluate its energy. Fortunately, most interactions decay at long distances, and we can use a cutoff radius when computing the energy. Any atoms further apart than this cutoff radius will not interact. The error ε that arise from the use of a cutoff radius r_c depends on the potential $\mathrm{V}(r)$ and the radial distribution function $g(r)$ of the current conformation:

$$\varepsilon(r_c) = \int_{r_c}^{\infty} r^2 \mathrm{V}(r)\, g(r)\, \mathrm{d}r \tag{3.38}$$

If the cutoff radius is large enough (around 10 Å is usually enough), then $g(r) \simeq 1$ and we get a simpler estimation that does not depends on the system conformation:

$$\varepsilon(r_c) = \int_{r_c}^{\infty} r^2 \mathrm{V}(r)\, \mathrm{d}r \tag{3.39}$$

For any function $\mathrm{V}(r)$ that goes to zero at infinity faster than $1/r^3$, this error is finite, and can be computed and applied afterward. Notably, the electrostatic potential decays as $1/r$, and this error does not converge. In order to describe electrostatic interactions properly in presence of periodic boundary conditions, we need to use other methods such as the Ewald summation, which I will present in more detail in section 6.3.

Finally, the *minimum-image convention* is often used to improve simulation speed. Under this convention, a particle in the central simulation box interacts only with neighbors from the central image or an image surrounding the central one. This restricts the number of images that has to be searched for neighbors inside the cutoff radius. To be able to enforce this minimum-image convention, the smallest inscribed sphere must have a radius bigger than the cutoff radius. For an orthorhombic simulation box, this means that the smallest box side length must be at least twice the cutoff radius.

3.3 Metropolis Monte Carlo

Now that we know how to compute the energy of any given configuration, our goal of computing macroscopic properties from microscopic states is getting closer. The idea is to evaluate the integral in equation (3.17) by using a finite set of configurations $\{\boldsymbol{r}^N\}_\alpha$ distributed according to the right distribution, which depends on the ensemble. The ensemble average $\langle A \rangle$ of a property A that only depends on the spatial conformation of the system gives us the value of this property at the macroscopic level:

$$\langle A \rangle \approx \frac{1}{m} \sum_{\alpha}^{m} A(\boldsymbol{r}_\alpha^N) \tag{3.40}$$

Metropolis Monte Carlo is an algorithm that enables us to generate new configurations following a given distribution, using only an initial configuration of the system. This means that we can directly sample any ensemble of interest.

3.3.1 The basic algorithm

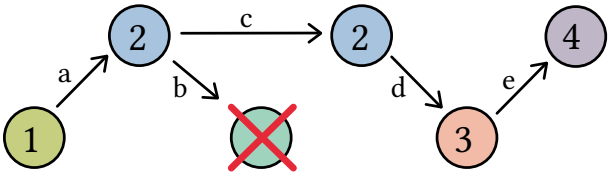


Figure 3.3 – Illustration of a Markov chain in Metropolis Monte Carlo. Configurations are numbered from 1 to 4, and letters from (a) to (e) are used for tentative moves. Step 2 appears twice in the chain, as the tentative move (b) was rejected.

The Metropolis Monte Carlo method is part of the family of Markov Chain Monte Carlo algorithms. A Markov chain is a set of configurations, such that the probability of generating a new configuration only depends on the previous configuration of the chain: it has no further *memory* of older states. A Markov chain is thus completely defined by the knowledge of the initial state and the procedure used to generate a new state.

When using a Markov chain to explore and sample the phase space, we want it to be able to explore all the possible states in the phase space. In order to ensure this, we want the chain to be *ergodic*, meaning that each state can be reached from any other state in a finite number of steps. This property guarantees the existence of a *stationary* equilibrium distribution of the generated states as the Markov chain grows, and that this distribution matches the probability to generate a new state from any given state. For us, this means that if we are able to create an ergodic Markov chain using the Boltzmann statistics to generate a new state, then the whole chain will follow the canonical ensemble distribution. The same reasoning also works in other statistical ensembles, following their own distribution of probability.

Micro-reversibility

The usual way to ensure we have an ergodic Markov chain is to impose that this chain is *micro-reversible*. Although this is not a required condition to have an ergodic chain, it is a sufficient one[92], and quite commonly used. Using $\mathcal{P}_i$ for the probability of being in a state i, and$\pi(i \to j)$ the probability to go from state i to state j, the condition of micro-reversibility is defined as:

$$\mathcal{P}_i \, \pi(i \to j) = \mathcal{P}_j \, \pi(j \to i) \tag{3.41}$$

This means that the probability for any transition in the Markov chain to be from state i to state j must be the same as the probability for this transition to be from state j to state i. In Metropolis Monte Carlo, the generation of a new state is a two-step process: first we generate a new configuration, and then we accept or reject this new configuration. If we accept the configuration, then the new state of the Markov chain is the new configuration; else the new state of the chain is the old configuration. This is illustrated in figure 3.3. This makes the probability $\pi(i \to j)$ a product of the probability $\alpha(i \to j)$ to generate a given new configuration and the probability $\text{acc}(i \to j)$ to accept it.

$$\pi(i \to j) = \alpha(i \to j) \times \text{acc}(i \to j) \tag{3.42}$$

The original Metropolis scheme[94] chooses the α probability to be symmetric, *i.e.* $\alpha(i \to j) = \alpha(j \to i)$, the probability of generating a configuration j from i is the same as the probability of generating the configuration i starting from j. This gives us

$$\frac{\text{acc}(i \to j)}{\text{acc}(j \to i)} = \frac{\mathcal{P}_j}{\mathcal{P}_i} \tag{3.43}$$

We want to set the resulting probability of the Markov chain to follow Boltzmann statistics, which gives us

$$\frac{\text{acc}(i \to j)}{\text{acc}(j \to i)} = e^{-\beta[\text{U}(j) - \text{U}(i)]} \tag{3.44}$$

There are multiple choices that would result in this same relation. The standard choice is to always accept the transition from i to j if the energy decreases ($\Delta\text{U} = \text{U}(j) - \text{U}(i) < 0$), or else accept it with probability $e^{-\beta\Delta\text{U}}$.

$$\text{acc}(i \to j) = \min\left[1, \exp\left(-\beta\Delta\text{U}\right)\right] \tag{3.45}$$

In order to accept a configuration change with probability $e^{-\beta\Delta\text{U}}$, we usually generate a random number r with uniform distribution in $[0, 1]$, and accept the new configuration if $r < e^{-\beta\Delta\text{U}}$.

When working in ensembles other than the canonical ensemble, we can still apply the same procedure, simply changing the acceptance probability to agree with the required distribution. For example, in the NPT ensemble, the probability to be in a state is proportional to $e^{-\beta\,\text{U}(\boldsymbol{r}) \,-\, \beta\text{PV}}$. This changes the acceptance probability to

$$\text{acc}^{\text{NPT}}(i \to j) = \min\left[1, \exp\left(-\beta\Delta\text{U} - \beta\text{P}\Delta\text{V}\right)\right] \tag{3.46}$$

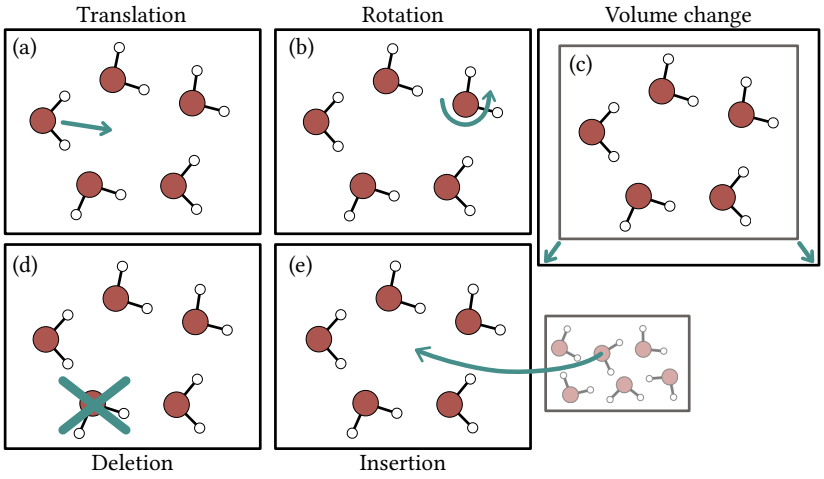


Figure 3.4 – Basic Monte Carlo moves: (a) translation of a single molecule; (b) rotation of a single molecule; (c) unit cell volume change; (d) molecule deletion; (e) molecule insertion from a reservoir.

3.3.2 Monte Carlo moves

The last remaining piece of the puzzle before being able to run Monte Carlo simulations is to specify a way to generate new configurations. We need to construct them such that the probability of creating a state j from i is the same as creating the state i starting from j; *i.e.* $\alpha(i \to j) = \alpha(j \to i)$. In implementations of Metropolis Monte Carlo for molecular simulation, basic algorithms used to generate a new conformation from the previous one are called *moves*. At each step of the simulation, a specific move is selected at random in the list of possible moves, ensuring $\alpha(i \to j) = \alpha(j \to i)$ globally as long as the move itself enforce $\alpha(i \to j) = \alpha(j \to i)$.

Canonical ensemble: translations and rotations

The simplest move imaginable is to randomly select an atom and move it by a random distance in a random direction. The newly generated state only differs from the previous one by the position of this atom. When working with rigid molecules, this move can also be used to translate a full molecule. But in that case, we also need to account for rotational degrees of freedom. We use random rotations (random rotation axis and magnitude) of molecules to do so. These moves are illustrated in figure 3.4, panels (a) and (b). By using uniform distributions for the cartesian components of the translation, the amplitude of the rotation, and uniform sampling of the unit sphere for the rotation axis; these moves have the same probability of going from a state i to j and vice versa.

The one remaining parameter is then the amplitude of the move, *i.e.* the interval in which to draw the translation or rotation amplitude with a uniform distribution. It is desirable to have an acceptance rate around 40 to 50%, as a lower acceptance rate means that we are wasting computation time to generate conformations that will not be used, and a higher acceptance rate usually means that the new conformation will be very close to the previous one, and the simulation will take longer to sample the phase space. We can adjust the amplitude of the rotations and translations to influence the value of $\alpha(i \to j)$: larger amplitude means a smaller acceptance rate, and smaller amplitude means a higher acceptance rate. But adjusting the amplitude breaks micro-reversibility,

and as such should be done beforehand to prevent affecting the simulation results. It is thus customary to break a simulation in two phases: a first equilibration phase to adjust the simulation parameters, and a later production phase that will be used for statistical analysis and averaging.

Translations and rotations, together with the acceptance scheme defined in equation (3.45), are enough to generate a Markov chain sampling the NVT ensemble. The temperature of the system is set by the value of β used in equation (3.45), and the number of particles and volume are set by the initial conformation. The conformations in the Markov chain can then be used with equation (3.40) to evaluate the macroscopic properties of the system.

Isobaric-isothermal ensemble: volume changes

If we want to work in another thermodynamic ensemble such as the NPT ensemble, we need first to change the acceptance scheme to equation (3.46), and also to generate moves that change the system's volume. One way to change the volume is to randomly pick a new volume using a uniform distribution centered around the old volume, and then resize the simulation unit cell while keeping intra-molecular distances constant. This is illustrated in figure 3.4, panels (c). Keeping the distances constant ensures that the change in energy ΔU will not be too big, and that the move will be accepted more often than with a simple scaling scheme. As for translations and rotations, we can adjust the acceptance rate of the move by changing the amplitude of volume changes.

Grand Canonical ensemble: insertion and deletion

Monte Carlo simulations in the μVT ensemble are called Grand Canonical Monte Carlo (GCMC) simulations. In this ensemble, the number of molecules is allowed to fluctuate in order to maintain the chemical potential constant. We need two moves to accomplish this, one where we select a molecule at random and remove it, and another one where a molecule is inserted from a fictitious reservoir at a random position and orientation (see figure 3.4, panels (d) and (e)). The acceptance probability is given as previously by the $\mathcal{P}_j/\mathcal{P}_i$ ratio; but this time the N! Λ^{N} terms from equation (3.25) do not cancel out. For the insertion of a new molecule we have:

$$\mathrm{acc}(\mathrm{N} \to \mathrm{N}+1) = \min\left[1, \frac{\mathrm{V}}{\Lambda(\mathrm{N}+1)} \exp\left(\beta\mu - \beta\Delta\mathrm{U}\right)\right]; \tag{3.47}$$

and for the removal of a molecule,

$$\mathrm{acc}(\mathrm{N} \to \mathrm{N}-1) = \min\left[1, \frac{\Lambda\mathrm{N}}{\mathrm{V}} \exp\left(-\beta\mu - \beta\Delta\mathrm{U}\right)\right]. \tag{3.48}$$

Unfortunately, the chemical potential μ is not easily known before starting a simulation and these expressions are hard to use. Instead, for chemical species in the gas phase we use the fugacity f of the gas, defined as the pressure of an ideal gas with the same temperature and molar free enthalpy as the real gas. It is related to the chemical potential

by $f = e^{\beta\mu}/\beta\Lambda$. Replacing in the equation above gives us the standard acceptance probabilities for GCMC simulations in the gas phase:

$$\text{acc}(\text{N} \rightarrow \text{N}+1) = \min\left[1, \frac{\beta \text{V} f}{\text{N}+1} \exp\left(-\beta\Delta\text{U}\right)\right] \tag{3.49}$$

$$\text{acc}(\text{N} \rightarrow \text{N}-1) = \min\left[1, \frac{\text{N}}{\beta \text{V} f} \exp\left(-\beta\Delta\text{U}\right)\right] \tag{3.50}$$

For other phases such as liquids, the relation between chemical potential and pressure $\mu(\text{P})$ for the bulk phase has to be established first, by separate series of Monte Carlo simulations.

3.3.3 Monte Carlo caveats

Monte Carlo simulations are very powerful, as they allow to directly sample the phase space of the ensemble of interest, and do not require that the moves used to generate a new trial configuration follow the laws of physics. This later property is used in some extensions of Monte Carlo simulations, where an atom identity is transformed as a Monte Carlo move, changing a sodium atom into a potassium atom for example. This method — called semi-grand canonical Monte Carlo[95] — allows to sample the equilibrium distribution during chemical reactions or at an interface without explicitly representing the interface.

They also have a few caveats, mainly it is not possible to sample the microcanonical ensemble, as this would require generating trial configurations with constant energy, and updating both the positions and velocities of the atoms. The main problem with Monte Carlo simulation is that they only deal with the conformations of the system, and do not contain any information on the kinetics of the transformations. Both of these caveats can be solved by using molecular dynamics instead to generate the states used to evaluate equation (3.17).

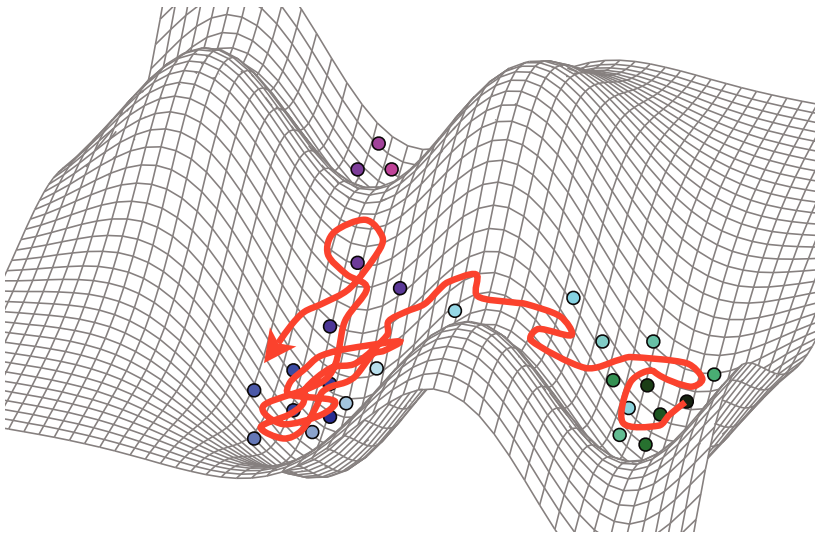


Figure 3.5 – Representation of the way the potential energy surface is explored by molecular dynamics (red line) and Monte Carlo (colored circles) simulations.

3.4 Molecular dynamics

The fundamental idea of molecular dynamics is to use Newton's laws of motion to integrate the positions and velocities of the atoms in the system as a function of time. In this work, I will consider that the atoms in the system behave as classical point particles. There are multiple methods to go beyond this approximation and incorporate the quantum nature of atoms, mostly based on Feynman path integrals[96].

Newton's laws give us differential equations on the time evolution of positions and velocities ($\boldsymbol{p}_i = m_i \boldsymbol{v}_i$):

$$\frac{\partial \boldsymbol{r}_i}{\partial t} = \boldsymbol{v}_i; \qquad \text{and} \qquad m_i \frac{\partial \boldsymbol{v}_i}{\partial t} = -\frac{\partial \mathrm{U}(\boldsymbol{r})}{\partial \boldsymbol{r}_i}. \tag{3.51}$$

We can discretize time in these equations by considering a succession of instants separated by the *time step* δt. If we then replace the derivatives with first order finite differences, we get the Euler integrator equations:

$$\boldsymbol{r}_i(t + \delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t)\delta t + \mathcal{O}(\delta t^2) \tag{3.52}$$

$$\boldsymbol{v}_i(t + \delta t) = \boldsymbol{v}_i(t) + \frac{1}{m_i}\boldsymbol{f}_i(t)\delta t + \mathcal{O}(\delta t^2) \tag{3.53}$$

where $\boldsymbol{f}_i = -\partial \mathrm{U}(\boldsymbol{r})/\partial \boldsymbol{r}_i$ is the force acting on atom i. The Euler integration scheme presented here have a high discretization error and is not used in actual simulations, but helps understanding molecular dynamics. Better integration schemes will be presented below.

Starting from an initial configuration defined by the positions and velocities of each atom at time t, these equations allow to compute the positions and velocities at the next time point $t + \delta t$. For this discretization to be valid, the time step δt has to be small enough so that both the $\mathcal{O}(\delta t^2)$ terms are negligible, and the forces acting on each atom are constant between t and $t + \delta t$. Depending on the system, typical values for the time step fall between 0.5 fs and 2 fs.

Time integration of positions and velocities follows the standard rules of classical mechanics, and in particular the conservation of mechanical energy. This means that by default, molecular dynamics trajectories evolve in the microcanonical NVE ensemble. There however is no *a priori* guarantees that a molecular dynamics trajectory will generate new configurations following the microcanonical distribution. We rely on the *ergodic* hypothesis to assume that the trajectory does sample the whole NVE ensemble. In the most intuitive form, this hypothesis is formulated as:

> *At equilibrium, the ensemble average of a property is the same as the average of the values of this property taken over a high number of measurements in different points in time.*

The ergodic hypothesis links together the statistical physics view of the world, and what can effectively be measured at the macroscopic level. Mathematically speaking, the ergodic hypothesis is a property of the Hamiltonian system: a system is said to be ergodic

if the point representing this system in the phase space can approach every point in the phase space as closely as desired. In this case, the Birkhoff ergodic theorem proves that the time average and ensemble average are the same almost everywhere. This condition has been verified for some systems, and is usually assumed to hold when working with statistical thermodynamics.

3.4.1 Integration schemes

Desirable properties

We want the integration scheme used for a molecular dynamics simulation to enforce the symmetry and conservation properties of a real mechanical system; namely the time reversal symmetry, conservation of energy, linear momentum, angular momentum and phase space volume. Conservation of phase space volume is a consequence of Liouville's theorem concerning the phase space distribution $\rho(\boldsymbol{r}^N, \boldsymbol{v}^N)$:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{\partial \rho}{\partial t} + \sum_i \left(\frac{\partial \rho}{\partial \boldsymbol{r}_i} \frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}t} + \frac{\partial \rho}{\partial \boldsymbol{v}_i} \frac{\mathrm{d}\boldsymbol{v}_i}{\mathrm{d}t} \right) = 0. \tag{3.54}$$

This means that if we consider a set of initial conditions inside a closed and bounded region of phase space and propagate them for a given time, the resulting points in phase space will still be a closed and bounded region of phase space with the same volume as the initial region — this is illustrated in figure 3.6. An integrator conserving the phase space volume is said to be *symplectic*.

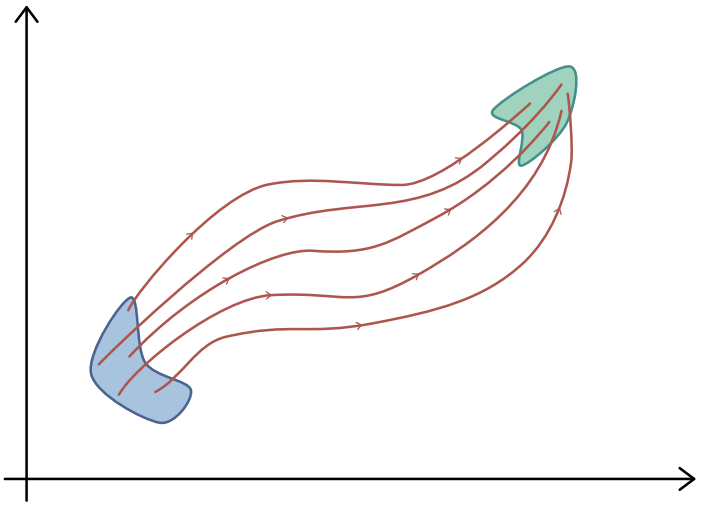


Figure 3.6 – Illustration of a set of trajectories (in red) in a 2 dimensional phase space. For a symplectic integrator, the initial volume (in blue) will be the same as the final volume (in green).

Because we are approximating the equations of motion by evaluating them on discrete time steps the energy might not be fully conserved, and an energy drift can appear. The amplitude of this drift will depend on the choice of time step; a bigger time step will introduce a bigger error. Good integrators will allow us to use bigger time steps while keeping the drift small. In turn, this makes the overall simulation more efficient by reducing the number of steps needed to simulate long processes. Another way to make the simulation more efficient is to evaluate the forces acting on the atoms as rarely

as possible. This is the case in multi-timesteps methods such as reversible Reference System Propagator Algorithms[97] (r-RESPA) where forces are separated by the time scale of their variations, and evaluated only when needed instead of being calculated at all steps.

Some examples of integration schemes

One of the simplest integrator respecting all the above defined constraints (symplectic, time-reversible, energy conservation) is the *leap frog* integrator[92]. The corresponding algorithm is reproduced below, using $\boldsymbol{r}(n)$ for $\boldsymbol{r}(n \times \delta t)$, and $\boldsymbol{f}(n)$ for the forces computed from the positions at step n.

$$\begin{aligned} \boldsymbol{a}(n) &= \frac{1}{m}\boldsymbol{f}(n), \\ \boldsymbol{v}(n+1/2) &= \boldsymbol{v}(n-1/2) + \boldsymbol{a}(n)\,\delta t, \\ \boldsymbol{r}(n+1) &= \boldsymbol{r}(n) + \boldsymbol{v}(n+1/2)\,\delta t, \end{aligned} \tag{3.55}$$

Evaluation of the forces is made clear by the $a = f/m$ line, to ensure it only occurs once. The leap frog integrator is more stable than the previously mentioned Euler integrator. Unfortunately, it computes the positions and velocities at interleaved instants, staggered in such a way that they jump over each other in a frog-like fashion.

Velocity-Verlet integrator fixes this issue while keeping the symplectic and time reversible nature of leap frog[92, 98]. The algorithm works as follow:

$$\begin{aligned} \boldsymbol{v}(n+1/2) &= \boldsymbol{v}(n) + \frac{1}{2}\boldsymbol{a}(n)\,\delta t, \\ \boldsymbol{r}(n+1) &= \boldsymbol{r}(n) + \boldsymbol{v}(n+1/2)\,\delta t, \\ \boldsymbol{a}(n+1) &= \frac{1}{m}\boldsymbol{f}(n+1), \\ \boldsymbol{v}(n+1) &= \boldsymbol{v}(n+1/2) + \frac{1}{2}\boldsymbol{a}(n+1)\,\delta t, \end{aligned} \tag{3.56}$$

Here, velocities are updated twice by a half step, but the forces are still computed only once. This algorithm also yields the positions and velocities at full time steps, allowing them to be used together when computing properties of the system.

Both integrators presented here have a $\mathcal{O}(\delta t^2)$ error, but they are other integration scheme with a smaller error, called higher-order integrators. The most famous ones are the Runge–Kutta methods, that can be derived for any even order, starting with $\mathcal{O}(\delta t^4)$. Runge–Kutta integrators are not used as often as the Velocity-Verlet or leap frog integrators, as they are more expensive to use and are not time-reversible.

3.4.2 Sampling other ensembles

By default, molecular dynamics samples the NVE ensemble since the mechanical energy, volume and number of particles are kept constant. To sample another ensemble, we need a way to control the variable that changes from this ensemble to the NVE ensemble. For example, going from NVE to NVT, we want to change the energy of the system as E and T are conjugated variables. Similarly, going from NVT to NPT requires changes to the volume.

Thermostats

Algorithms that control the temperature by changing the energy of the system are called *thermostating* algorithms, or *thermostats* for short. They work by changing the kinetic energy of the system, or more generally by changing the velocities of the atoms. As illustrated in figure 3.7, for the same system, different thermostats will produce different temperature response.

In the simplest (and crudest) approach, if we want to change the system temperature from T_0 to T_1, we need to multiply all velocities by $\sqrt{T_1/T_0}$. As the system evolves, the temperature needs to be updated again, as kinetic energy is transferred into potential energy and reciprocally by the integrator. This approach is sometimes called a *rescaling* thermostat. While this algorithm does fix the system temperature to the target value, the integrator is no longer time-reversible and symplectic. This also introduces many non-physical artifacts in the dynamics. One example of such non-physical effect is the so-called *flying ice cube*; where all the thermal energy of the system is transferred into the global translation velocity, while removing all vibrational energy from the atoms, resulting in a simulation of flying frozen atoms. Finally, the simulation does not sample the true NVT ensemble, since the kinetic energy is not allowed to fluctuate around its equilibrium value.

The *weak coupling* or Berendsen[99] thermostat is a slightly better version of the rescaling thermostat that imposes a first-order (exponential) relaxation of the temperature such that

$$\frac{\mathrm{d}T(t)}{\mathrm{d}t} = \frac{T_0 - T(t)}{\tau_T}; \tag{3.57}$$

where $T(t)$ is the instantaneous system temperature, T_0 the target external temperature, and τ_T the relaxation time constant. The instantaneous temperature is computed from the system's kinetic energy $E_{kin} = \sum_i m_i \boldsymbol{v}_i^2(t)$ as

$$T(t) = \frac{2\, E_{kin}(t)}{n_f\, k_B}, \tag{3.58}$$

where n_f is the number of degrees of freedom in the system, 3N for fully flexible molecules. The Berendsen thermostat is implemented by rescaling the velocities at each time step by

$$\alpha = \sqrt{1 + \frac{\delta t}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)}. \tag{3.59}$$

The Berendsen thermostat models a system weakly coupled to a heat bath at temperature T_0. In order to sample the canonical ensemble, we need an algorithm that is able to reproduce both the average value of the temperature and the fluctuations around the equilibrium of this temperature. Because it diminishes fluctuations of the kinetic energy of the system, the Berendsen thermostat is unable to produce trajectories in the canonical NVT ensemble[100]. It is still useful for the equilibration phase of a simulation, where one wants the system to reach approximate equilibrium as fast as possible.

To probe the correct canonical ensemble, we can instead use a Nosé-Hoover[101, 102] thermostat. In this approach, the Lagrangian (and thus the Hamiltonian) of the system is modified to include an additional variable s controlling the temperature of the system by

changing the time scale of the extended system such that $\delta t' = s\,\delta t$. This new variable is coupled to the system by changing all the velocities to $\boldsymbol{v}' = s^{-1}\,\mathrm{d}\boldsymbol{r}/\mathrm{d}t$. It can be shown that sampling the extended system in the microcanonical ensemble is equivalent to sampling the real system in the canonical ensemble, including the fluctuations. The intensity of the coupling is can be tuned by changing the fictitious mass associated with s. The extended Lagrangian can be propagated through time with any existing time-reversible and symplectic integrator.

Instead of using such extended Lagrangian methods, it is also possible to modify the rescaling methods to sample canonical ensemble. This is the basis of the Canonical Sampling through Velocity Rescaling[103] (CSVR) thermostat which I also used during my PhD. The main idea is to replace the α scaling parameter from Berendsen thermostat by a parameter picked randomly using the expected distribution for the fluctuations of kinetic energy. The trajectory will then explore the canonical ensemble, by construction.

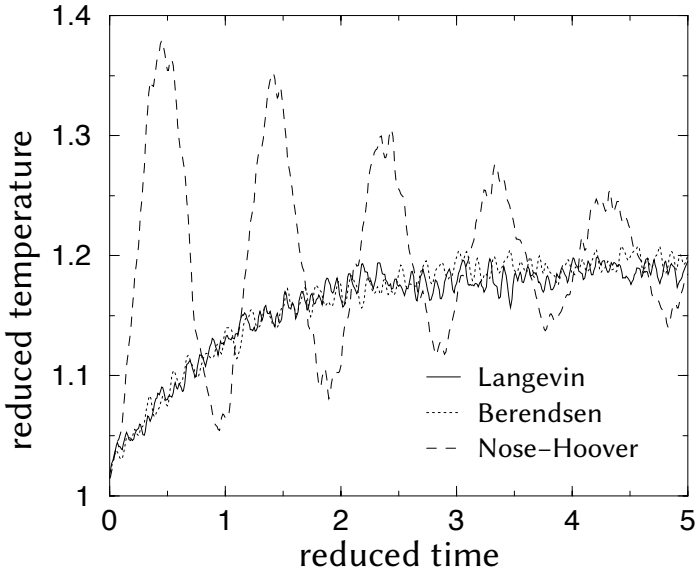


Figure 3.7 – The temperature response of a Lennard–Jones fluid under control of three thermostats (solid line: Langevin, not presented here; dotted line: Berendsen; dashed line: Nosé–Hoover) after a step change in the reference temperature. Image adapted from reference [104].

Barostats

Controlling the pressure can be done with the same approaches as the temperature to generate the NPT ensemble. In this case, we want to fix the value of the pressure P by dynamically changing the volume V of the system. For example, the Berendsen barostat[99] rescales the system volume at every step to ensure first-order relaxation to the target pressure. For a cubic unit cell subject to isotropic pressure, this means multiplying the cell volume by

$$\eta(t) = 1 - \frac{\beta\,\delta t}{\tau_\mathrm{P}}\left(\mathrm{P}_0 - \mathrm{P}(t)\right); \tag{3.60}$$

where P_0 is the target pressure, $\mathrm{P}(t)$ the instantaneous system pressure, τ_P the pressure relaxation time constant and β the compressibility of the system; often approximated by

the compressibility of liquid water. We then multiply the atomic positions by $\sqrt[3]{\eta}$. If the system is subject to an anisotropic stress tensor $\underline{\mathrm{S}}_0$, then the unit cell matrix should be resized by:

$$\underline{\eta} = \mathbb{1} - \frac{\beta\,\delta t}{\tau_\mathrm{P}}\left(\underline{\mathrm{S}}_0 - \underline{\mathrm{S}}(t)\right). \tag{3.61}$$

As for thermostats, the Berendsen barostat does not sample the actual isothermal-isobaric ensemble, as it cannot accurately reproduce the volume fluctuations of the NPT ensemble. It is still very useful for system equilibration when starting a simulation. The Nosé-Hoover barostat is based on an extended Lagrangian, and can sample the real NPT ensemble[91]. To my knowledge, there is no equivalent to the CSVR thermostat for controlling the pressure.

When using both a thermostat and a barostat to control the system pressure and temperature, one must take care that both control algorithms have sufficient time to react to changes. In particular, every time the barostat changes the simulation cell, the thermostat must change the velocities to adapt the kinetic energy to the new potential energy. This means that the time step used for the barostat should be higher than the one used by the thermostat. In my simulations, I found that setting $\tau_\mathrm{P} = 10 \times \tau_\mathrm{T} = 1000 \times \delta t$ works well in practice.

Can osmostats exist? Grand canonical molecular dynamics

As we have seen, it is possible to go from an ensemble to another by controlling the conjugated variables: change E to control T, V to control P. It is then natural to consider whether it is possible to control the chemical potential of a system to sample the grand canonical μVT potential. The obvious way to do so is to change the number of particles in the system. Multiple versions of such *Grand Canonical Molecular Dynamics* have been presented over the years, either using an extended Hamiltonian approach[105–107], or a mixture of Monte Carlo moves and time integration in the same molecular dynamics trajectory[108–110].

The latter methods — randomly inserting or removing particles within molecular dynamics integration — have not been proven, to my knowledge, to generate any specific statistical ensemble. While Monte Carlo simulations and molecular dynamics can be made to work together so under certain hypotheses, we lack such mathematical proofs for the generic hybrid case. For example, it is possible to mix molecular dynamics and Monte Carlo under the framework of Hybrid Monte Carlo simulations, which I will discuss in more details in chapter 6.2. The main idea in hybrid Monte Carlo is to use short Molecular Dynamics simulations as a new Monte Carlo move to generate new configurations. Under some conditions on the molecular dynamics integrator, the resulting Markov chain is proven to be ergodic. In this case, the simulation is a Monte Carlo simulation, and not a molecular dynamics one: molecular dynamics is only used as a mean to generate new configurations, and the underlying trajectory is thrown away. Because of this, the resulting simulation suffers from the same limitations as all Monte Carlo simulations, and in particular the absence of any dynamic information.

For the methods using an extended Hamiltonian approach, the general idea is to have one or more *fractional* particles in the system, represented by a parameter λ going from 0 to 1. The interactions of this particle with the system are scaled down by λ, and the

value of λ evolves through time. When it reaches 1, the particle is converted to a full particle and a new one is inserted with a small λ, and when it reaches 0, the particle is removed and a new one is picked to have the value of λ change.

This approach has several limitations, starting with the fact that time-reversibility is broken when λ reaches 0 or 1. Additionally, Liouville's theorem (the fact that the phase-space distribution function is constant along the trajectories of the system) is only proven for a constant number of atoms, and might not be valid in the grand canonical ensemble [111]. This can be understood intuitively: in both NVE, NVT and NPT simulations, the system explores a certain sub-space of the phase space, the hyper-surfaces at a constant energy/temperature, or the hyper-volume at constant pressure respectively. In the grand canonical ensemble, the underlying number of dimensions of the ensemble would change depending on the position in the extended phase space. As all of our current theory and understanding of Hamiltonian dynamics is based on a constant number of particles, using molecular dynamics to sample the grand canonical ensemble might not be possible. Fortunately, we can still use Grand Canonical Monte Carlo for this.

3.4.3 Molecular dynamics caveats

The main advantage of molecular dynamics simulations is the explicit description of time. This allows to study time-dependent phenomena, and make the simulation trajectories easier to interpret as they form a "movie" of what is happening in the system.

But these simulations also have a few limitations. As they are based on Hamiltonian dynamics, they can only sample ensembles with a constant number of particles. At the same time, it is harder to ensure that the trajectory is sampling the right ensemble, compared with Monte Carlo simulations where the ensemble is built directly into the moves and their probabilities. Generating the right average value — for example when using a Berendsen thermostat — is not enough, and one must check that the full distribution is reproduced.

In general, Monte Carlo and molecular dynamics simulations complete each other, one having strengths where the other is weaker, or harder to use. The main difference between them remains the way they explore the phase space, represented figure 3.5. Molecular dynamics will generate a smooth trajectory in the phase space, while Monte Carlo trajectories will exhibit jumps and discontinuities. This ability to use non-physical moves (such as a change in the number of particles in the system) make Monte Carlo simulations a more versatile tool for molecular simulations, able to jump over local energy barriers and improve sampling rate of the phase space.

3.5 Free energy methods

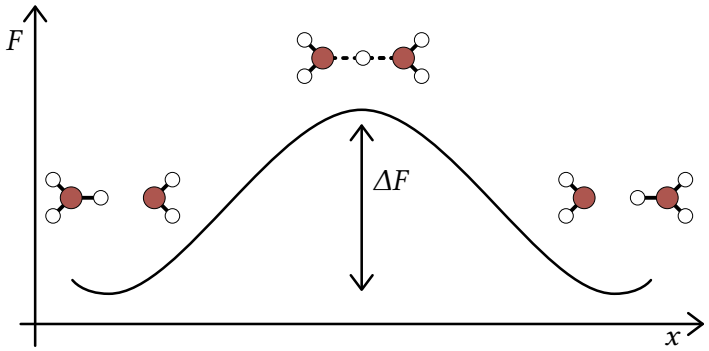


Figure 3.8 – Representation of a free energy profile along a reaction coordinate x, on the example of proton exchange in water.

Sometimes, the knowledge and sampling of the whole phase space is not necessary to acquire insight into the system of interest. For example, a reaction occurring in a water solution will only involve a few degrees of freedom (the distance between the reacting molecules) among the thousands of degrees of freedom composing the phase space — this is illustrated in figure 3.8 where the distance of the proton to the first water molecule is used as reaction coordinate. In these cases, it can be more efficient to only consider these degrees of freedom and extract the corresponding free energy surface. This is what free energy methods are designed to do. The idea is to bias the system to sample preferentially the degrees of freedoms of interest, and extract the free energy profile from the resulting trajectory.

Another application of free energy methods is the study of rare events, *i.e.* any event involving a high free energy barrier. If we were to leave the simulation evolve by itself, very long simulation time would be needed to obtain correct sampling of the entire phase space. This is true of both molecular dynamics where the barrier crossing is explicit, and Monte Carlo where even if a move can "jump" over the barrier, moves approaching the barrier have a higher chance of being rejected. In both cases, free energy methods — also called *biased simulations* or *non-Boltzmannian sampling* simulations — can help overcome the issue.

It is important to note that free energy methods do not allow to compute absolute free energies, but only free energy differences. This is because in statistical thermodynamics, free energy is defined as a function of the partition function:

$$\mathrm{F} = -\frac{1}{\beta} \ln \left(\iiint_{\Phi} \mathrm{d}\boldsymbol{r}^{\mathrm{N}} \mathrm{d}\boldsymbol{p}^{\mathrm{N}} \; e^{-\beta \, \mathrm{H}(\boldsymbol{r}^{\mathrm{N}}, \boldsymbol{p}^{\mathrm{N}})} \right) \tag{3.62}$$

which cannot be computed as an ensemble average using equation (3.17). It is however possible to express free energy differences as quantities based on ensemble average.

I will focus here on the Umbrella Sampling method which is the main method I used during my work. Other methods to extract free energy from a simulation include Thermodynamic Integration[92], Metadynamics[112], Blue Moon ensemble simulations[113], Adaptive Biasing Force[114], Wang Landau Monte Carlo[115], and many others.

3.5.1 Umbrella sampling

Umbrella sampling is a method to extract the free energy profile along a reaction coordinate x. The idea is to use a series of independent simulations where the system is constrained at specific positions x_i with a harmonic potential, such that the total potential energy is:

$$\mathrm{U}_i(\boldsymbol{r}^\mathrm{N}) = \mathrm{U}(\boldsymbol{r}^\mathrm{N}) + \frac{1}{2} k \left(x - x_i\right)^2 \tag{3.63}$$

With this energy, the probability that the system is in a state described by x is given by

$$\mathcal{P}_i(x) = \frac{1}{\mathrm{Z}_i} \iiint_\mathrm{V} \mathrm{d}\boldsymbol{r}^\mathrm{N} \ \exp[-\beta\, \mathrm{U}_i(\boldsymbol{r}^\mathrm{N})] \ \delta(x - x(\boldsymbol{r}^\mathrm{N})). \tag{3.64}$$

The above expression uses $x(\boldsymbol{r}^\mathrm{N})$ as the value of the collective variable x for a given conformation, and the δ function to filter out a specific value of x. Z_i is the modified conformational partition function:

$$\mathrm{Z}_i = \iiint_\mathrm{V} \mathrm{d}\boldsymbol{r}^\mathrm{N} \ \exp[-\beta\, \mathrm{U}_i(\boldsymbol{r}^\mathrm{N})]. \tag{3.65}$$

The unbiased system is described by a distribution probability $\mathcal{P}_0$:

$$\mathcal{P}_0(x) = \frac{1}{\mathrm{Z}_0} \iiint_\mathrm{V} \mathrm{d}\boldsymbol{r}^\mathrm{N} \ \exp[-\beta\, \mathrm{U}(\boldsymbol{r}^\mathrm{N})] \ \delta(x - x(\boldsymbol{r}^\mathrm{N})). \tag{3.66}$$

$$\mathrm{Z}_0 = \iiint_\mathrm{V} \mathrm{d}\boldsymbol{r}^\mathrm{N} \ \exp[-\beta\, \mathrm{U}(\boldsymbol{r}^\mathrm{N})]. \tag{3.67}$$

And the overall goal of the method is to evaluate the free energy of the system as a function of x, using

$$\mathrm{F} = -k_\mathrm{B}\mathrm{T} \ln\left(\mathcal{P}_0(x)\right) \tag{3.68}$$

Umbrella sampling relies on running multiple simulations for different values of x_i, and approximating $\mathcal{P}_i(x)$ by a histogram of the values taken by x during the constrained simulations. In theory, it is then possible to reconstruct $\mathcal{P}_0$ from any of these histograms:

$$\mathcal{P}_0(x) = \exp\left(\beta \frac{1}{2} k \left(x - x_i\right)^2\right) \frac{\mathrm{Z}_i}{\mathrm{Z}_0} \mathcal{P}_i(x) \tag{3.69}$$

In practice, this requires having very good statistics in regions where $\mathcal{P}_i$ is near zero, and thus prohibitively long simulation time. Instead, we try to reconstruct $\mathcal{P}_0$ as a linear combination of such estimates:

$$\mathcal{P}_0(x) = \sum_i w_i(x) \exp\left(\beta \frac{1}{2} k \left(x - x_i\right)^2\right) \frac{\mathrm{Z}_i}{\mathrm{Z}_0} \mathcal{P}_i(x) \tag{3.70}$$

where the $w_i(x)$ are weight functions to be determined, normalized such that:

$$\forall x, \quad \sum_i w_i(x) = 1 \tag{3.71}$$

It is then possible to solve these equations for $w_i(x)$ and $\mathrm{Z}_i/\mathrm{Z}_0$ using a self-consistent approach, as expressions for $w_i(x)$ depend on $\mathrm{Z}_i/\mathrm{Z}_0$ and *vice versa*. This approach is called the *Weighted Histogram Analysis Method*[116] (WHAM).

4

FIRST PRINCIPLES SIMULATIONS

As discussed in the previous chapter, to be able to compute the properties of a system using statistical physics, we need to be able to compute the total energy of this system in every possible configuration. It is possible to do so by considering the quantum nature of the electrons, and solving the Schrödinger equation for this system. We can still model the atomic nuclei as classical particles, as this is usually sufficient for the description of the system. Solving the Schrödinger equation to compute the energy of the system presents the advantage that no additional hypotheses have to be made about the system, and the functional form of the interaction potential $U(\boldsymbol{r})$.

In this chapter, I will first introduce the Density Functional Theory (DFT) approach used to solve the Schrödinger equation. I will then explain how I used it to run *ab initio* molecular dynamics to study deformations of ZIF-8 during N_2 adsorption. Finally, I will discuss how we can use DFT simulations to parametrize classical force fields, allowing us to study bigger systems on longer timescales at the cost of some accuracy.

4.1 From Schrödinger to Density Functional Theory

4.1.1 The Schrödinger equation

In quantum chemistry, the state of a system is represented by a complex-valued function of positions and time $\Psi(\boldsymbol{r}, t)$ called the *wave function*. The Schrödinger equation describes the time evolution of this wave function given the Hamiltonian $\hat{\mathcal{H}}$ of the system[117]:

$$\hat{\mathcal{H}}\Psi(\boldsymbol{r}, t) = j\hbar\frac{\partial}{\partial t}\Psi(\boldsymbol{r}, t), \tag{4.1}$$

where j is the imaginary number ($j^2 = -1$), and $\hbar$ the reduced Plank's constant. The quantum Hamiltonian operator $\hat{\mathcal{H}}$ can be expressed as a sum of quantum kinetic energy and potential energy operators:

$$\hat{\mathcal{H}} = \sum_i \frac{-\hbar^2}{2m_i}\nabla_i^2 + \mathrm{U}(\boldsymbol{r}) \tag{4.2}$$

where the m are the masses of the particles in the system, ∇ is the differential nabla operator, and $\mathrm{U}(\boldsymbol{r})$ is the potential energy of the system. When the potential energy does not explicitly depend on time, we can look for solution with separated variables of the form:

$$\Psi(\boldsymbol{r}, t) = \psi(\boldsymbol{r}) \times y(t). \tag{4.3}$$

Replacing in equation (4.1), and separating the spatial and time quantities, we get

$$-\sum \frac{\hbar^2}{2m}\frac{1}{\psi(\boldsymbol{r})}\nabla^2\psi(\boldsymbol{r}) + \mathrm{U}(\boldsymbol{r}) = j\hbar\frac{1}{y(t)}\frac{\mathrm{d}y(t)}{\mathrm{d}t}. \tag{4.4}$$

For this equation to stand for all t and all r, there must exist a constant value E such that:

$$j\hbar\frac{1}{y(t)}\frac{\mathrm{d}y(t)}{\mathrm{d}t} \quad = \quad \mathrm{E} \quad = \quad -\sum \frac{\hbar^2}{2m}\frac{1}{\psi(\boldsymbol{r})}\nabla^2\psi(\boldsymbol{r}) + \mathrm{U}(\boldsymbol{r}) \tag{4.5}$$

Putting everything together, any solution of the general equation (4.1) can be written as a (potential infinite) linear combination of special solutions:

$$\Psi(\boldsymbol{r}, t) = \sum_i^{\infty} c_i\, \psi_i(\boldsymbol{r})\, e^{-j\,\mathrm{E}_i\, t\,/\hbar}, \tag{4.6}$$

with c_i being the complex coefficients of the combination; and the $(\mathrm{E}_i\ ;\ \psi_i)$ pairs are solutions of the *time-independent* Schrödinger equation:

$$\hat{\mathcal{H}}\psi(\boldsymbol{r}) = \mathrm{E}\psi(\boldsymbol{r}). \tag{4.7}$$

Ψ is called a *wave function* because its time dependence has a wave-like form, and it propagates through space and time. Solving equation (4.7) gives us stationary states $(\mathrm{E}_i\ ;\ \psi_i)$, where E_i is the energy of the state; and ψ_i is the wave function of the state. Although there is no easy interpretation of the complex valued function ψ_i, its squared norm $|\,\psi_i(\boldsymbol{r})\,|^2$ is the probability density of finding the electrons around a given position.

4.1.2 Density Functional Theory

To compute the energy of an atomic system, we need to solve equation (4.7) for a collection of N electrons carrying a negative charge $-e$ and M nuclei carrying a positive charge $e\,\mathrm{Z}_j$, all interacting through a Coulombic potential. As the electrons are much more lightweight than the nuclei (an electron is roughly 2000 times lighter than a proton), it is customary to work under the Born-Oppenheimer approximation[118]. In this approximation, the degrees of freedom of electrons and nuclei are decoupled, and the electrons move much faster than the nuclei, reacting instantly to changes in the nuclei's positions. This effectively means that at a given point in time, the electrons evolve in the constant electrostatic potential generated by the fixed nuclei. The corresponding Hamiltonian (using atomic units, *i.e.* $e^2/4\pi\varepsilon_0 = 1$; $\hbar = 1$; and $m_{\text{electron}} = 1$) is:

$$\hat{\mathcal{H}} = \left[-\frac{1}{2}\sum_{i}^{\mathrm{N}} \nabla_i^2 - \sum_{i}^{\mathrm{N}}\sum_{j}^{\mathrm{M}} \frac{\mathrm{Z}_j}{|\boldsymbol{r}_i - \boldsymbol{R}_j|} + \sum_{i}^{\mathrm{N}}\sum_{j>i}^{\mathrm{N}} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \right]. \tag{4.8}$$

We can rewrite this by defining the potential an electron feels due to the presence of all the nuclei $\mathrm{V}_{\text{ext}}(r) = -\sum_j \mathrm{Z}_j/|\boldsymbol{r} - \boldsymbol{R}_j|$, and the electron-electron interaction potential $\mathrm{U}(r_i, r_j) = 1/|\boldsymbol{r}_i - \boldsymbol{r}_j|$:

$$\hat{\mathcal{H}} = \left[-\frac{1}{2}\sum_{i=1}^{\mathrm{N}} \nabla_i^2 + \sum_{i=1}^{\mathrm{N}} \mathrm{V}_{\text{ext}}(\boldsymbol{r}_i) + \sum_{i=1}^{\mathrm{N}}\sum_{j>i}^{\mathrm{N}} \mathrm{U}(\boldsymbol{r}_i, \boldsymbol{r}_j) \right], \tag{4.9}$$

$$\hat{\mathcal{H}} = \hat{\mathrm{T}} + \hat{\mathrm{V}}_{\text{ext}} + \hat{\mathrm{V}}_{ee}. \tag{4.10}$$

The three terms in the Hamiltonian are the kinetic energy $\hat{\mathrm{T}}$, the total external potential $\hat{\mathrm{V}}_{\text{ext}}$, and the electron-electron interaction potential $\hat{\mathrm{V}}_{ee}$. This electron-electron interaction makes the state of any electron depend on the state of all other electrons; thus making the wave function depends on 3N coupled variables. Looking directly for analytic or numerical solutions of the Schrödinger equation is hard because the search space is huge: even for small systems, they are already thousands of electrons interacting.

The Density Functional Theory (DFT) is a strategy to solve the Schrödinger equation without having to explicitly determine the wave function ψ. It is based on the fact that the square of the norm of the wave function is the probability for the system to be in a given state. For a system of electrons evolving in the fixed potential created by the nuclei, the wave function $\psi(\{\boldsymbol{x}\})$ depends on the Cartesian coordinates of all the electrons $\{\boldsymbol{x}\}$. The electronic density $n(\boldsymbol{r})$ is the probability to find an electron in a small neighboring of $\boldsymbol{r}$, defined as:

$$n(\boldsymbol{r}) = \mathrm{N} \iiint |\psi(\boldsymbol{r}, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{\mathrm{N}})|^2 \, \mathrm{d}\boldsymbol{x}_2 \cdots \mathrm{d}\boldsymbol{x}_{\mathrm{N}} \tag{4.11}$$

The whole idea of DFT is to transform the problem so that we can look for a solution of the Schrödinger equation as a function of 3 variables $n(\boldsymbol{r})$ instead of a function of 3N variables ψ.

Hohenberg and Kohn showed in 1964[119] that every electronic density correspond exactly to one and only one external potential V_{ext}, *i.e.* the knowledge of the external potential or the density are equivalent. This is the first Hohenberg and Kohn theorem. This also means that knowing the electronic density is equivalent to knowing the wave

function of the system, as one can reconstruct it using equation (4.7). As a consequence, all the observables of the system only depend on the electronic density, and can be written as functionals of this density: $\mathrm{E}[n]$, $\Psi[n]$, *etc.*

The second Hohenberg and Kohn theorem[119] states that the ground state electronic density $n_0(\boldsymbol{r})$ is the one that minimize the energy functional: $\mathrm{E}[n] \geq \mathrm{E}_0 = \mathrm{E}[n_0]$. From the previous relations, the total energy functional contains three terms:

$$\mathrm{E}[n] = \mathrm{T}[n] + \mathrm{V}_{\mathrm{ext}}[n] + \mathrm{V}_{ee}[n]; \tag{4.12}$$

where the potential terms can be expressed as integrals of the one and two electrons density:

$$\mathrm{V}_{\mathrm{ext}}[n] = \int \mathrm{V}_{\mathrm{ext}}(\boldsymbol{r})\, n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}; \tag{4.13}$$

$$\mathrm{V}_{ee}[n] = \iint \frac{n_2(\boldsymbol{r}_1, \boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, \mathrm{d}\boldsymbol{r}_1 \mathrm{d}\boldsymbol{r}_2. \tag{4.14}$$

Here, the two electrons density $n_2(\boldsymbol{r}_1, \boldsymbol{r}_2)$ is the probability for an electron to be in a small neighboring of $\boldsymbol{r}_1$, while another electron is in a small neighboring of $\boldsymbol{r}_2$:

$$n_2(\boldsymbol{r}_1, \boldsymbol{r}_2) = \mathrm{N} \iiint |\psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_{\mathrm{N}})|^2\, \mathrm{d}\boldsymbol{x}_3 \cdots \mathrm{d}\boldsymbol{x}_{\mathrm{N}}. \tag{4.15}$$

While we can compute and minimize $\mathrm{V}_{\mathrm{ext}}[n]$ given the positions of the nuclei, the two other terms $\mathrm{T}[n]$ and $\mathrm{V}_{ee}[n]$ are harder to evaluate.

Kohn and Sham reformulated the problem in 1965[120] by considering a set of *non-interacting* electrons evolving in a specific external potential, such as the density arising from these electrons is the same as the one we look for. The new system of non-interacting electrons is described by a new set of independent orbitals φ_i, such as the total electronic density of the system is

$$n(\boldsymbol{r}) = \sum_i^{\mathrm{N}} \left|\varphi_i(\boldsymbol{r})\right|^2. \tag{4.16}$$

For these non interacting electrons, the kinetic energy $\mathrm{T}_s[n]$ is known:

$$\mathrm{T}_s[n] = \sum_i^{\mathrm{N}} \int \varphi_i^*(\boldsymbol{r}) \left(-\frac{1}{2}\nabla^2\right) \varphi_i(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}. \tag{4.17}$$

We also know that most of the $\mathrm{V}_{ee}[n]$ term can be approximated as a classical classical Coulombic interaction — the so called *Hartree energy* — which we can compute directly from the electronic density:

$$\mathrm{V}_{\mathrm{H}}[n] = \iint \frac{n(\boldsymbol{r}_1) n(\boldsymbol{r}_2)}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}\, \mathrm{d}\boldsymbol{r}_1 \mathrm{d}\boldsymbol{r}_2. \tag{4.18}$$

The functional to minimize in order to find the electronic density is now

$$\mathrm{E}[n] = \mathrm{T}_s[n] + \mathrm{V}_{\mathrm{ext}}[n] + \mathrm{V}_{\mathrm{H}}[n] + \mathrm{E}_{xc}[n], \tag{4.19}$$

in which we can compute all of the terms expect for the exchange-correlation contribution $\mathrm{E}_{xc}[n]$.

$$\mathrm{E}_{xc}[n] = (\mathrm{T}[n] - \mathrm{T}_s[n]) + (\mathrm{V}_{ee}[n] - \mathrm{V}_{\mathrm{H}}[n]). \tag{4.20}$$

The form of this contribution is generally unknown, and it describes the quantum nature of electrons that are able to interact with themselves. The "exchange-correlation" name comes from the two different physical terms that it encompasses. The exchange energy is a quantum-only effect due to the fact that electrons are not distinguishable, and that exchanging two of them should change the sign of the wave function. This exchange effect is sometimes called Pauli repulsion, as it increases the distance between the electrons. The correlation term describes how much the presence of other electrons influences the position of a given electron; *i.e.* how much the two electrons density $n_2(\boldsymbol{r}_1, \boldsymbol{r}_2)$ differs from the product of single electrons densities $n(\boldsymbol{r}_1)n(\boldsymbol{r}_2)$.

It can be shown that the minimizing the energy functional is equivalent to solving a set of differential equations called the Kohn-Sham equations:

$$\left[-\frac{1}{2}\nabla^2 + \mathrm{V}_{\mathrm{ext}}(\boldsymbol{r}) + \frac{\partial \mathrm{V}_{\mathrm{H}}[n](\boldsymbol{r})}{\partial n(\boldsymbol{r})} + \frac{\partial \mathrm{E}_{xc}[n](\boldsymbol{r})}{\partial n(\boldsymbol{r})}\right] \varphi_i = \varepsilon_i \varphi_i \tag{4.21}$$

These are non-linear equations, as both $\mathrm{V}_{\mathrm{H}}[n]$ and $\mathrm{E}_{xc}[n]$ depends on the φ_i through $n(\boldsymbol{r})$. The usual algorithm to solve them is a self-consistent iterative algorithm. We start by making an initial guess for the Kohn-Sham orbitals φ_i^0, using for example a linear combination of atomic orbitals. Using this initial guess, we can:

1. calculate the corresponding electron density at step α, $n^\alpha(\boldsymbol{r})$;
2. calculate all the terms in the left hand side of equation (4.21): kinetic, external, Hartree and exchange-correlation energies using this electron density;
3. solve the Kohn-Sham equation to find new Kohn-Sham orbitals $\varphi_i^{\alpha+1}$;
4. compute the new energy using the energy functional (4.19).

If the new energy $\mathrm{E}^{\alpha+1}$ and density $n^{\alpha+1}(\boldsymbol{r})$ are the same as the energy E^α and density $n^\alpha(\boldsymbol{r})$ in the previous step, then we have found the fixed point of the Kohn-Sham equation, and the converged density is the solution of the Schrödinger equation. Else, we go back to step 1 using the new Kohn-Sham orbitals as the initial guess. This cycle is called the *Self-Consistent Field* (SCF) method, as we iteratively solve the equations until the solution is self-consistent.

When a solution is found, the ground state energy $\mathrm{E}^{\alpha+1}$ and electronic density $n^{\alpha+1}(\boldsymbol{r})$ will be the values minimizing the energy functional from equation (4.19). The Kohn-Sham energies $\{\varepsilon_i\}$ on the other hand have no physical meaning. They are only the solution of equation (4.21) for a system of non-interacting electrons that happens to have the same electronic density as the real system of interest.

4.1.3 Exchange, correlation, and dispersion

A good approximation for the exchange-correlation functional is required to solve the Kohn-Sham equations. There are a number of approaches routinely used in theoretical chemistry, usually classified along the so-called Perdew's ladder[121] from less precise and

cheaper to compute to more precise and expensive to compute. Going up the ladder, we encounter Local Density Approximation (LDA) where the exchange-correlation functional only depends on the local density; Generalized Gradient Approximation (GGA), where the functional also depends on the local gradient of the density, meta-GGA functionals incorporating the second derivative of the density (the Laplacian of the density), or hybrid functionals. Hybrid functionals mix exchange expressions from LDA or GGA with exact exchange term coming from other methods for solving the Schrödinger equation, such as Hartree-Fock methods.

Generally, we use exchange-correlation functionals that only depend on the local density, and maybe some of its derivatives. This means that DFT calculations will have trouble reproducing non-electrostatic long-range correlation effects, such as the dispersion interaction. Fortunately, we have a simple analytic formulation for these interactions, which Grimme and coworkers[122, 123] proposed to include when computing the total energy. This empirical correction is added to the energy obtained by DFT, and does not directly modify the electronic density.

$$\mathrm{E_{tot}} = \mathrm{E_{DFT}} + \mathrm{E_{disp}} \tag{4.22}$$

In my PhD, I used the DFT-D3 version of this correction[123], with zero damping and only the two-body term. The correction to the energy is expressed as:

$$\mathrm{E_{disp}} = -\sum_{i}^{\mathrm{M}} \sum_{j>i}^{\mathrm{M}} \left[\frac{\mathrm{C}_6^{ij}}{\mathrm{R}_{ij}^6} f_6^{ij}(\mathrm{R}_{ij}) + \frac{\mathrm{C}_8^{ij}}{\mathrm{R}_{ij}^8} f_8^{ij}(\mathrm{R}_{ij}) \right], \tag{4.23}$$

where the sum is over all the pairs of atoms, R_{ij} is the distance between the atoms, the C_6^{ij} are parameters tabulated for each element pairs and f_n^{ij} is a damping function.

$$f_n^{ij}(\mathrm{R}) = \begin{cases} \dfrac{s_n}{1 + 6\left(\mathrm{R} \,/\, \sigma_n \sqrt{\mathrm{C}_8^{ij}/\mathrm{C}_6^{ij}}\right)^{-\alpha_n}} & \text{for } \mathrm{R} < \mathrm{R}_c \\ 0 & \text{for } \mathrm{R} \geq \mathrm{R}_c \end{cases} \tag{4.24}$$

In this expression, s_n and σ_n are parameters specifically adapted to the exchange-correlation functional one uses. For example, when using the PBE functional one should set $s_6 = 1.0$, $s_8 = 0.722$, $\sigma_6 = 1.217$, and $\sigma_8 = 1.0$. α_6 is always set to 14, and α_8 to 16.

4.1.4 Ab initio molecular dynamics

DFT can be used to compute the energy and forces acting on the atoms during an *ab initio* molecular dynamics (AIMD) simulation. Different types of dynamics can be used for the nuclei depending on the time precision needed, I only used the Born-Oppenheimer dynamic. In this framework, the nuclei are supposed to behave as classical particles, and DFT is used to optimize the electronic density at each step of the dynamic. All the algorithms devised for molecular dynamics (see section 3.4) can be used. In particular, it is possible to use AIMD to sample the NVE, NVT and NPT ensembles; although simulation in the NPT ensemble are more costly than simulations in the NVT ensemble, because the proper calculation of the pressure requires more precise atomic forces, and a stricter convergence criterion in the SCF cycles.

4.2 Adsorption of N_2 in ZIF-8 and its derivatives

In this section, I will show an example of how the use of *ab initio* molecular dynamics enables a precise description of the adsorbed phase in flexible porous materials. This study has been published in *The Journal of Physical Chemistry C* (2018)[124].

4.2.1 ZIF-8 and its cousins

The flexible porous material used in this study is ZIF-8. It is a representative of the Zeolitic Imidazolate Framework (ZIF) family of MOFs, with a sodalite (SOD) topology, methyl-imidazolate linkers, and zinc metallic centers. The structure of ZIF-8 is represented in figure 4.1.

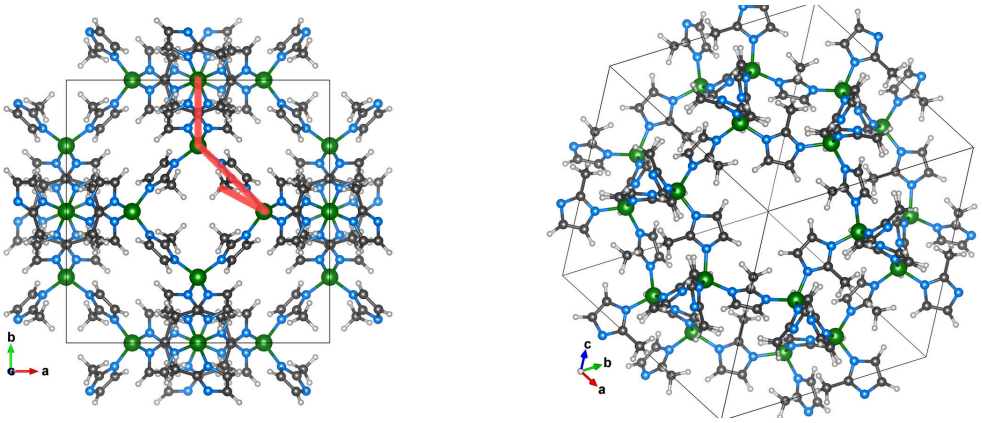


Figure 4.1 – Structure of the methyl-imidazolate ZIF-8. On the left, the front-most opening is the 4-member ring window (the structure is represented along the 001 axis); and on the right the front-most opening is the 6-member ring window (the structure is represented along the 111 axis). The atomic color code is green for zinc, blue for nitrogen, gray for carbon and white for hydrogen. The red line connects the atoms of the "swing" Zn – Zn – Zn – X dihedral angle.

ZIF-8 is known for it relatively high surface area of 1.947 m^2/g and large cavity with a diameter of 11.6 Å[19], separated by 6-ring windows of small diameter ($\approx$ 3.4 Å). It is also commercially available and have excellent performance at the lab scale for the separation of gas mixtures such as C_2/C_3 hydrocarbons, CO_2/CH_4, or CO_2/N_2[125, 126].

It exhibits some flexibility of its framework through torsional rotation of the linkers around the plane defined by the 6-member ring windows. The first evidence of such flexibility came from the fact that molecules bigger than the window aperture can diffuse through the structure. When a molecule bigger than the geometric window size tries to go through the window, the linkers will rotate, effectively increasing the apparent window size. In particular, it has been shown that we need to account for the flexibility of the structure to be able to properly reproduce experimental diffusion coefficients of small hydrocarbons[127] when using molecular simulation.

Moreover, there is evidence of a structural transition occurring in ZIF-8 upon loading, from an ambient pressure or AP phase to a high pressure or HP phase. This was demonstrated using *in situ* X-ray diffraction both during methanol/ethanol intrusion at 1.47 GPa[128]; and during the adsorption of N_2 at 77 K[129]. In the latter case, the authors

also demonstrated that the stepped isotherm (see figure 4.3) could not be reproduced by Grand Canonical Monte Carlo (GCMC) simulations in the rigid AP structure. Instead, they showed that the lower pressure regime is reproduced by GCMC simulations in the rigid AP structure, while the higher regime corresponds to GCMC simulations in the rigid HP structure.

The AP and HP phases have the same space group ($I\bar{4}3m$), and mainly differ by the orientation of the linkers around the 4 and 6-member windows. This orientation is defined by the dihedral angle $Zn-Zn-Zn-CH_3$, also called the *swing angle*; which goes from an average value of 7° in the AP phase to an average value of 35° in the HP phase.[129] This angle is represented in figure 4.1. A recent, more detailed characterization of the structural evolution upon adsorption showed that the transition from AP to HP is a continuous rotation of the linkers, and not an abrupt change.[130].

Other studies[131, 132] showed that the adsorption isotherms of CO, O_2 and Ar at 77 K and 90 K all present similar stepped isotherm and hysteresis loop in the high pressure regime. Again, these isotherms could not be fully described by GCMC simulations in the AP structure only, instead we also need to account for the HP phase.

Nitrogen adsorption isotherms in ZIF-8, ZIF-8(Cl), and ZIF-8(Br)

Two new materials analogs to ZIF-8 have been synthesized recently by Li et al.[125] using chloro- and bromo-substituted imidazolate linkers instead of the original methyl-imidazolate in ZIF-8. These new materials, that we will call ZIF-8(Cl) and ZIF-8(Br), are represented in figure 4.2.

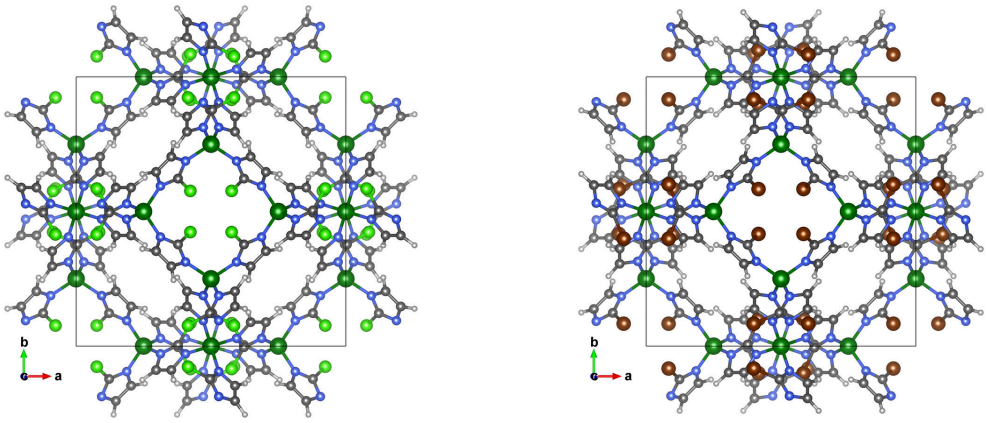


Figure 4.2 – Structure of the chloro-imidazolate ZIF-8(Cl) on the left and the bromo-imidazolate ZIF-8(Br) on the right. The atomic color code is the same as in figure 4.1, with a bright green for chlorine atoms and dark red for bromide atoms.

The nitrogen adsorption isotherms at 77 K in ZIF-8(CH_3), ZIF-8(Cl) and ZIF-8(Br), are presented in figure 4.3. They were measured as part of a collaboration by the experimental team IS2M laboratory in Université de Haute Alsace, France.

Looking at these isotherms, we first notice that for ZIF-8(CH_3), there are two jumps in loading: a first one corresponding to the initial filling of the pores up until around 300 cm^3/cm^3, and a second one up until 400 cm^3/cm^3. This shape of isotherm is known as a type IV isotherm in the International Union of Pure and Applied Chemistry (IUPAC)

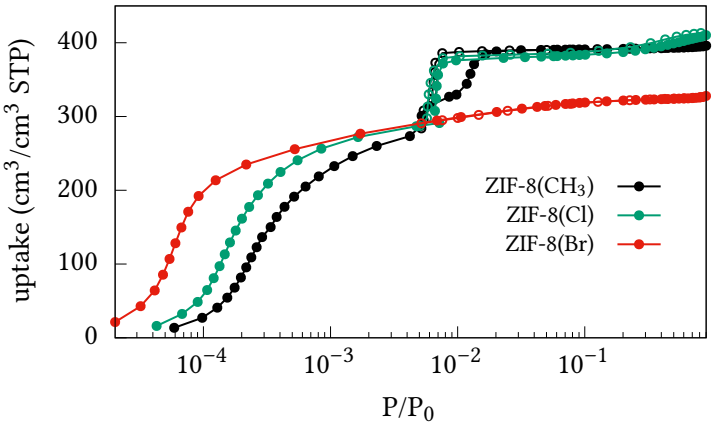


Figure 4.3 – Nitrogen adsorption isotherms at 77 K in ZIF-8(CH_3) (black) and the chloro (green) and bromo (red) derivatives. Full circles are used for the adsorption branch, and empty ones for the desorption branch. Experimental data acquired by Chaplais et al.[124]

classification[43] (see figure 1.12). Type IV isotherms usually appear when two kinds of porosity exist, and the second, larger kind is filled at higher pressures; for example when there are both micropores and mesoscopic pores in the material. In the case of ZIF-8(CH_3), there is no evidence for the existence of such mesoscopic pores, and instead the step in the isotherm is attributed to the phase transition between the AP and HP phases.

Concerning the two new materials, ZIF-8(Cl) shows the same behavior with a stepped, type IV adsorption isotherm. ZIF-8(Br) however presents a different isotherm shape, without the second jump in loading, known as a type I isotherm in the IUPAC classification. This kind of isotherm usually occurs in materials where there is a single kind of micropores.

Ab initio molecular dynamics study

In order to gain better understanding of the relation between the structure changes and the adsorbed molecules, I used molecular dynamics simulations. To describe fully the flexibility of the frameworks without any assumption, I favored *ab initio* molecular dynamics over force field-based molecular dynamics — as there are currently no force fields available for the ZIF-8 variants studied here, and existing force fields for ZIF-8(CH_3) have questionable accuracy. For each framework — ZIF-8(CH_3), ZIF-8(Cl), ZIF-8(Br)— I ran five simulations corresponding to different numbers of nitrogen molecules N inside the porous space, going from an empty framework (N = 0) to the fully loaded host material. The maximal loading was determined from the experimental isotherms to be close to N = 50 molecules per unit cell for ZIF-8(CH_3) and ZIF-8(Cl) and N = 40 molecules per unit cell for ZIF-8(Br).

To create the starting configurations, I started from the energy-minimized configuration of the empty frameworks, and randomly placed the selected number of nitrogen molecules in the unit cell using the packmol software[133]. The whole {ZIF, adsorbate} system was then minimized again at the DFT level before starting the molecular dynamics simulations. I used the Quickstep module[134] of the CP2K software package (version 2.5.1, available

online at http://www.cp2k.org/) for all the simulations. I used a PBE exchange-correlation functional with D3 dispersion corrections, a double zeta polarizable valence (DZVP) basis set, and an energetic cutoff of 600 Ry. All the systems were simulated with a 1 fs timestep giving a total of 12 to 22 ps of simulation. Temperature was held constant at 77 K with a CSVR thermostat, using a thermostat time constant of 1000 fs. I used the last 5 ps of simulation for analysis, leaving 7 to 17 ps to the system to reach equilibrium.

We used constant volume (NVT) simulations instead of constant pressure (NPT) simulations because NPT molecular dynamics simulations require a higher computation time to correctly converge the calculation of forces. In general, when global deformations are expected in the system, one should use NPT simulation. In this case, I was able to use NVT simulations because the volume change between the different phases of ZIF-8(CH_3) is very small. The low-pressure phase unit cell volume is 4900.5 $Å^3$, while the high-pressure phase volume is 4974.8 $Å^3$, making the difference less than 2%.

4.2.2 Deformation under adsorption

The first indicator of deformation of the framework is the Zn – Zn – Zn – X dihedral angle, where X is the group on the imidazolate linkers: methyl, chlorine or bromide. This angle — presented on figure 4.1 — is also called the swing angle, noted φ. The swing angle represents the rotation of the linker around the 6-member window plane, with 0° being the point where the linker is completely in this plane.

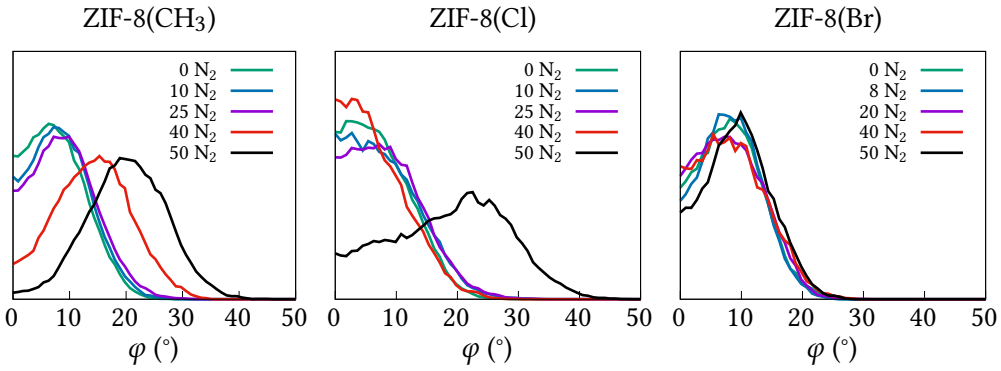


Figure 4.4 – Distribution of linker swing angle (Zn – Zn – Zn – X dihedral angle, where X stands for CH_3, Cl or Br for ZIF-8(CH_3), ZIF-8(Cl) or ZIF-8(Br), respectively) at various values of nitrogen loading.

Following the evolution of the distribution of swing angles as the number of nitrogen molecules increases allow understanding the deformations of the three frameworks under adsorption. This evolution is represented on figure 4.4, where we observe for ZIF-8(CH_3) a gradual increase of the mean angle value as loading increases while the distribution remains Gaussian-like.

These results are consistent with the already published ones for ZIF-8(CH_3) [130], but using a different functional (PBE here instead of BLYP in reference [130]). Interestingly, the two other frameworks behave differently. For ZIF-8(Cl), almost no change is noticed in the distribution profile upon adsorption until the highest value of loading (*i.e.*, N = 50). In this case, the distribution shifts and the profile is no longer of Gaussian type, but instead looks like the sum of two Gaussian distribution, one centered around 25°, and the

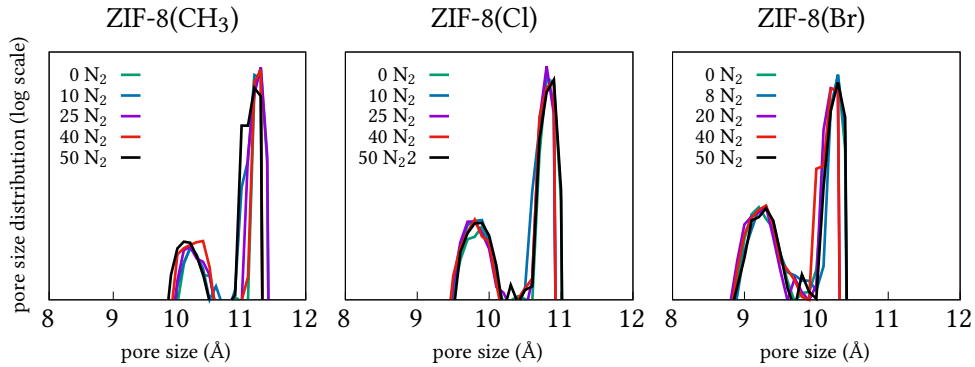


Figure 4.5 – Pore size distribution (no unit) under increasing loading for the three structures.

other one around 10°. This indicates that some of the linkers do not rotate (swing) even at high loading. The non-zero value of the distribution at $\varphi = 0°$ contrasts with the case of ZIF-8(CH_3). Finally, for ZIF-8(Br), no significant change occurs for the dihedral angle distribution as the loading increases, meaning that the linkers do not rotate.

Although this behavior appears to be correlated with the presence or absence of the adsorption step in the isotherms, it is not however sufficient to explain it. A first hypothesis to understand the link between linkers swing and the isotherms step is that the swinging motion could lead to an increase of the accessible porous volume in the structure, thereby increasing nitrogen uptake. In order to check whether this is or not the case, I computed the pore size distribution (see figure 4.5) and the accessible porous volume (see figure 4.6) from the trajectories using Zeo++[135] version 0.3. In order to compute these values, I first emptied the structure of all the nitrogen molecules, and then computed the pores sizes distribution and accessible volume of the remaining frameworks. The algorithm used by Zeo++ starts by computing the Voronoï decomposition of the framework, which provides a representation of the void space in the structure. Then, it uses Monte Carlo sampling of the Voronoï network to extract the volume accessible to a probe molecule of a given radius and the pore size distribution. I used the standard probe radius of 1.2 Å, corresponding to a Helium atom.

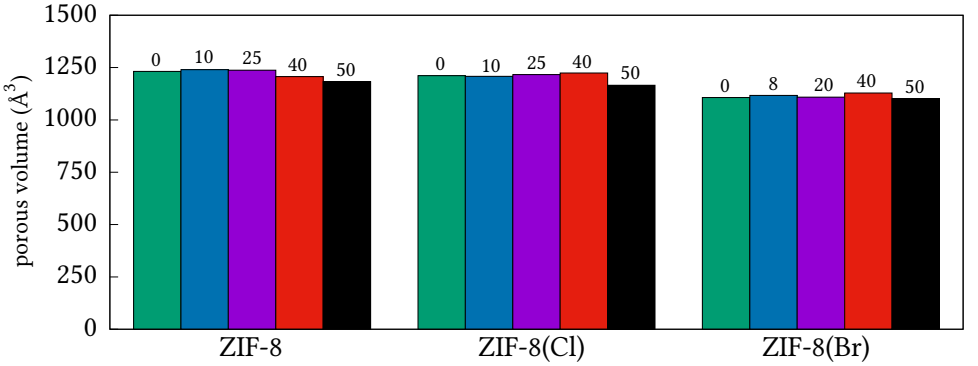


Figure 4.6 – Accessible porous volume changes in the three isoreticular structures during adsorption. The numbers on top of the columns are the number of adsorbed N_2 molecules.

The results presented in figure 4.5 highlight that the pore size distribution (PSD) remains largely unchanged in all cases as the loading increase and the linkers swing. The PSDs present two peaks, corresponding to two types of cavities: a small one corresponding to the 6 member windows around 10.3 Å for ZIF-8(CH_3) and a second one, with a much larger contribution to the overall porous volume around 11.2 Å in ZIF-8(CH_3). The three PSDs are very similar, only changing in the absolute the size of pores, in the order ZIF-8(CH_3) > ZIF-8(Cl) > ZIF-8(Br). This is also reflected in the accessible volume figure 4.6 which remains roughly constant as the loading increase and the linkers swing.

4.2.3 Changes in the adsorbed phase

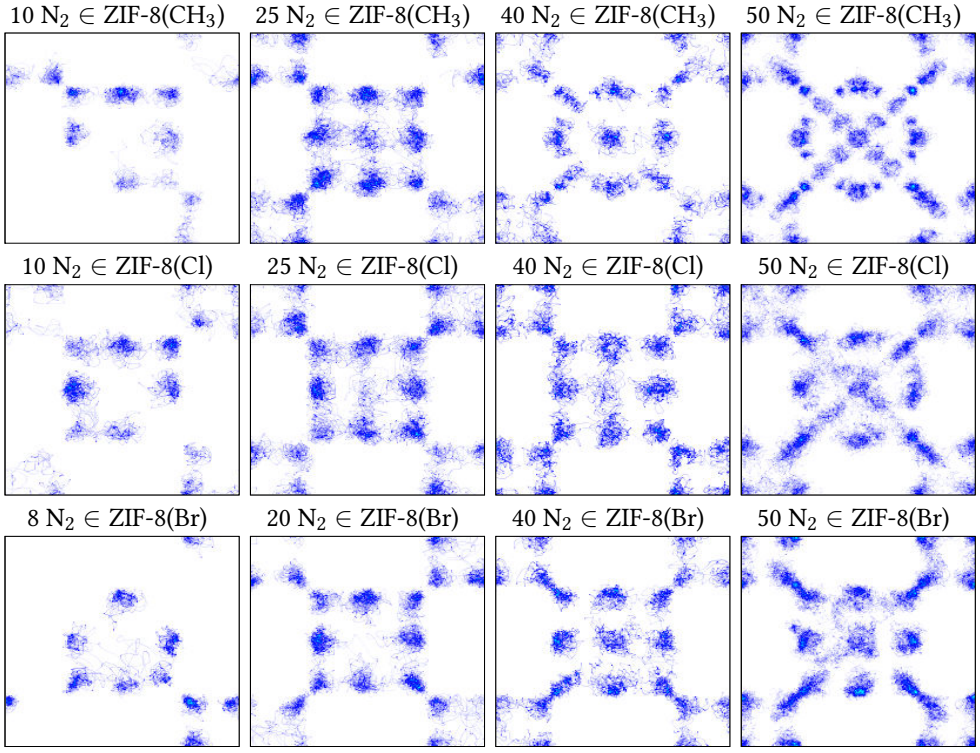


Figure 4.7 – 2D density maps of the adsorbed nitrogen atoms positions in the xy plane at various loadings in ZIF-8(CH_3) (top), ZIF-8(Cl) (middle), and ZIF-8(Br) (bottom). The loading increases from left to right.

The hypothesis that the linkers swing leads to an increase of the accessible porous volume in the structure does not seem to be able to explain the stepped adsorption isotherm. We need now to take a look at the other chemical species participating in adsorption: the nitrogen fluid. Another hypothesis we can formulate to explain the stepped isotherm is that nitrogen molecules undergo a reordering or repacking in the cavity, thereby leading to an increase in adsorbed molecules in the same total pore volume as well as the swing of the linkers aiming to accommodate the new packing. This has already been proposed by Ania et al.[132], but the intermediate adsorption regime was not probed or interpreted. In order to visualize this packing, I have projected the positions of all adsorbed nitrogen atoms in the *xy* plane and created a density map of the adsorbed phase. This density map is shown figure 4.7 both at various loadings and for the three frameworks.

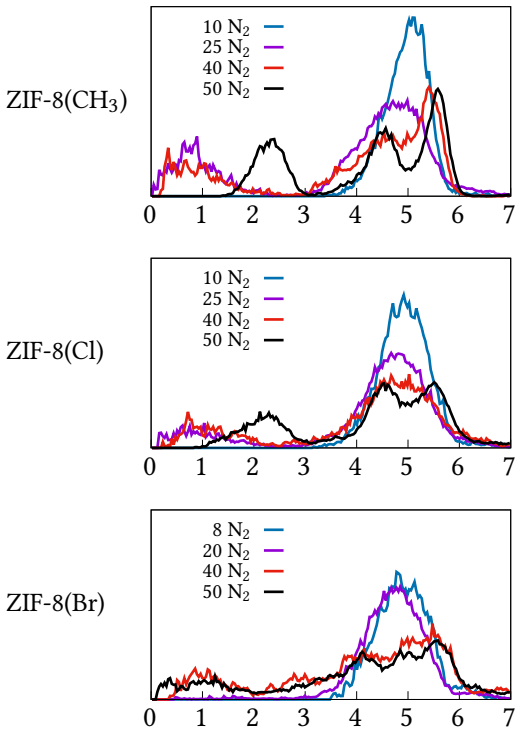


Figure 4.8 – Radial distribution function of nitrogen atoms in the three frameworks for various nitrogen loading. This represent the number of atoms at a given distance from the center of the cavity.

For ZIF-8(CH_3), two different molecular packing are encountered according to the loading. With 10 or 25 molecules in the unit cell, the density maps show clear delimited positions on a cubic-like arrangement, whereas with 40 and 50 molecules, they show a tetragonal-like arrangement of the molecules. This reordering of the adsorbed nitrogen molecules from a cubic-like phase to a tetragonal-like phase is at the origin of the steep uptake at low relative pressure. For ZIF-8(Cl), the behavior is roughly similar: the molecules first pack in a cubic-like fashion in the cases of 10, 25 and 40 molecules per unit cell, before a reordering toward a tetragonal-like arrangement at 50 molecules per unit cell. This is consistent with the dihedral angles distributions as shown in figure 4.4, and evidence that the molecular packing rearrangement happens conjointly with the swing of the linkers. It is interesting to note that some disorder in the tetragonal arrangement remains — even at a loading of 50 molecules per unit cell — as the molecules' positions are not as well defined as the one in ZIF-8(CH_3). Again, this is consistent with the dihedral angles distribution at 50 molecules per unit cell for ZIF-8(Cl) which is not of a single Gaussian type, indicating that this disorder can also be found in the framework structure. For ZIF-8(Br), the behavior is different. A same cubic-like arrangement is found at the lower loadings of 8 and 20 molecules per unit cell, whereas the arrangement at the higher loadings of 40 and 50 molecules per cell differs from the tetragonal-like one seen for ZIF-8(CH_3) and ZIF-8(Cl). Indeed, the latter appears as a mix of the cubic and the tetragonal organization as the molecules are mostly distributed on a cube with additional molecules in the [111] channels on the diagonals of the cube. Consequently, because of

the absence of a clear reordering of adsorbed nitrogen molecules in ZIF-8(Br), its sorption isotherm does not display the S-shaped adsorption step.

This repacking of nitrogen molecules is also visible on the radial distribution of nitrogen atoms (not to be confused with the pairs radial distribution function $g(r)$) in figure 4.8. The radial distribution counts how many nitrogen atoms are at a given distance from the center of the sodalite cage. On the graphs for the ZIF-8(CH_3) framework, we can see the change from the initial packing at N = 10 and N = 25 with a main peak at 5 Å and a secondary peak at 1 Å; the final packing at N = 50 with three well defined peaks at 2.5 Å, 4.5 Å, and 5.5 Å; and the N = 40 distribution which present characteristics of both extremes. Again, for ZIF-8(Cl), we observe the first bi-modal distribution for N = 10, 25 and 40; and the tri-modal distribution for N = 50. And in ZIF-8(Br) we never see the tri-modal distribution: instead at high loadings (N = 40 and N = 50) the curve is not well defined in the 3 to 6 Å range, as already seen in figure 4.7.

4.2.4 Mechanism of nitrogen adsorption in ZIF-8

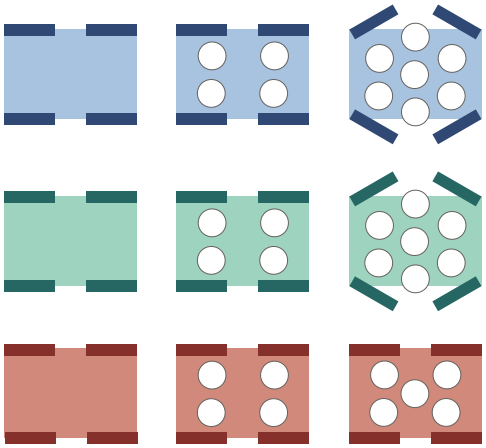


Figure 4.9 – Schematic overview of the adsorption process in the three ZIF-8 frameworks. Blue is ZIF-8(CH_3), green is ZIF-8(Cl) and red is ZIF-8(Br); loading increases from left to right.

To summarize this section, the proposed mechanism of nitrogen adsorption in the three ZIF-8 frameworks with different functionalization is illustrated in a schematic way in figure 4.9. When the pores are empty, the linkers are in their equilibrium position, around 0°. As the loading increases, the pores start to fill in a given, cubic-like arrangement. Then, as loading continues to increase, in ZIF-8(CH_3) and ZIF-8(Cl) the nitrogen molecules arrangement changes to a tetragonal-like arrangement and the linkers rotate to accommodate for this repacking. In ZIF-8(Br), the nitrogen molecules do not repack and the linkers do not rotate.

Given that the change in functional group from methyl to chloro and bromo does not significantly impact partial atomic charges or the strength of the Zn – N coordination bond, the origin of the different sorption features between the three ZIF-8 derivatives does not come from a difference in flexibility or stiffness of the linkers rotation. This is confirmed by the dihedral angles distribution in figure 4.4, where all the Gaussian distributions have the same width, which means that all the linkers rotation have the same

stiffness. Instead, the differences in the adsorption isotherms come from the differences in pore size and shape, the pores in ZIF-8(Br) being smaller than in ZIF-8(Cl) and ZIF-8(CH_3), thus preventing a molecular reordering to occur.

4.3 Classical force fields from *ab initio* data

As mentioned in section 4.2.1, I used *ab initio* molecular dynamics to study the adsorption in the three ZIF-8 derivatives to be able to describe the full flexibility of the frameworks without making any other assumptions; and because there was no available force field for the studied materials. During my PhD, I have also started a collaboration with Johannes Dürholt and Rochus Schmid from the Ruhr Universität in Bochum, Germany to parametrize a classical force field for ZIF-8(CH_3), ZIF-8(Cl), and ZIF-8(Br) using data from the *ab initio* simulations presented above. The corresponding work is published in *Journal of Chemical Theory and Computation* (2019)[136].

4.3.1 Fitting a force field

Classical force fields are fully defined by two things: (1) the choice of functional forms used to represent the different terms (*e.g.* using Lennard-Jones, Buckingham or another form for non-bonded pair interactions), and (2) the atom-dependent parameters used in these functional forms (like the ε and σ parameters for Lennard-Jones). See section 3.2.2 for a more complete description of force fields and their use in computational chemistry. The process of optimizing the parameters of a force-field is called the parametrization of the force field. The idea is to use some reference data, either from experimental properties or *ab initio* computations, and optimize the parameters to reproduce the reference data in the best possible way. Both static properties such as crystalline structure from X-ray diffraction, and dynamic ones such as vibrations frequencies as given by infrared spectroscopy can be used as reference data when fitting a force field. The created force field can then be evaluated by checking how well it can reproduce physical properties that were not used for creating it — for example the melting point, or solvation energy. A good force field will be able to reproduce most properties of a system reasonably well.

Historically, optimizing the parameters of a force-field has most often been a manual and ad hoc process. First, we would start by setting the parameters to an initial value by making an educated guess, then we would run a few static or dynamic calculations with these parameters, compute and compare the properties of interest with the reference value, and finally we would adjust the parameters and start over, until some convergence was achieved. Convergence is measured by a cost function that we try to minimize when changing the parameters. Creating a force field this way is difficult and takes a lot of time. Additionally, it is not reproducible: starting the process all over again might lead to a different set of parameters.

Transferable or accurate?

Ideally, we want the force fields we use to be as accurate as possible when computing the energy of a system, to be as confident as possible that the model described by the force field describe the chemical reality as well as possible. This means creating a specific force field for every molecule, and for every combination of molecules. However, having separate force fields for each system we are interested in can prevent us to compare the predicted properties. We cannot ascribe with certainty the predicted differences to the

underlying chemical reality, and not the model we used. The origin of the discrepancies might also be the use of different reference data or different functional forms, as well as a different parametrization procedure.

This is the reason why generic force fields have also been developed. They trade some accuracy for a better transferability: we are able to use them for multiple molecular systems with roughly the same accuracy everywhere. The Universal Force Field[137] (UFF) and the AMBER family[138] of force fields for bio-molecules are well-known examples of such generic force fields. Instead of defining a force field for a whole molecule, they are based on *fragments*. For example methanoic acid HCOOH would contain C – H, C – O, C = O, and O – H bond fragments; as well as H – C = O, O – C = O, C – O – H angles and O = C – O – H dihedral angle. The same fragments would also be used in ethanoic acid CH_3COOH, or any other carboxylic acid.

Existing generic force fields are not always optimal for simulating MOFs, mainly because of the coordination bonds existing between the metallic centers and the organic linkers. As metal centers are relatively rare in bio-molecules, the interactions between them and organic molecules are not well described by existing force fields such as the ones in AMBER family. There also have been adaptations of UFF for MOFs[139, 140], that are able to reproduce static properties within 10% for most of the structures. However, reproducing dynamics properties and especially the flexibility of MOFs is harder.

4.3.2 Systematic parametrization of force fields

One way to overcome the previously mentioned issues (long parametrization time, non-reproducible parametrization, trade-off between transferable and accurate force fields) is to use a systematic parametrization algorithm. Using an algorithm will improve the parametrization time although the generated force field should still be validated. We can also design the algorithm to give us reproducible parametrization. Finally, using the same algorithm for all the systems of interest should help with accuracy — as each system has parameters specifically fitted for it — while still allowing comparison between different systems because they would share the functional form and kind of input data with all the other systems.

Several approaches have been developed to generate new force fields for MOFs, such as Quick-FF[141] and MOF-FF[142]. Both use *ab initio* reference data, such as the optimized geometry of the system, and the Hessian matrix at this optimized geometry; usually represented on a base on internal coordinates. The Hessian matrix H contains the second derivatives of the energy with respect to two internal coordinates. To optimize the parameters, they use machine learning algorithms, such as genetic algorithms. Genetic algorithms start from a population of randomly created parameter sets, evaluate a cost function to rank them all. Then — mimicking evolution and natural selection — the weakest sets (the ones with the highest cost) are eliminated, and the remaining sets are combined to create a new generation of parameters set. New generations are created until the cost function is minimal over the whole generation. In this last generation, the best parameter set is set to be the new force field.

MOF-FF is based on the MM3 functional form[143], and some of the parameters are fixed before starting the optimization. In particular, atomic charges are computed directly

from the *ab initio* reference and represented as atom-centered spherical Gaussians, while the dispersion interaction is modeled with a Buckingham potential, using the tabulated parameters for MM3[144]. The remaining intra-molecular parameters are optimized in order to make sure the overall potential reproduces the reference data. The score function Z used for optimization is defined for a parameter set P as:

$$\begin{aligned}
\mathrm{Z(P)} = {} & \alpha_{\text{bonds}} \sum_{\text{bonds}} w_i \left[r_i(\mathrm{P}) - r_i^{ref} \right]^2 + \alpha_{\text{angles}} \sum_{\text{angles}} w_i \left[\theta_i(\mathrm{P}) - \theta_i^{ref} \right]^2 \\
& + \alpha_{\text{dihedrals}} \sum_{\text{dihedrals}} w_i \left[\varphi_i(\mathrm{P}) - \varphi_i^{ref} \right]^2 + \alpha_{\text{impropers}} \sum_{\text{impropers}} w_i \left[\delta_i(\mathrm{P}) - \delta_i^{ref} \right]^2 \\
& + \alpha_{\text{diag}} \sum_{m} w_{mm} \left[\mathrm{H}_{mm}(\mathrm{P}) - \mathrm{H}_{mm}^{ref} \right]^2 + \alpha_{\text{non-diag}} \sum_{k} \sum_{m \neq k} w_{km} \left[\mathrm{H}_{km}(\mathrm{P}) - \mathrm{H}_{km}^{ref} \right]^2 \\
& + \frac{1}{9} \sum \left[\mathrm{C}^{-1} \cdot (\mathrm{S} \cdot \mathrm{V})^2 \right]^2
\end{aligned} \tag{4.25}$$

All terms are squared mean deviations, weighted both by w_i to set the relative importance of similar terms; and α_x to set the relative importance of different terms. The first four terms account for equilibrium values of geometric parameters, the following two account for the diagonal and out of diagonal contributions to the Hessian of the system, *i.e.* the dynamics of the system around equilibrium. The last term contains the unit cell matrix C, the stress S and the unit cell volume V, and is used to minimize the resulting total stress on the periodic system.

Johannes Dürholt used MOF-FF to generate new force fields for ZIF-8(CH_3), ZIF-8(Cl), and ZIF-8(Br) from *ab initio* input data. I contributed to this work by providing the previously described simulations of empty ZIFs; and by helping to run analysis on the trajectories in order to validate the force field. Previous versions of MOF-FF used finite representative clusters in vacuum to parametrize the force field. During this work, the fitting strategy was improved to allow the use of periodic boundary conditions in the reference data. This removes the need to find representative and charge neutral finite clusters, which is not always possible depending on the MOF topology. Interested readers should refer to the corresponding article[136] for more information and the force field parameters.

4.3.3 Validating the generated force-fields

After generating the force field, we ran multiple classical simulations to validate it. We first checked the static properties of an energy-minimized structure, starting with the unit cell lattice parameters, reported in table 4.1. MOF-FF is able to reproduce very well the DFT reference used to parametrize it. Since we did not use lattice parameters when optimizing MOF-FF, this was not guaranteed.

We also extracted geometric parameters such as bonds lengths, or angles and dihedral angles equilibrium value; as well as vibrational normal modes from the same energy-minimized structures. We computed vibrational normal modes by using finite differences to compute the full system Hessian in Cartesian coordinates; and then diagonalizing the mass-weighted Hessian to extract normal modes frequencies. These properties are compared to DFT calculations in figure 4.10; and overall the new MOF-FF force field is able

Table 4.1 – Unit cell lattice parameters for the three ZIF-8-based frameworks, comparing the experimental values to DFT and MOF-FF.

		a, b, c (Å)	α, β, γ(°)
ZIF-8(CH_3)	exp[19]	16.99	90.00
	DFT	17.03	90.00
	MOF-FF	17.08	90.00
ZIF-8(Cl)	exp[124]	17.04	90.00
	DFT	17.20	90.00
	MOF-FF	17.21	90.00
ZIF-8(Br)	exp[124]	17.08	90.00
	DFT	17.25	90.00
	MOF-FF	17.25	90.00

to reproduce all of them very well. It overestimates the frequency of vibrational modes after 1500 cm^{-1}. These modes are very localized and involve distortions of the linkers aromatic rings, which are not explicitly described in the force field. A way to improve the description of these modes would be to incorporate cross-terms (terms coupling multiples geometric parameters, *i.e.* a bond length and an angle, …) in the force field.

We then computed elastic constants of the frameworks with both DFT and MOF-FF, the values are presented in table 4.2. Zheng et al. predicted recently the elastic constants of differently functionalized ZIFs in the sod topology[146]. They found that electron withdrawing groups improve the mechanical stiffness of the materials (ZIF-8(CH_3) < ZIF-8(Cl) < ZIF-8(Br)). Although the absolute numbers of our DFT calculations differ up to a few GPa, this trend is reproduced both by our DFT and force field calculations.

Finally, we also ran constant temperature classical molecular dynamics using the generated force field with the DL_POLY software. From these simulations, we computed the same swing angle distribution as in section 4.2.2, which are represented in figure 4.11. The general shape of the DFT distributions is reproduced reasonably well by MOF-FF, but the central value of the Gaussians is shifted by 5° for ZIF-8(Cl) and ZIF-8(Br); and by 10° for ZIF-8(CH_3). However, these differences are of the same order of magnitude as the differences between the distributions generated by using a different DFT functional such as BLYP[130].

In conclusion, the generated force field is able to reproduce well both geometric properties and dynamic properties of the three ZIF-8, even if it was not explicitly parametrized with these properties. It should thus be possible to use these force fields in simulations where a good description of the flexibility of ZIF-8 is required.

Table 4.2 – Elastic constants for the three ZIF-8-based frameworks, comparing the values computed by MOF-FF and DFT.

		C_{11} (GPa)	C_{12} (GPa)	C_{44} (GPa)
ZIF-8(CH_3)	DFT[145]	11.04	8.32	0.94
	MOF-FF	8.54	6.55	0.62
ZIF-8(Cl)	DFT	9.23	7.35	0.86
	MOF-FF	9.92	7.84	0.46
ZIF-8(Br)	DFT	10.33	8.31	0.88
	MOF-FF	10.51	8.65	0.19

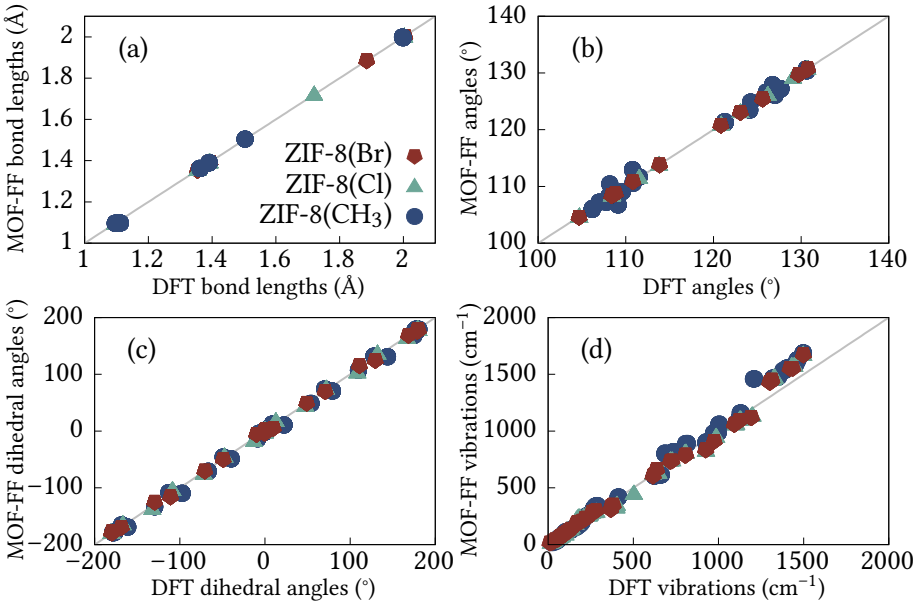


Figure 4.10 – Comparisons of (a), (b) and (c): geometric parameters and (d): vibrational normal modes between the reference DFT data and the new MOF-FF force field.

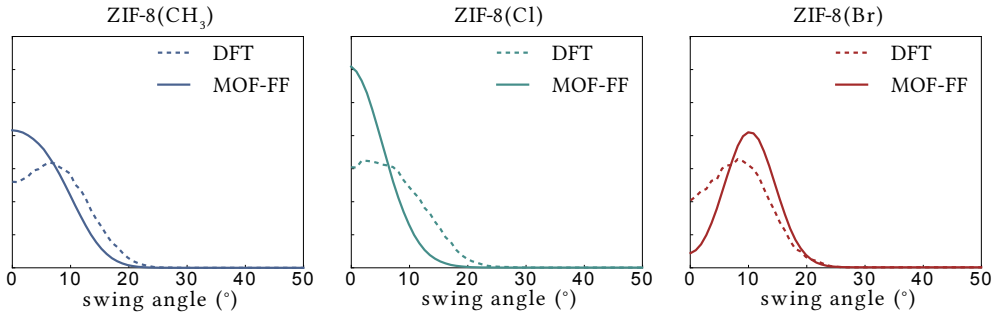


Figure 4.11 – Comparison of dihedral swing angle distribution in empty frameworks for the reference DFT simulations and the new MOF-FF force field.

Conclusions

First principles simulation methods are based on the Schrödinger equation. Because solving this equation directly or numerically is difficult and limited to very small systems, we use alternative methods such as DFT to find approximated solutions. The main advantage of using such first principles methods is that we don't need to make additional a priori approximations on the shape of the energy surface.

I used DFT with molecular dynamics to study the adsorption of nitrogen in the ZIF-8 MOF, and two other functionalized ZIF: ZIF-8(Cl) and ZIF-8(Br). I have shown that in ZIF-8(CH_3) and ZIF-8(Cl), the linkers rotate around the plane of the 6 member window when the nitrogen loading increases, creating a second step in the adsorption isotherm. However, this linker rotation does not increase the accessible porous volume in ZIF-8. Instead, it allows the nitrogen molecules to re-organize inside the pore, from a packing to another. This repacking *in the same accessible volume* is at the origin of the second step in the nitrogen adsorption isotherm. At the same time, the whole process of repacking and linker rotation does not happen in ZIF-8(Br), probably because the pores are smaller in ZIF-8(Br) compared to ZIF-8(CH_3) and ZIF-8(Cl). This is also seen in the nitrogen adsorption isotherm which does not present the second step.

I also used DFT calculations in collaboration with Johannes Dürholt and Rochus Schmid to parametrize a classical force fields for the same structures. Classical forces fields have the drawback of being less precise than *ab initio* calculations, and are usually not able to describe bonds creation and rupture. They nonetheless have the advantage of being a lot cheaper to use to compute the energy of large structures, allowing a better sampling of the phase space. In the next chapter I will describe how I used classical simulations to study the adsorption and intrusion of water and related fluids in hydrophobic porous systems.

5

CLASSICAL SIMULATIONS

First principles molecular simulations allow us to describe and simulate any system while making a minimal number of hypotheses on the shape of the potential energy surface. Unfortunately, they are limited by the computing power currently available to systems containing less than a few hundred atoms, and to time scales of a few hundred picoseconds. If we want to study larger systems (up to a hundred of thousands of atoms) or processes occurring over larger time spans(up to a few microseconds) we can switch to classical simulations.

During my PhD, I used classical molecular simulations — both classical molecular dynamics and classical Monte Carlo — to study the behavior of water and electrolyte solutions confined in porous materials. In the first part of this chapter, I will present the intrusion of aqueous electrolytes liquids in ZIF-8, from static, dynamic and thermodynamic points of view. In the second part, I will discuss the adsorption of water in alumino-silicate nanotubes called imogolite.

5.1 Intrusion of electrolytes in ZIF-8

One of the novel applications that have been proposed for hydrophobic nanoporous materials is related to mechanical energy storage or dissipation through water intrusion [47, 147]. In a hydrophobic porous material, the pressure at which external bulk water will enter the pore space is greater than the vapor pressure of water. That is, adsorption happens only in the liquid phase — and this high-pressure adsorption is called *intrusion* [148]. This phenomenon has been extensively studied in inorganic nanoporous materials, such as zeolites[149–152], and more recently evidenced in hydrophobic metal–organic frameworks[51, 153, 154]. Depending on the nature of the nanoporous material and the strength of the host–guest interactions, intrusion curves can have different shapes and be classified as either as a molecular spring, a shock absorber, or a bumper. These behaviors are illustrated in figure 5.1. This classification depends on the level of hysteresis during an intrusion–extrusion cycle: springs do not present hysteresis, shock absorber present a moderate amount of it, and bumpers do not present a reversible behavior.

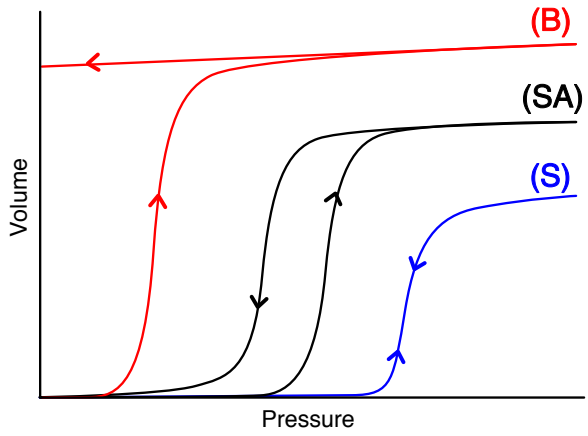


Figure 5.1 – Schematic representation of the 3 different types of energy storage behavior: spring (S), shock absorber (SA) and bumper (B). The intruded volume of liquid is represented as a function of applied pressure.

A very sought-after property related to the intrusion of water in hydrophobic frameworks is the ability to tune the intrusion pressure and the amount of hysteresis present. This can be achieved by chemical modifications of the structure of the host material[155], or by changing the nature of the liquid —for example by adding ions to the water[55]. Depending on the size of the ions and porous channels, in some cases only water can enter the nanopores while ions stay in the bulk liquid. In this case, the change in intrusion pressure is directly related to additional osmotic pressure the fluid has to overcome to enter the structure[51]. But in other cases, the addition of ions has a more complex impact on both the intrusion pressure and the shape of intrusion curves. For example, in the pure-silica analog of the β zeolite[156], increasing the electrolyte concentration from 5 mol/L to 10 mol/L only shifts the intrusion pressure, but increasing it again to 15 mol/L changes the overall behavior from a shock-absorber to a spring[52], removing hysteresis in the intrusion-extrusion cycle. Such changes indicate that the interaction between the electrolyte fluid and the host structure is not merely reduced to a simple

effect of size-based exclusion. At the same time, *in situ* X-ray diffraction measurements performed during intrusion–extrusion cycles showed that $MgCl_2$ ions can enter a pure-silica ferrierite[57]. All this evidence points to a more complex effect than pure osmotic pressure when using an electrolyte fluid for intrusion, but these effects are currently not well understood.

Moreover, we know that adsorption of water in the gas phase can induce large structural changes in nanoporous materials[157, 158]. This is in particular true of soft porous crystals[12], such as flexible MOFs[159], exhibiting dynamic frameworks that are able to respond to external stimuli. This has been studied, by both experimental and computational means, on the MIL-53 family of *breathing* frameworks: the presence of water influences the structure of MIL-53(Cr)[160], and is responsible for the occurrence of numerous structural transitions in MIL-53(Ga), as a function of both water vapor pressure and temperature[161, 162]. Relatively little is known, in contrast, on the impact of the intrusion of liquid water — and aqueous solutions — on the structure and properties of flexible MOFs.

During my PhD, I have studied high-pressure electrolyte intrusion in ZIF-8. ZIF-8 is hydrophobic[155] and presents interesting behavior upon intrusion. While osmotic pressure effects do not depend on the chemical nature of the ions but only of their concentration, ZIF-8 shows changes from one energetic behavior to another when the ion nature changes while keeping concentration constant[55]. ZIF-8 behavior can also be tuned chemically, by modifying the nature of the linkers. For example, changing the methylimidazolate to a chloroimidazolate increases the intrusion pressure[163].

However, the exact mechanism and behavior at the molecular level of the electrolytes intrusion in ZIF-8 are still unknown. In this work, I used classical molecular simulations to study the structure, dynamics and energetic implications of confining water and aqueous solutions of LiCl in ZIF-8. Using these simulations, I was able to explore different aspects of the {water, ZIF-8} and {electrolyte, ZIF-8} systems. I describe below the structure of the liquids and the influence of confinement, their dynamics, the mechanical properties of ZIF-8 and the impact of liquid intrusion on them. I also look at the energetic behavior of intrusion, and the thermodynamics of the ions' entry in ZIF-8. This work is published in *The Journal of Physical Chemistry C* (2019)[164].

Computational methods

I ran classical molecular dynamics (MD) simulations and umbrella sampling simulations using the LAMMPS[165] software. The umbrella sampling simulations additionally employed the COLVARS[166] module for collective variables. I used a combination of different force fields for the component of the system: a rigid SPC/E[167] for water, for its ability to describe the dynamics of liquid water and the solvation of ions; a flexible force-field adapted from AMBER by Zheng et al.[168] for the description of the ZIF-8 framework; and a combination of electrostatic and Lennard-Jones potentials for the ions[169]. I used Lorentz-Berthelot mixing rules for cross-terms in Lennard-Jones potential, and Ewald summation to account efficiently for electrostatic interactions. I used a cutoff of 8.5 Å for both the Lennard-Jones potentials and the separation between real space and Fourier space in the Ewald summation.

After initial energy minimization, I carried all simulations in the isothermal-isobaric NPT ensemble with a timestep of 1 fs, using a Nosé-Hoover thermostat with a time constant of 1 ps and a Nosé-Hoover barostat with a time constant of 10 ps. I allowed the barostat to make arbitrary changes to unit cell lengths and tilt factors (fully flexible anisotropic cell), while imposing an isotropic pressure to the system. Unless specified otherwise, I ran all simulations in the NPT ensemble for 10 ns, and only used the last 4 ns for analysis.

I used three different types of systems in this study. First, bulk liquids at different LiCl concentration: 0 mol/L (pure water), 1 mol/L, 5 mol/L, 10 mol/L, 15 mol/L and 20 mol/L — the experimental solubility of LiCl in water at 25 °C is 19.87 mol/L. Then a $3 \times 3 \times 3$ super-cell of ZIF-8, with liquid confined inside the pores, at the same concentrations as the bulk liquid. Finally, I used a $2 \times 2 \times 3$ ZIF-8 super-cell containing pure water together with a 34 Å cubic reservoir of water on top (this system featuring an explicit ZIF-8/liquid interface) for the umbrella sampling simulations. This last system is represented in figure 5.10.

I generated the initial configurations using the packmol software[133], randomly placing the desired number of particles in the system. I started with bulk electrolyte in a 32 Å cubic box containing 750 water molecules for the pure water, and added ions for the different LiCl concentrations. To generate initial configuration of the liquids confined in ZIF-8, I needed to know how many molecules would fit in the ZIF-8 pores. For that, I ran a constant pressure simulation with a reservoir of pure water outside an empty $2 \times 2 \times 3$ ZIF-8 super-cell at 0.5 GPa for 10 ns and counted 75 water molecules by unit cell that entered ZIF-8. I then ran simulations of the bulk liquid at various values of concentration at the constant pressure of 0 GPa and recorded the corresponding particles density. Using the density of pure water, I mapped the 75 molecules per unit cell to an accessible porous volume of 2.286 nm^3, or 46% of the unit cell. From this volume and the density of the bulk liquids, I could compute the number of molecules to put in the $3 \times 3 \times 3$ super-cell for each concentration. The resulting system was a cubic box of 51 Å, containing roughly 13 000 atoms.

5.1.1 Structuration of the liquid

From the point of view of the liquid, the main effect of intrusion is the confinement of the fluid to a pore space of nanometric dimensions, *i.e.* the nanoporous material acts as a host matrix — although a flexible one. I looked at the effects of this confinement on the liquid structure, as a function of the electrolyte concentration. In order to characterize the structure of the liquid and the solvation of ions confined in ZIF-8, I computed radial distribution function for each pair of atom types in the system. Integrating the radial distribution function until the first minimum gives the number of neighbors in the first solvation shell of each atom. The evolution of this number of neighbors as a function of LiCl concentration is presented in figure 5.2, for both the bulk liquid and the confined liquid.

The first thing we can see here is that, for all species (Li^+, Cl^-, and water), the number of neighboring water molecules — *i.e.*, the solvation — decreases as the LiCl concentration increases. At 0 mol/L and 1 mol/L, there are more than enough water molecules for each ion (roughly 55 water molecule per ion at 1 mol/L) so that they can be in their ideal solvation state: 4 H_2O around each Li^+, and 6 H_2O around each Cl^-. But as the concentration

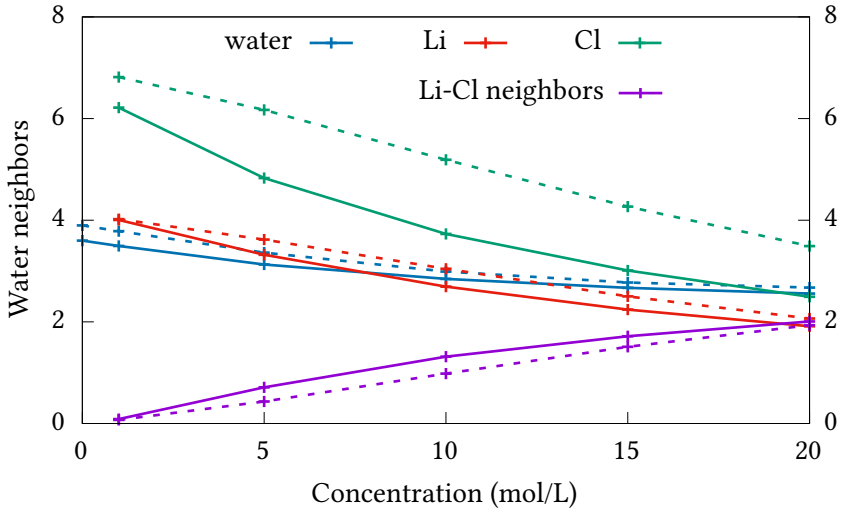


Figure 5.2 – Number of water neighbors in the first solvation shell in the confined liquid (plain lines) and the bulk liquid (dotted lines) as function of the concentration, at the constant pressure of 0 GPa. *The number of chlorine neighbors for lithium ions is also represented.*

increases, there are fewer available water molecules and the ions have to accommodate by having fewer molecules in their first solvation shell. At 20 mol/L, there are only 2.7 water molecules available for each lithium ion; and each cation is thus surrounded by 1.9 water molecules on average, less than half its complete solvation state. The same is true for water/water coordination through hydrogen bonds, as the water molecules compete with the ions to surround themselves with other water molecules.

In the bulk liquid, the loss of neighbors is seemingly linear with the concentration, as all the molecules in the system are able to adapt to find the state of largest possible solvation. In the confined liquid, however, the molecules are geometrically constrained by the presence of the ZIF-8 framework, which manifests as an excluded volume, and thus are not able to adapt as well to the increase in LiCl concentration. Therefore the number of neighbors drops faster. We also note a slight increase in the number of chlorine neighbors of lithium ions at intermediate concentrations, compared to the bulk: the effect of confinement, by diminishing the solvation of the ions, favors the occurrence of anion–cation pairs. The number of lithium neighbors for chlorine ions is the same as the number of chlorine neighbors for lithium ions.

Going from bulk liquid to confined liquid also changes the number of neighbors for water and Cl ions, even at 1 mol/L. In addition to preventing a full reorganization of the water molecules when the concentration increases, the presence of the framework also affects the structure of the solvation shells. Molecules close to the framework can only have neighbors from the liquid on one side — this is an excluded volume effect. However, the framework also has an effect at longer range, the available space in the pores dictating the arrangement of molecules. Instead of being widely distributed, the molecules are restricted to specific preferential locations, due to host–guest interactions. This effect is particularly visible on figure 5.3, and is stronger on Cl/water pairs than water/water or Li/water: the Cl has a larger radius and binds to the hydrogen atoms in water, making its solvation sphere both bigger and “softer”.

I also computed in figures 5.3 and 5.4 the density profiles of atoms in the confined liquid, represented in the xy plane, averaged over z and the $3 \times 3 \times 3$ super-cell. To account for cell deformation during the NPT simulations, I used fractional coordinates to represent the positions of atoms.

Here we clearly see the long distance structuration of water inside the ZIF-8 pores. At low concentration (1 mol/L), the water molecules occupy very well-defined sites, in particular inside the windows between two neighboring cages. As the concentration increases, this organization is perturbed by the ions inserted in the water molecules' network — however, this effect is relatively small, and the water distribution is not greatly affected. The same can be observed in the distribution of chlorine ions, with a well defined, high symmetry distribution at low concentration, but as the concentration increases the distribution of ions becomes more and more distributed over the whole pore space. When it comes to lithium ions, they present a looser arrangement inside the pores and are distributed relatively evenly. Yet they present a preferential occupation next to the water molecules in the 6-member windows (in the diagonal in figure 5.3).

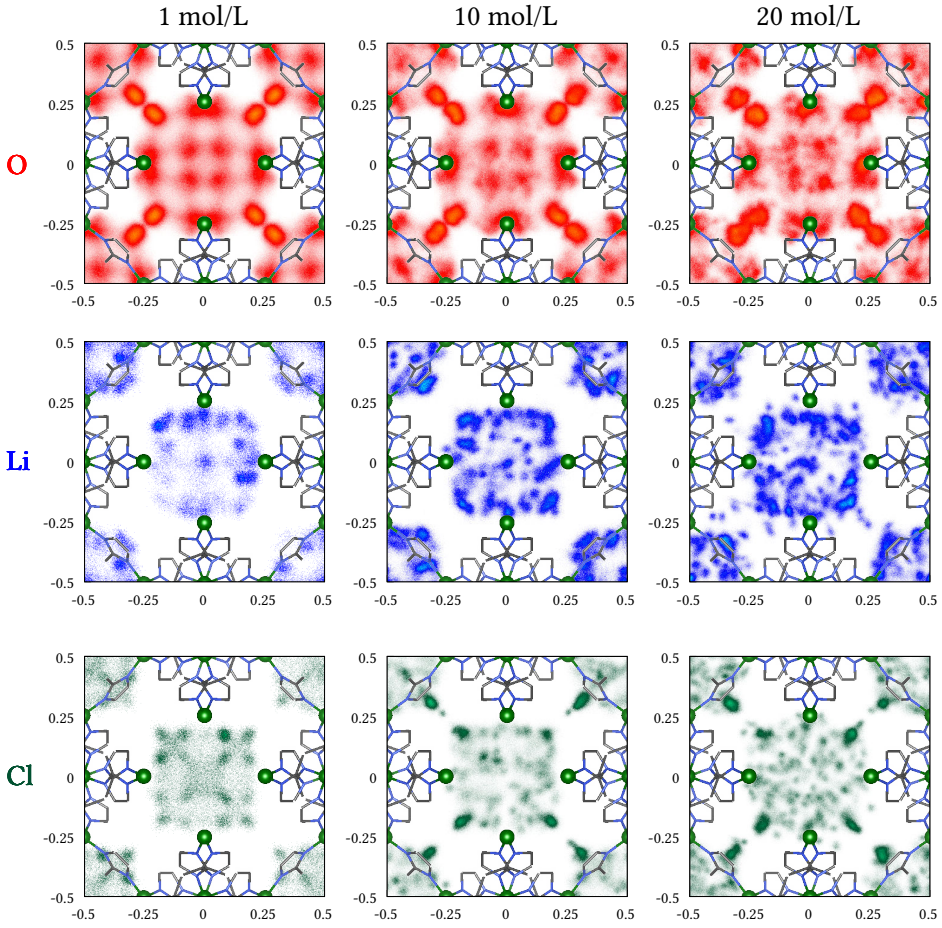


Figure 5.3 – Two-dimensional density profile of oxygen (top), lithium (middle) and chlorine (bottom) atoms in LiCl electrolyte confined in ZIF-8 at 0 GPa *as a function of LiCl concentration (left to right). The atoms from the ZIF-8 framework are superimposed, with some linkers omitted for clarity between the four central zinc atoms.*

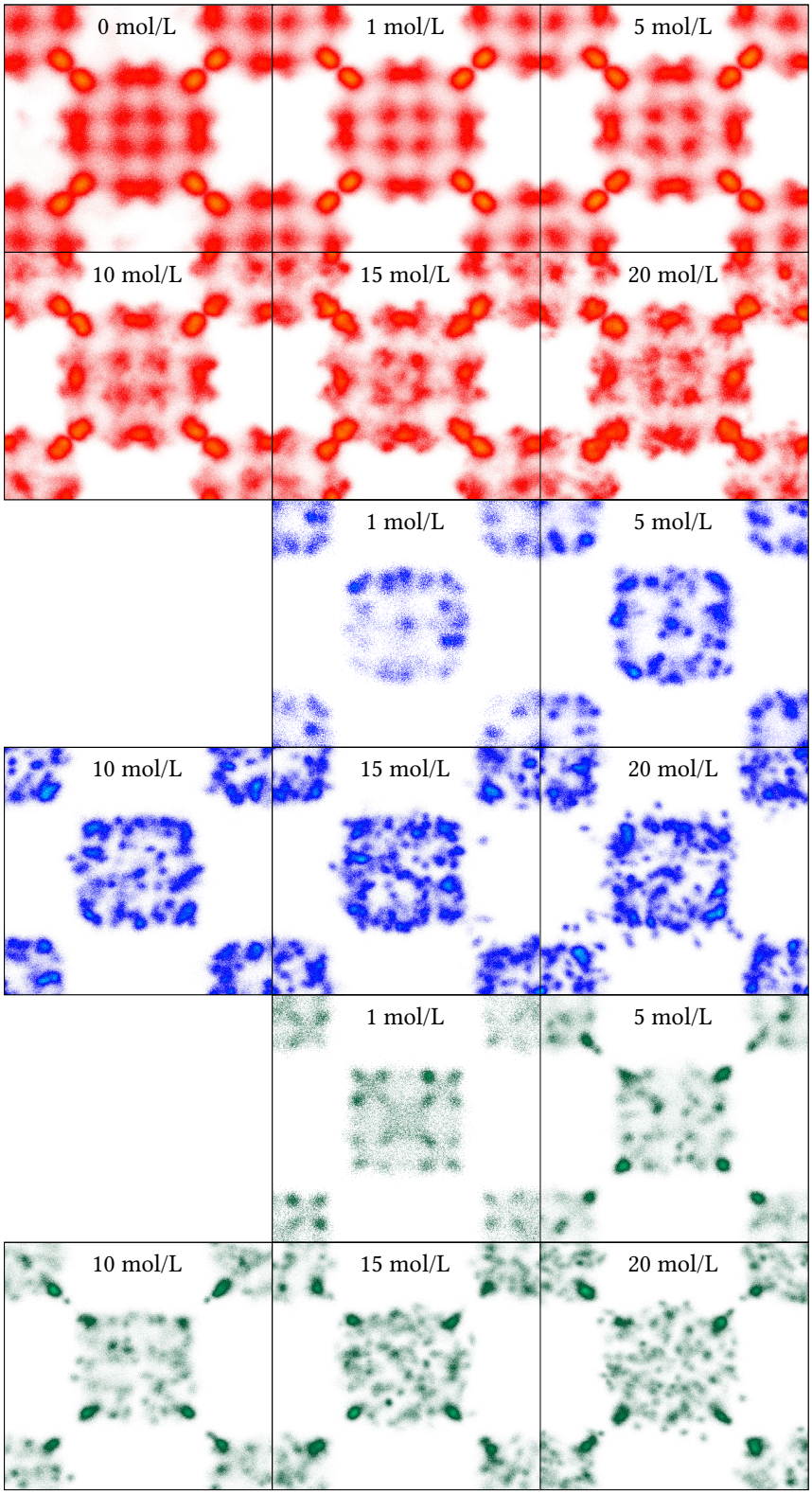


Figure 5.4 – 2 dimensional density profile of oxygen (red), lithium (blue) and chlorine (green) atoms in LiCl electrolyte confined in ZIF-8 at 0 GPa *as a function of LiCl concentration. Every graph is presented in fractional coordinates between -0.5 and 0.5.*

These results can be explained by both the difference in kinetic radius[170] for water molecules (2.65 Å), chlorine ions (3.2 Å) and lithium ions (2.1 Å), as well as the strong attraction between water oxygen atoms and lithium. The difference in size allows lithium ions to fit in smaller spaces, and an even distribution of ions will increase the total entropy of the system. At the same time, chlorine ions and water molecules are more polarizable than lithium, and as such will have stronger interactions with the aromatic linkers, making it preferable for them to take the highly organized arrangement we observe.

5.1.2 Dynamics under confinement

I used two different indicators to quantify the dynamics of water molecules confined in ZIF-8 and the impact of LiCl concentration and pressure on this dynamics. The first one is based on the lifetime of hydrogen bonds between water molecules. Following Luzar et al.[171], I characterize the presence of a hydrogen bond between two water molecules based on a purely geometric definition: two oxygen atoms separated by less than 3.5 Å, with an oxygen–oxygen–hydrogen ($\widehat{\mathrm{OOH}}$) angle less than 30°. I then computed the time autocorrelation function of the hydrogen bond existence functions $\mathrm{H}(t)$ — set to 1 if the bond exists at time t, 0 if it does not exists — as:

$$\mathrm{C}_{\mathrm{hbonds}}(t) = \left\langle \mathrm{H}(t_0) \cdot \mathrm{H}(t_0 + t) \right\rangle_{t_0} \tag{5.1}$$

The decay of this autocorrelation function, presented in figure 5.5, is characteristic of the dynamics of the hydrogen bond network and the lifetime of individual hydrogen bonds.

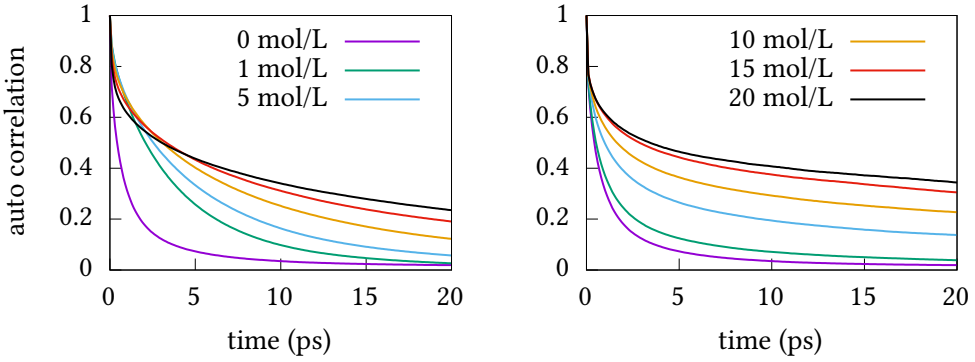


Figure 5.5 – Hydrogen bonds existence autocorrelation in bulk (left) and confined (right) electrolyte as function of LiCl concentration.

This decay is not adequately described by a pure exponential model, and therefore I fitted all the autocorrelation functions with bi-exponential functions:

$$f(t) = \mathrm{A}_1 e^{-t/\tau_1} + \mathrm{A}_2 e^{-t/\tau_2}, \tag{5.2}$$

where τ_1 and τ_2 are the two time scales of decay, and A_1 and A_2 are their relative weights. The resulting fit parameters are presented in table 5.1 and figure 5.6.

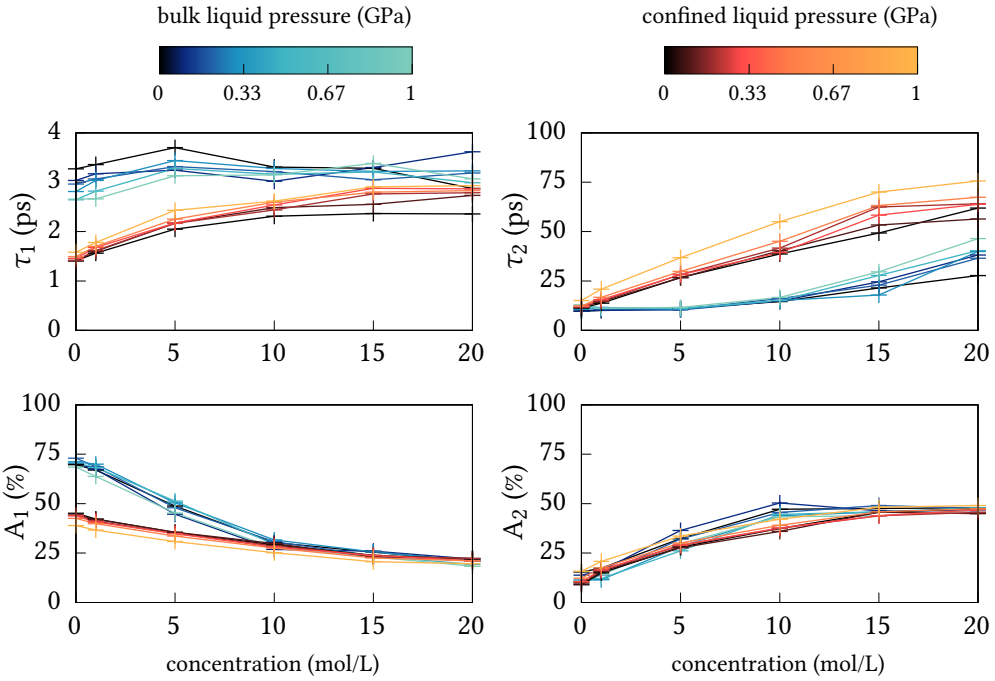


Figure 5.6 – Variations of the time constant and weights of the bi-exponential hydrogen bonds autocorrelation decay as function of the pressure in bulk (blue shades) and confined (red shades) liquids.

This geometric definition of hydrogen bonds has a minor drawback: small fluctuations of the atomic positions, near the cut-off values, can be mistaken for hydrogen bond forming and breaking. To overcome this issue, I also computed the time autocorrelation of the orientation vector $\boldsymbol{u}(t)$ of water molecules:

$$C_{rot}(t) = \left\langle P_2(\boldsymbol{u}(t_0) \cdot \boldsymbol{u}(t_0 + t)) \right\rangle_{t_0}, \tag{5.3}$$

where $P_2(x)$ is the second order Legendre polynomial $P_2(x) = 1/2\ (3x^2 - 1)$[172]. The resulting curves and fit coefficients are presented in figure 5.7 and table 5.2 respectively. As water is a strongly associated liquid, breaking a hydrogen bond is predominantly correlated to rotational jumps, both autocorrelation decay in very similar ways, and I will mainly focus the discussion on hydrogen bonds dynamics.

Table 5.1 – Fit coefficients for the hydrogen bonds autocorrelation decay at 0 GPa.

	Bulk				Confined			
	τ_1 (ps)	τ_2 (ps)	A_1	A_2	τ_1 (ps)	τ_2 (ps)	A_1	A_2
0 mol/L	3.27	9.84	69.7%	15.4%	1.41	11.3	45.2%	8.94%
1 mol/L	3.36	10.1	67.0%	17.3%	1.56	13.8	42.3%	15.0%
5 mol/L	3.70	11.1	48.8%	32.4%	2.05	26.8	35.4%	28.0%
10 mol/L	3.31	14.6	30.3%	47.2%	2.31	38.7	29.1%	37.4%
15 mol/L	3.28	21.5	25.5%	47.6%	2.37	49.3	23.8%	45.7%
20 mol/L	2.87	27.7	20.9%	47.9%	2.36	61.9	22.0%	47.5%

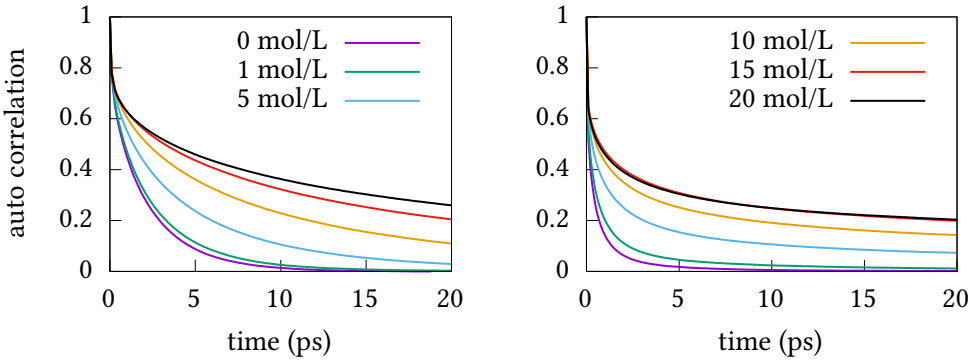


Figure 5.7 – Rotational autocorrelation in bulk (left) and confined (right) electrolyte as function of LiCl concentration.

Table 5.2 – Fit coefficients for the rotational autocorrelation decay at 0 GPa.

	Bulk				Confined			
	τ_1 (ps)	τ_2 (ps)	A_1	A_2	τ_1 (ps)	τ_2 (ps)	A_1	A_2
0 mol/L	2.69	1.05	55.6%	17.2%	0.85	4.65	32.2%	5.0 %
1 mol/L	1.84	4.02	40.3%	30.0%	1.23	10.9	31.5%	6.2 %
5 mol/L	2.42	7.69	33.5%	37.1%	1.92	24.3	30.6%	15.9%
10 mol/L	2.80	13.8	24.8%	45.8%	2.33	33.6	27.1%	25.3%
15 mol/L	3.24	22.8	21.4%	48.5%	2.62	47.1	24.7%	30.1%
20 mol/L	3.34	31.8	20.8%	48.3%	2.54	51.0	23.7%	29.7%

First, we can see on figure 5.6 that the pressure has a relatively small influence on the hydrogen bonds dynamics. Pressure could be expected to have an impact on the dynamics through excluded volume effects[172]: a molecule needs to rotate and temporarily occupy "more space" when creating and destroying hydrogen bonds. This additional space might not be available under high pressure. The small impact of pressure on hydrogen bonds dynamics here has to be related with the relatively small volume changes as the pressure increase seen in figure 5.9: they are too small for an excluded volume effect to have a significant impact.

In the bulk liquid, the shortest lifetime is almost constant around 3 ps as the concentration increases, but the weight of this fast process (A_1) decreases. This suggests that this fastest lifetime is associated with hydrogen bonds between water molecules surrounded only by other water molecules. As more and more ions are added in the system, water molecules are less likely to be surrounded only by other water molecules — as is indicated by the results I presented in section 5.1.1. This lifetime is smaller in the confined liquid than it is in the bulk phase, which has already been shown for water at interfaces: Fogarty et al.[172] showed that it is linked to the librational motions of the O – H bonds, and the dynamics of these *dangling* OH groups at interfaces is faster than in the bulk liquid[173]. On the other hand, the second lifetime, associated with the slowest process, increases with concentration, as well as the corresponding weight (A_2). This points to hydrogen bonds between water molecules bounded to ions.

In the confined liquid, weights still evolve in the same way with respect to the concentration, which points to them being associated with the same kind of hydrogen bonds. The second lifetime increases in the confined liquid compared to the bulk liquid. This slowdown of water dynamics under confinement is well-known[172], and has been observed in many classes of nanoporous materials[160, 174–176]. It is attributed to the stronger organization of water as well as water molecules finding fewer partners for hydrogen bond exchange. The same arguments also apply to the increase in LiCl concentration, both in the bulk and confined liquid. This is coherent with what we see on the changes of bulk modulus as a function of concentration in table 5.3.

5.1.3 Deformations of the framework

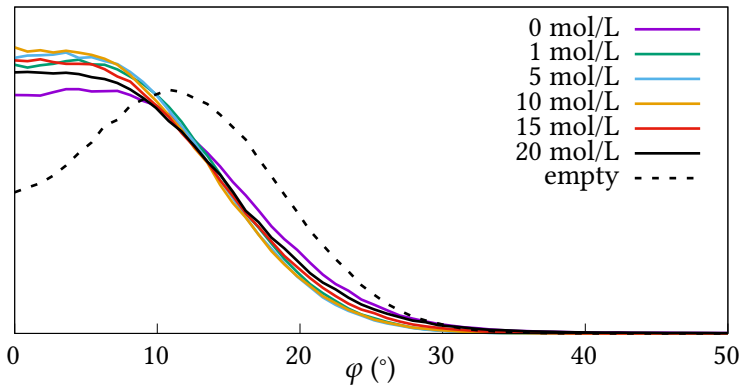


Figure 5.8 – Swing dihedral angle distribution in empty and intruded ZIF-8 at all the LiCl concentrations.

So far, I have studied the properties of the confined liquid inside the ZIF-8 framework, and the changes in the structure and dynamics of this liquid as it goes from a bulk to a confined state. It is also interesting to look at the changes the framework undergoes when going from an empty state to an intruded state. Macroscopic changes to the volume are relatively small (less than 4%), and presented in the next section, especially in figure 5.9. In this section, I will look at the internal deformations of the framework under intrusion. I used the distribution of the Zn – Zn – Zn – CH_3 dihedral swing angle to characterize the deformations of the framework under intrusion. The distributions for empty and intruded ZIF-8 are presented in figure 5.8.

The distribution of swing angles for the empty framework reproduces the results from chapter 4, obtained at a different level of theory: it is a Gaussian distribution, centered around the equilibrium value of 15°, and with fluctuations of the order of 10°. The distributions for ZIF-8 containing the electrolyte liquid do not seems to depend on the concentration. Instead, they are all centered around 5° and still have the same order of magnitude for fluctuations. This is to be contrasted with the effect of nitrogen adsorption at 77 K in ZIF-8; where the equilibrium value goes from 10° to 25° as the pores fill up. It is however interesting to note that the presence of ions does not impact the structure more than the presence of pure water does. This is coherent with the results presented in figure 5.3, where the water molecules are located inside the 6-member windows. Water molecules take the place of linkers that rotate to accommodate them in the windows. As ions never enter this window, they don't have any effect on the linkers rotation.

5.1.4 Elastic properties

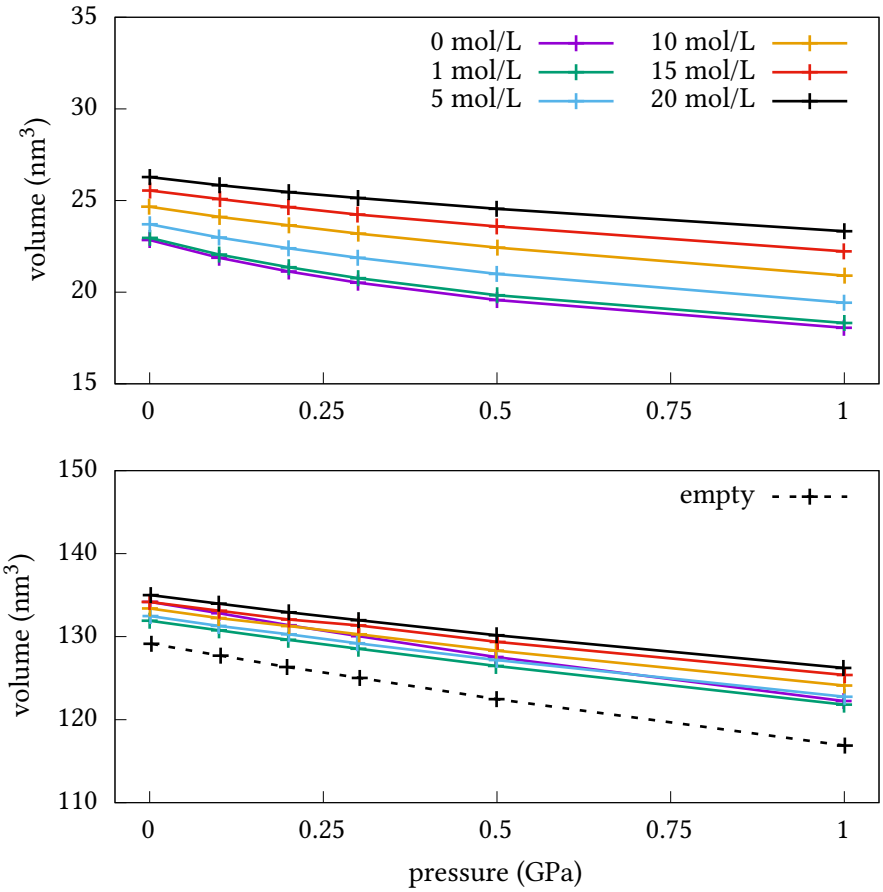


Figure 5.9 – Changes in volume of bulk liquid (top) and ZIF-8 with confined electrolyte (bottom) as function of the pressure. Lines connect the different simulations using the same concentration.

Given its potential applications in mechanical energy storage, one of the most important properties of the ZIF-8/electrolyte system is its mechanical behavior, and notably its stability under high pressure. As increasing stress is applied to a material, it will first deform in a reversible and linear fashion, in what is called the linear elastic regime. Under higher stress, the material will start to deform in an irreversible way — called *plastic deformations* — and finally break. Information on the linear elastic regime and its extent are useful to study the stability of materials under stress, as stiffer materials are generally tougher and able to bear higher stress before failing. This is particularly important for soft nanoporous materials, where the mechanical stability range is lower than in inorganic porous materials such as zeolites, and where pressure-induced amorphization is common at moderate (sub-gigapascal) pressures. For example, ZIF-4 can undergo reversible amorphization between 0.35 and 0.98 GPa[177], and ZIF-8 becomes amorphous after ball-milling operations[178, 179].

I probed the mechanical response of the ZIF-8/electrolyte system using direct simulations of the system under explicit hydrostatic stress. We can first note that even though I allowed arbitrary changes to the unit cell lengths and tilt factors (anisotropic strain), all

the simulations cells remained orthorhombic on average. Figure 5.9 presents the changes in volume as the pressure increases, for all the liquid concentrations. We can see that up to 1 GPa, the deformation remains in the elastic regime, and that the response is almost linear. Moreover, confining an electrolyte in the ZIF-8 structure do not drastically affect the mechanical properties of the system.

From these curves, I extracted the bulk modulus K of the system, defined as

$$K = -V \left(\frac{\partial P}{\partial V} \right)_{N,T} \tag{5.4}$$

The values obtained for all concentrations are reported in table 5.3. The value for the empty ZIF-8 is close to the experimental[145] value of 7.8 GPa. The bulk moduli of pure water and the 1 mol/L liquid are further away from the experimental values[180] of 2.4 GPa and 2.6 GPa respectively. These differences are likely coming from the force-fields we used, which were not parametrized on the mechanical properties of the framework or the liquids. I will mainly discuss the trends here, as they should be reproduced by the force field even if the absolute values are not.

Table 5.3 – Bulk modulus of the bulk electrolyte liquids and of ZIF-8 containing a confined electrolyte liquid.

Concentration	Bulk liquid	Liquid ∈ ZIF-8
0 mol/L	4.3 GPa	11.2 GPa
1 mol/L	4.4 GPa	13.0 GPa
5 mol/L	5.0 GPa	13.6 GPa
10 mol/L	6.1 GPa	14.4 GPa
15 mol/L	7.2 GPa	15.2 GPa
20 mol/L	8.4 GPa	15.4 GPa
Empty ZIF-8		10.5 GPa

Adding water to the pores of ZIF-8 only changes the bulk modulus by a moderate amount (10%), meaning that most of the stiffness comes from the ZIF-8 framework — the stiffer component of the two. However, adding ions to the liquid has a larger effect, both in the bulk state and the confined state, with the bulk modulus increasing by up to 50% at 20 mol/L with respect to the empty framework. Since I generated the structures in such a way that the volume occupied by the liquid is always the same, the increase in bulk modulus is not related to changes in the size occupied by the ions relative to water. Rather, this increase in bulk modulus comes from the stronger interactions between ions and water molecules, compared to interactions between water molecules. Interactions between lithium cations and water molecules are stronger than water–water hydrogen bonds. This will make the liquid less compressible, especially so at high concentrations all water molecules (statistically speaking) are bonded to at least one lithium atom. This is further supported by the fact that the bulk modulus increases by the same order of magnitude ($\approx$ 5 GPa) in both the bulk liquid and the confined liquid in ZIF-8.

Table 5.4 – Average interaction energy in kcal/mol per unit cell for various sub-systems. See the text for the definition of each sub-system. For reference, $k_B T$ is 0.6 kcal/mol at 300 K.

Concentration	E_{Total}	E_{ZIF}	E_{LiCl}	$E_{LiCl/ZIF}$	E^{bulk}_{LiCl}
0 mol/L	−1617	−758.1	−703.1	−155.5	−822.9
1 mol/L	−1859	−739.5	−974.0	−145.8	−1080
5 mol/L	−2753	−738.3	−1864	−149.7	−1995
10 mol/L	−3628	−735.5	−2744	−148.6	−2898
15 mol/L	−4295	−732.5	−3416	−146.0	−3581
20 mol/L	−4818	−728.4	−3949	−140.9	−4117
Empty	−746.8	−746.8			

5.1.5 Thermodynamics of the intrusion

In order to shed light into the thermodynamics of electrolyte intrusion in ZIF-8, I extracted the potential energy for various sub-components of the total system by taking the average value of the interaction energy of the corresponding sub-components. The resulting average energies are presented in table 5.4; where E_{total} is the total potential energy of the electrolyte confined in ZIF-8; E_{ZIF} is the interaction of ZIF-8 with itself; E_{LiCl} is the interaction of the confined electrolyte with itself; and $E_{ZIF/LiCl}$ is the interaction of the electrolyte with the ZIF-8. E^{bulk}_{LiCl} refers to the total potential energy of the bulk electrolyte with the same number of particles as the confined one. Every quantity is expressed for one unit cell of ZIF-8, plus the confined liquid inside.

From these values, I can extract a few thermodynamic quantities of interest, presented in table 5.5:

$$\begin{aligned} \Delta E_{ZIF}(c) &= E_{ZIF}(c) - E^{empty}_{ZIF}; \\ \Delta E_{LiCl}(c) &= E_{LiCl}(c) - E^{bulk}_{LiCl}(c); \\ \Delta H_{intr}(c) &= E_{total}(c) - E^{bulk}_{LiCl}(c) - E^{empty}_{ZIF}. \end{aligned} \tag{5.5}$$

ΔE_{ZIF} is the energetic change in ZIF-8 during intrusion; ΔE_{LiCl} is the energetic change in the electrolyte during intrusion; and ΔH_{intr} is the intrusion enthalpy, i.e. the enthalpy change during the ZIF-8 + liquid $\rightarrow$ liquid $\in$ ZIF-8 process. The sign convention is taken so that all of these energies are negative when the confined state is more stable.

We can see in table 5.5 that the intrusion process has a relatively small impact on the ZIF: the energy difference between the empty and intruded states ΔE_{ZIF} is in the range of tens of kT (at 300 K, $kT \approx 0.6$ kcal/mol). The ZIF-8 framework is slightly destabilized (energetically) in presence of the intruded liquid except at 0 mol/L, where it is slightly stabilized. The overall effect is small compared to the two next trends we see. I have already shown that the presence of ions has little to no effect on the ZIF-8 structure in section 5.1.3.

First, the liquid is always more energetically stable in the intruded phase than in the bulk phase (ΔE_{LiCl}). This might seems strange as ZIF-8 is a hydrophobic material, but the values presented only account for energetic contributions, and do not contain entropy. Figure 5.3 shows that the entropy of the confined liquid can indeed be expected to be

Table 5.5 – Derived thermodynamic quantities in kcal/mol per unit cell. See the text for the definition of each quantity.

Concentration	ΔE_{ZIF}	ΔE_{LiCl}	ΔH_{intr}
0 mol/L	−11.3	−120	−47.3
1 mol/L	7.3	−106	−32.2
5 mol/L	8.5	−131	−11.2
10 mol/L	11.3	−154	16.8
15 mol/L	14.3	−165	32.8
20 mol/L	18.4	−168	45.8

lower than the entropy of the bulk liquid, because of the strong organization of the confined fluid. As the LiCl concentration increase, the intruded phases become more and more stabilized, the ions adding additional rigidity and strong interactions in the pores network.

Secondly, we see that the energetic behavior of the whole process (ΔH_{intr}) is more complex: the intrusion process is energetically favorable for low concentrations ($\leq$ 5 mol/L), and becomes unfavorable at higher LiCl concentrations ($\geq$ 10 mol/L). The interaction between the liquid and ZIF-8 ($E_{LiCl/ZIF}$ in table 5.4) makes for the difference between ΔE_{ZIF} + ΔE_{LiCl} and ΔH_{intr}. Even if the process is energetically favorable at low concentrations, intrusion is not spontaneous because of the entropy contribution to the Gibbs free energy, which makes the adsorption overall thermodynamically unfavorable. This balance of effects, computed here from molecular simulations, could be measured experimentally using high-pressure calorimetry, as it has been done for the purely siliceous silicalite-1 zeolite[181, 182] in previous works.

5.1.6 Thermodynamics of ion entry into the nanopores

In the previous sections, I described the behavior of intruded electrolytes in the pores of ZIF-8, using full periodic boundary conditions. Here, I want to investigate the thermodynamics of the process by which individual species (water molecules and ions) can actually enter the nanopores' space, i.e. pass through the windows of the material and its external surface. I have thus modeled an explicit water/ZIF-8 interface, depicted in figure 5.10: the system here contains both water in the bulk state in a 34 Å×34 Å×30 Å reservoir, and water confined inside ZIF-8. I used umbrella sampling simulations and the WHAM analysis method[183] to reconstruct the free energy profile of a single species (Li^+, Cl^-, or H_2O) entering ZIF-8 along the (111) crystallographic axis. I placed the molecule of interest at 5 Å of a window between the bulk and confined water, and the corresponding counter ion on the other side of the ZIF-8 slab (to keep the system neutral yet minimize ion–ion interactions). I ran a total of 121 umbrella sampling simulations for each species, spaced every 0.33 Å. Each simulation ran for 500 ps in the NVT ensemble, using the last step of the previous simulation as the starting configuration. The resulting free energy profile is presented in figure 5.11, together with the average number of neighbors at a given position on the axis (lower panel).

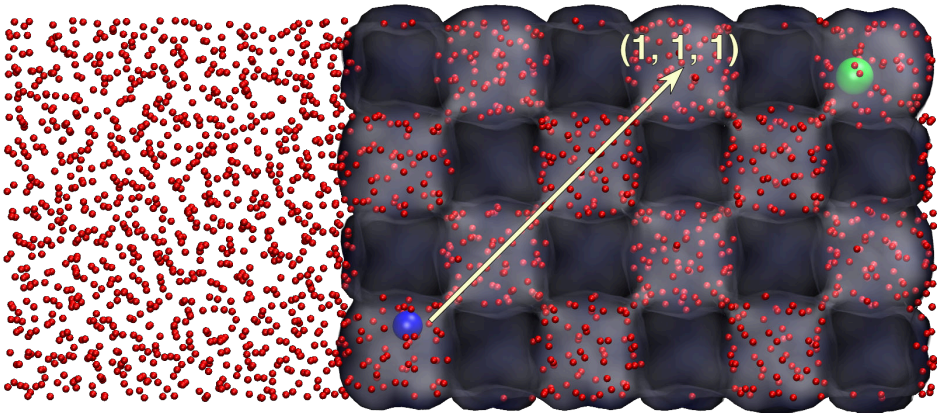


Figure 5.10 – Representation of the system used for umbrella sampling simulations to compute the free energy profile of entry of lithium ions in ZIF-8. We used similar systems for the free energy profile of entry of chlorine and water. Water molecules are represented in red, lithium ions in pink and chlorine ion in green. ZIF-8 is represented as a transparent matrix.

The first conclusion is that there is no free energy barrier for entry of the water molecule: the energy profile is flat and the number of neighbors is constant and around 4. This is consistent with the results already presented on the location of water molecules inside the ZIF-8 windows and on the number of neighbors for water molecules inside ZIF-8. It also confirms that the nature of the liquid-phase intrusion process is not a kinetic limitation of water adsorption, but actually due to thermodynamic hydrophobicity of the framework. For chlorine anions, we observe two barriers on the free energy profile, which correspond to the ZIF-8 windows at 0 and 15 Å. These barriers are correlated to a lower number of neighbors for the anion, dropping to a value of 4: there is not enough space inside the window to fit a chlorine ion and 7 water molecules, and the anion has to partially desolvate to pass through the window — explaining the existence of the free energy barrier. Outside of these barriers, the profile is flat and at the same level as in the bulk liquid, meaning that while the entry of a single chlorine ion is a rare event, at long thermodynamic time scale, Cl ions should be able to enter in ZIF-8. Generally speaking, the Cl ions have a kinetic barrier to entry in the ZIF-8.

The results for Li are more surprising. We see both a high barrier at the first (x = 0 Å) and second (x = 15 Å) windows; and an energetic difference between outside and inside the pores of roughly 15 kcal/mol. This energy difference is not only due to the bulk liquid to confined liquid transition, as it is also present in the transition between before and after the second window. At the same time, these barriers and energy differences are not linked to a difference in solvation as in the chlorine case, as the number of neighbors of lithium stays constant and around 4. Indeed, the solvation of Li^+ by water is much stronger, and its solvation sphere is smaller in size than Cl^-. As lithium does not partially desolvate or rearrange to pass the barrier, the whole solvation sphere needs to go through a relatively small window, thus making the barrier higher. This points to a difference in nature between the Li^+ and Cl^- ions, which will have to be probed further, for example by studies on other ions of different size.

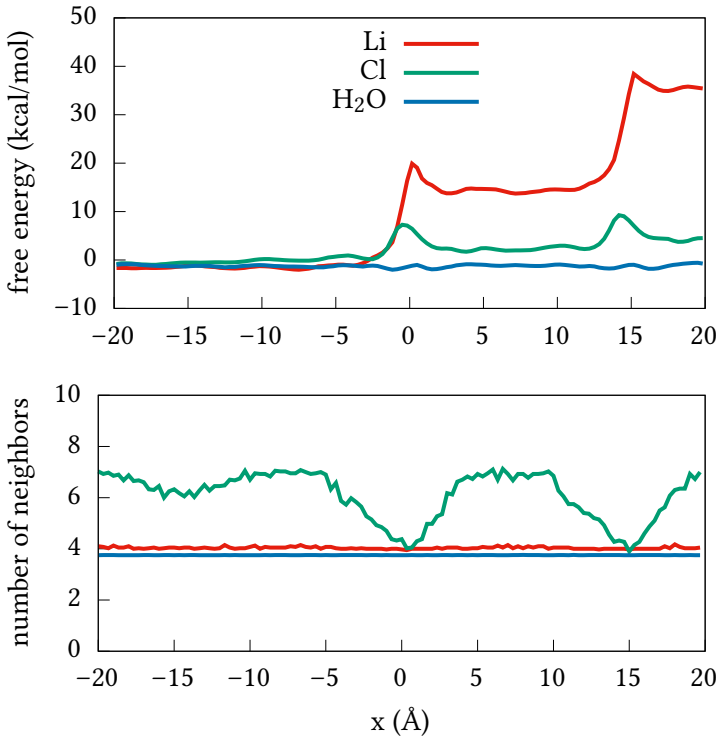


Figure 5.11 – Free energy profile (top) of a single molecule entering ZIF-8 and corresponding number of neighbors (bottom) in the first solvation shell as function of the position of the molecule along the (111) crystallographic axis. The first ZIF-8 windows is at $x = 0$; the $x < 0$ area corresponds to bulk water, and the $x > 0$ area to water-filled ZIF-8. We evaluated the uncertainty on the free energy profile using Monte Carlo bootstrapping[183], and found it to be at most 0.08 kcal/mol *for* H_2O*, and* 0.3 kcal/mol *for* Cl^- *and* Li^+*.*

Conclusion

Intrusion of liquid water and concentrated aqueous solutions in hydrophobic materials have been proposed for applications in mechanical energy storage and dissipation, and recently ZIF frameworks have been highlighted for the high energy density that they can store. However, while the process of intrusion has been well studied in various zeolitic materials over the last 20 years, there is relatively little information available on the behavior — at the microscopic scale — of water and electrolytes in hydrophobic metal–organic frameworks. These systems are difficult to probe experimentally because liquid intrusion occurs under high pressure. Therefore, I have used molecular dynamics simulations to shed some light onto the properties of LiCl aqueous solutions at various concentrations confined inside the pores of the ZIF-8 metal–organic framework. I show that the presence of the electrolyte has a moderate impact on the ZIF-8 framework, while the presence of the ZIF-8 matrix strongly influences the behavior of the confined aqueous solution, affecting the overall properties of the system. I also computed the free energy profiles for the entry of water molecules and ions into the nanopores, showing a difference between anions and cations.

While this work provides an interesting picture of the LiCl electrolytes in ZIF-8, it also opens a few venues for future research. The main one is the impact of the ion size on the properties of the confined liquid. Experiments have been performed with other ions of larger size, including KCl, and there it is not even clear what fraction of the larger cations (K^+) actually can diffuse inside the nanopores. Computational approaches to these systems will be of great help in rationalizing the experimental results and provide a view of the microscopic mechanisms that are behind them.

Another one is to give a deeper look at the free energy barriers for ions passing through the windows of ZIF-8. While the windows are found, in the gas phase, to be very flexible and let diffuse molecules of large diameter (up to butane), we find that the entry of solvated species, such as ions in water, can be linked to a significant free energy barrier. Our free energy simulations of this process will have to be extended to other ions, in order to probe the influence of the size of both the ion and its solvation shell, but also to look at the influence of electrolyte concentration on the free energy profiles. Initial tests in this direction have shown that it should be technically possible, but convergence in such highly constrained systems is very difficult to achieve.

Finally, this work focused on the ZIF-8 framework, perhaps the most archetypal of the ZIF materials. The influence of framework functionalization with various imidazolate derivatives, which has shown to greatly impact adsorption in the gas phase, will surely also manifest itself in the liquid-phase intrusion processes. This is currently a wide open question, which is just starting to get studied[163].

5.2 Adsorption of water in imogolites

I worked with Laura Scalfi during her master's internship (*stage de M2* in France) in the group to study water adsorption, confinement, and structuration in alumino-silicate nanotubes called imogolites. This study was published in *Langmuir* (2018)[176], and is described in this section.

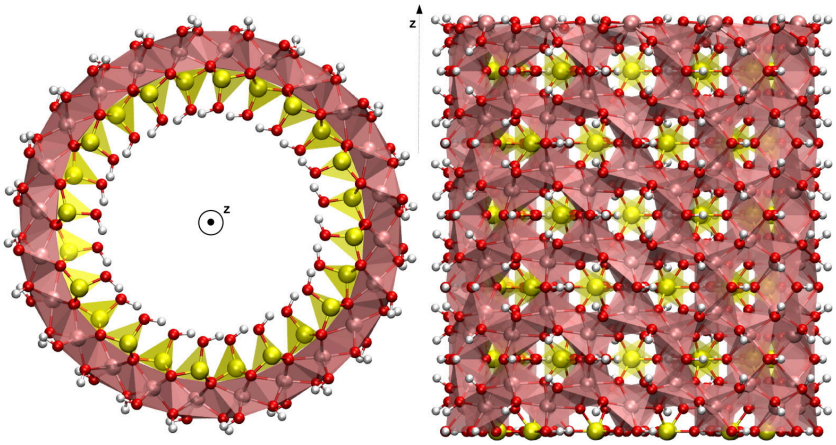


Figure 5.12 – Imogolite nanotube ($n = 12$) with three unit cells along the tube axis z. Color code: aluminum (pink), silicon (yellow), oxygen (red), hydrogen (white). Note the hydrogen bonds between silanol groups on the internal surface (left panel).

Imogolite is a natural alumino-silicate material, with formula $Al_2SiO_3(OH)_4$, that was first discovered in volcanic ashes in Japan[184]. It is the only known alumino-silicate material that spontaneously forms inorganic nanotubes, and there exists no planar equivalent material in nature — unlike carbon nanotubes, of which graphene is the planar form. Imogolite nanotubes are monodisperse in diameter, and can be readily synthesized with controlled length and diameter, for example by substituting silicon with germanium [185]. This precise control of the imogolite nanotube dimensions is interesting for applications that rely on one-dimensional pore channels, in fields such as nanofluidic devices, membranes for filtration and separation, desalination, etc. Moreover, from a theoretical point of view, its hollow cylindrical topology and tunable size make it a very attractive model to study the properties of fluids under confinement. Their structure was first described by Cradwick et al.[186] from electron diffraction measurements as a cylindrical assembly of silicon tetrahedra and aluminum octahedra (see figure 5.12). This sheet spontaneously folds into a nanotube because of bond length mismatch and formation of intra-molecular hydrogen bonds[187, 188]. The imogolite nanotubes are characterized using the same nomenclature as carbon nanotubes. Both natural and synthetic imogolite nanotubes exhibit a zigzag folding, with a variable number n of $Al_2O_3SiOH(OH)_3$ units (called gibbsite units) along the nanotube circumference. The cylindrical unit cell then contains $2n$ gibbsite units. Natural imogolite is a nanotube with size $n = 12$, while depending on the details of the synthesis conditions, synthetic imogolites correspond to values of n typically between 12 and 14, and sometimes larger.

Although imogolites have been the subject of several experimental studies, only a few theoretical works have been carried out until now. Most of the theoretical studies on imogolite nanotubes focus on the energetics of empty nanotubes. The sharp monodisper-

sity in nanotube diameters has been explained by both quantum chemistry and classical studies, that show the strain energy has a minimum for a given nanotube diameter, contrary to carbon nanotubes where the energetically favorable structure is that of infinite diameter (*i.e.*, the graphene slab[189–192]). Several studies also focused on computing vibrational spectra[189, 193, 194] and studying the energetics and dynamics of the rolling of the nanotubes themselves[187, 188, 195]. Finally, other works have pointed at more complex aspects, such as defects[196] and deformation of the nanotubes[189, 197, 198]. There is still relatively little data — experimental or computational — on the hydration of these nanotubes and the behavior of the confined water molecules inside the pore space. Imogolite nanotubes differ markedly from the more common carbon nanotubes in both geometry and chemistry, and the behavior of water inside their pores is still very much open. It is experimentally challenging to address these issues, especially differentiating the water inside the nanotubes from the water outside, and a computational approach is therefore a natural complement to the published experimental results.

A single imogolite nanotube presents two surfaces, both available for adsorption. The outer surface is composed of a gibbsite-like sheet of aluminum octahedra. The internal surface is formed of silicon tetrahedra with one hydroxyl group exposed and the three other corners of the tetrahedron linked to three aluminum octahedra. Therefore both surfaces are covered with hydroxyl groups that are expected to have a hydrophilic behavior. Based on electrostatic calculations, authors have suggested that the internal surface is strongly hydrophilic whereas the outer surface might be more hydrophobic [190, 196]. Because adsorption on the outer surface of the nanotubes depends strongly on experimental conditions affecting the bundling of the nanotubes, and the spacing between them, we focused on the adsorption and behavior of water confined *inside* imogolite nanotubes. Moreover, the strong curvature and limited space inside the nanotube provide for strong confinement effects on which little experimental data is available.

The internal surface of an imogolite nanotube can be described as a periodic sequence of silicon rings along the nanotube axis. Two adjacent rings are rotated of π/n. When dry, silanol groups within a ring form hydrogen bonds so that all the hydroxyl groups are in a plane normal to the nanotube axis, as shown in figure 5.12. The internal diameter, computed between internal oxygen atoms (noted O_{int}), is 12.8 Å and the external diameter, computed between external oxygen atoms (noted O_{ext}), is approximately 22 Å. As a consequence, the internal pore is not smooth and uniform as for the carbon nanotube, but the pore is rugged: narrower where there are silanol groups and larger between these silicon rings. One can picture the internal cavity as a hollow cylinder with equally spaced circular furrows in the circumference.

5.2.1 Simulation methods

In order to produce imogolite nanotube models, we started from the structure of Cradwick et al.[186] for a $n = 10$ nanotube with C_{2n} symmetry. We formed a flat gibbsite-like sheet by unfolding the nanotube and adding hydrogen atoms (which were not present on the original structure derived from electron diffraction data). We relaxed the slab geometry and unit cell at DFT level, using the PBE-SOL exchange-correlation functional[199] in the CRYSTAL14 software[200]. From this relaxed planar structure, we rolled back a nanotube of size $n = 12$, and relaxed the nanotube structure and cell parameters using periodic DFT

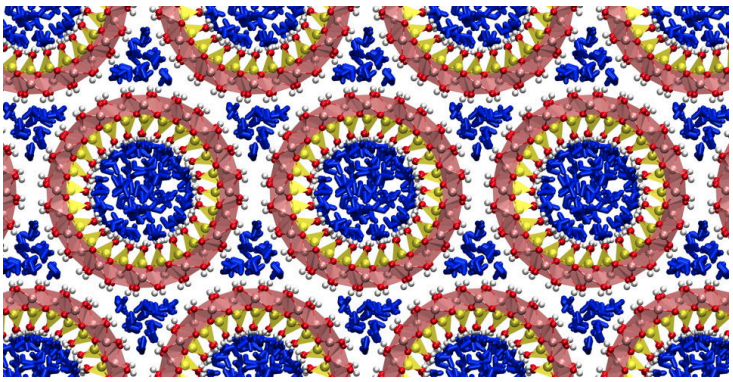


Figure 5.13 – Fully hydrated imogolite nanotube ($n = 12$) with hexagonal packing. Water molecules are represented as blue sticks.

calculations, with large inter-nanotube spacing (60 Å) so that there are no interactions between nanotubes. I show on figure 5.12, two views from top and side of the $n = 12$ nanotube with three unit cells along the z axis.

Imogolite nanotubes can have a length from 10 nm up to a few micrometers and experiments show that they tend to pack into bundles when dry. The periodic pattern seems to be a monoclinic assembly [201, 202] with $\beta \approx 78°$. Given our focus on adsorption inside the nanotubes, following other theoretical studies, we chose to represent the nanotubes in a hexagonal packing that is close to the monoclinic one[193, 194, 203]. Lattice constants are chosen to ensure close contact (but no overlap) between neighboring nanotubes, at $a = b = 24.2$ Å. This choice is consistent with experiments and other theoretical studies. Along the tube axis, we studied super-cells with lattice parameters $c = n \times 8.486$ Å where $n = 1$, 3 or 5 (where the value of 8.486 Å was determined by geometry optimization with variable cell).

We used classical simulations to study the structure and dynamics of water in imogolites at large timescales. We ran MD simulations in the canonical NVT ensemble with the LAMMPS code[165], using a timestep of 0.5 fs. After 100 ps of equilibration, we collected trajectories from 200 ps to 50 ns, depending on the properties studied in each case. We also performed a series of Grand Canonical Monte Carlo (GCMC) simulations to study the adsorption behavior of the nanotubes. In the gas phase, we considered water to be an ideal gas, and the chemical potential is then easily linked to the gas pressure[150].

To describe the interactions of the nanotube, we relied on the CLAY-FF force field[204], which has been extensively used in the literature for clay and other alumino-silicates[194, 195, 203]. Water was described in the flexible Single-Point Charge (f-SPC) model[205], which is naturally suited for coupling with CLAY-FF. CLAY-FF is a general force field developed to model clay minerals, that have the same chemical nature as the imogolite. It relies almost exclusively on non-bonded interactions using mostly electrostatic (Coulombic) and Lennard-Jones potential to describe the system's interactions. In addition to these non-bonding interactions, CLAY-FF includes a harmonic bond term for hydroxyl bonds (O – H stretching); and a M – O – H harmonic bending potential.

5.2.2 Water adsorption

In order to characterize water adsorption in the gas phase and obtain physically meaningful water uptake values, we performed GCMC simulations for water pressure between 2 and 3600 Pa using two different water models (SPC and TIP4P/2005). For computational reasons, rigid water models were preferred in the GCMC calculations. We observe in figure 5.14 that the imogolite nanotube is hydrophilic, *i.e.* the filling occurs at a pressure below the bulk saturation pressure. This is in agreement with the affinity of water for silanol-rich surfaces, or zeolites with many silanol defects[79]. The value of the pore filling pressure varies strongly with the nature of the model chosen for the description of the water, and the movement allowed for the silanol groups — consistent with the findings of Zang et al.[203]. We find transition pressures of 0.1 kPa for TIP4P/2005, and 1 kPa for the SPC water model. The order of magnitude of those values is within the range of experimental data, for example the gravimetric study of Konduri et al.[194], which features a smooth isotherm with water uptake in the same range.

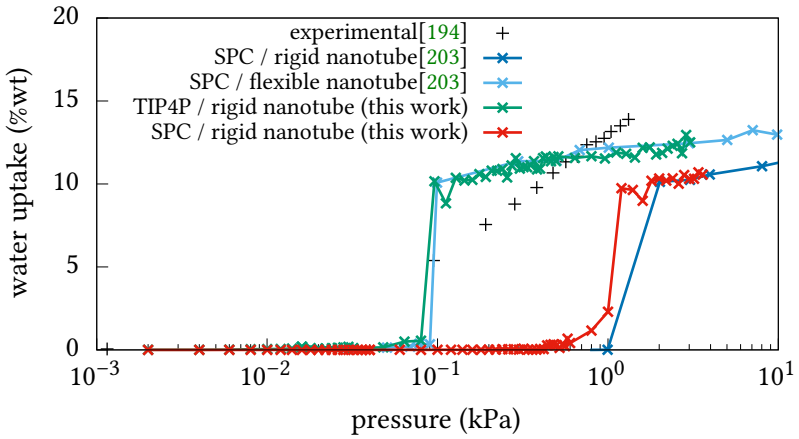


Figure 5.14 – Water adsorption isotherms in imogolite, taken from different experimental and computational sources.

In contrast to the adsorption pressure model dependence, there is excellent agreement between the various models on the saturation uptake of water once the pore is filled, *i.e.* on the density of the water adsorbed inside the imogolite nanotube. Uptake is found to be 10% by weight after the transition, both for the SPC and TIP4P/2005 water models. We thus derived from the GCMC calculations an initial configuration for the simulation of the fully filled imogolite nanotube, from the plateau of the SPC adsorption isotherm. This initial configuration contains three nanotube (n = 12) unit cells along the z axis, 98 water molecules inside the nanotubes and 18 water molecules outside. It is represented in figure 5.13. We used this fully hydrated tube as initial configuration for classical molecular dynamics simulations, with the extended CLAY-FF force field and f-SPC water model.

5.2.3 Structure of confined water

From the analysis of the MD trajectories of the fully hydrated imogolite, we computed density profiles of all atom types (water O_w and H_w; silanol groups Si, O_{int}, and H_{int}). They are plotted in figure 5.15 in the xy plane, *i.e.* as if viewed from the top of the (infinite)

nanotube. From the water oxygen distribution O_w, it is clear that there are two different populations of water molecules: strongly structured water adsorbed next to the internal surface (a first hydration layer, at distances between 4.5 and 5.8 Å from the nanotube's center), and more disordered water filling up the center of the nanotube. This is also clearly visible in the cylindrical distribution functions figure 5.16, where we see that the two populations are not completely separate, *i.e.* the density does not fall to zero in-between. Furthermore, we see a slight deviation of the nanotube from a purely circular form: the position of the Si atoms show a slightly hexagonal deformation, due to the symmetry of the packing of nanotubes in bundles[197]. This deformation is only very small, with Si displacements of 0.1 to 0.2 Å at most, due to the rigid nature of the imogolite nanotube, especially with such a small diameter. This deformation is both linked to the hexagonal packing and the adsorption stress exerted by the water molecules — as shown previously for water/quartz interfaces[206], and more generally for adsorbates in soft porous materials[207, 208]. In the case of imogolite, the stress exerted is counterbalanced by the relative stiffness of the nanotube, although the extent and details of the deformation depend on the packing of the nanotubes. For example, Creton et al.[198], used a larger distance between the nanotubes and observed a transition between an ellipsoid shape to a more cylindrical shape when increasing water densities.

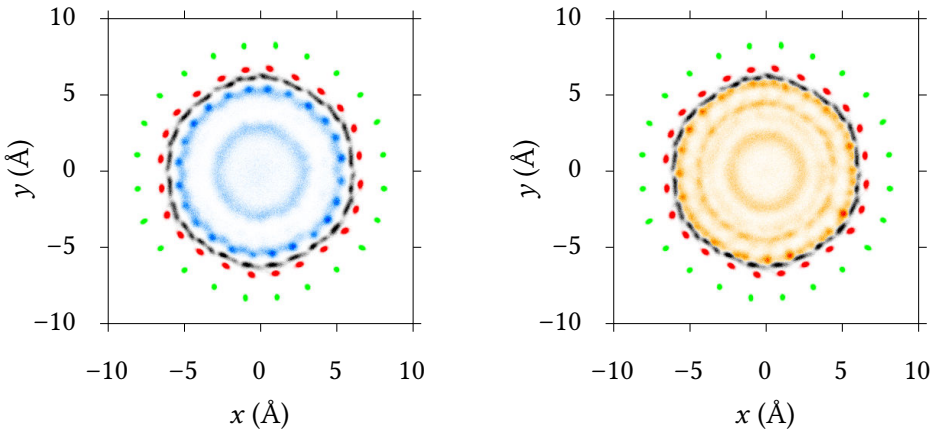


Figure 5.15 – Two-dimensional density profiles in the xy plane for atoms from the silanol groups (Si in green, O in red, H in black) and water molecules (O_w in blue and H_w in orange).

To better understand the structuration of the first adsorbed layer, we plot in figure 5.16 the same densities, but this time as a function of cylindrical coordinates z and r, where r is the distance to the central axis of the nanotube. There, the strongly preferential location for the confined water oxygen atoms becomes clear: the water molecules sit in the wider furrows, situated in-between two rings of silanol groups (SiOH rings are located every 4.25 Å along the z axis, and are depicted on figure 5.16 by dotted lines). This preferential localization of the adsorbed water contrasts sharply with that of water inside carbon nanotubes, whose internal surface is smoother. It was, however, demonstrated in other nanopores of small dimensions with dangling hydroxyl groups accessible to the water for hydrogen-bonding[160].

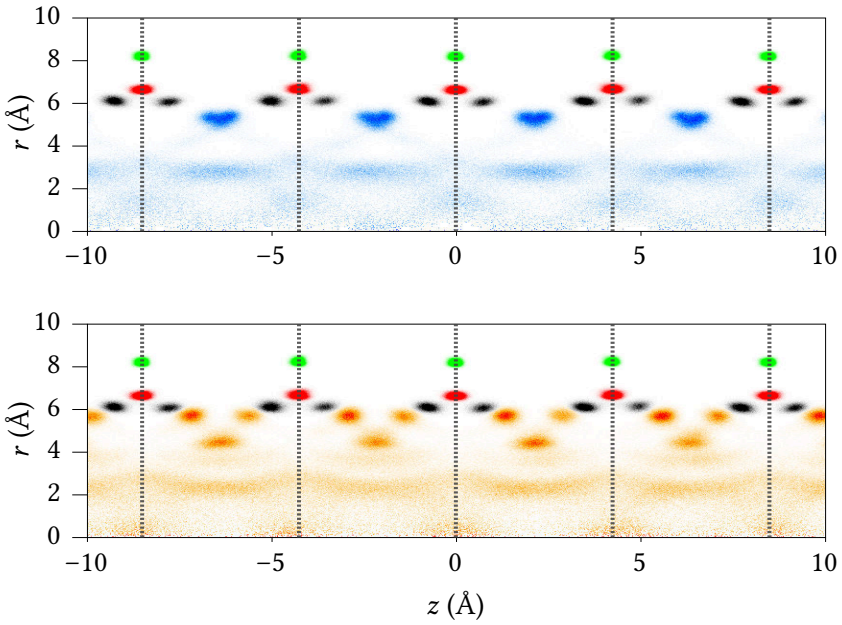


Figure 5.16 – Two-dimensional density profiles (the radial distance to the center of the nanotube r versus the projection along the z axis) for surface atoms (Si in green, O in red, H in black) and water molecules (O_w in blue and H_w in orange). Silicon rings are indicated by dashed lines.

These water molecules are strongly hydrogen-bonded to the silanol groups, as indicated by the well-defined positions for H_{int} atoms, which are rearranged compared to their relaxed position. Indeed, in the dry imogolite nanotube, the most stable position for the internal hydrogen atoms is in the silanol ring plane ($z = 0$), where they are pointing toward a neighboring O_{int} atom. This conformation allows the formation a relatively weak hydrogen bond, because the $O_{int} - H_{int} \cdots O_{int}$ distance is large. In the hydrated tube, the hydrogen atoms' most common position is shifted out of that silanol ring plane (see figure 5.16).

We thus propose the following model to better visualize and understand the structure of the adsorbed water. We model the internal surface of the nanotube as a periodic sequence of triangles, where each silanol group SiOH is a vertex. The Si–Si and O–O radial distribution functions show two first neighbor peaks at 4.2 and 4.7 Å for silicon, and at 3.3 and 4.6 Å for internal oxygen atoms. This means that the triangles are isosceles with two angles of 66.5° and one angle of 47°. This pattern is depicted on figure 5.17. Above the center of each triangle is a potential water adsorption site. However, analysis of the sites shows that no neighboring sites in the same *xy* plane can be occupied at the same time, due to short-distance inter-molecular repulsion of water molecules, so that at most half of the adsorption sites are occupied in the filled nanotube. Then, each of the hydrogen atoms of the SiOH groups that form the vertices of a triangle will point toward one of the three occupied neighboring adsorption sites.

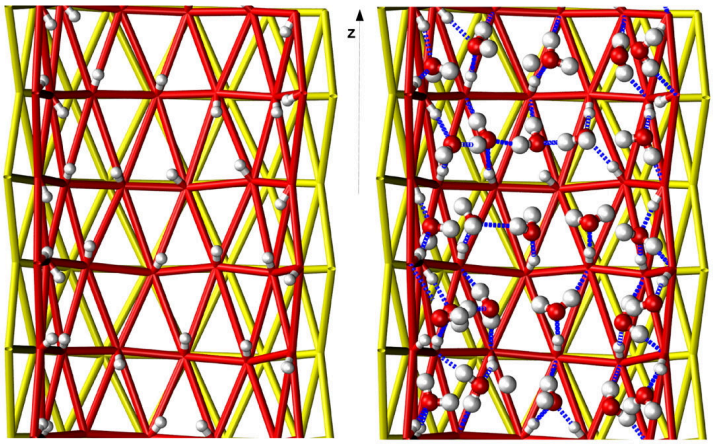


Figure 5.17 – Illustration of the triangular adsorption sites, highlighted by drawing yellow triangles between silicon atoms and red triangles between oxygen atoms.

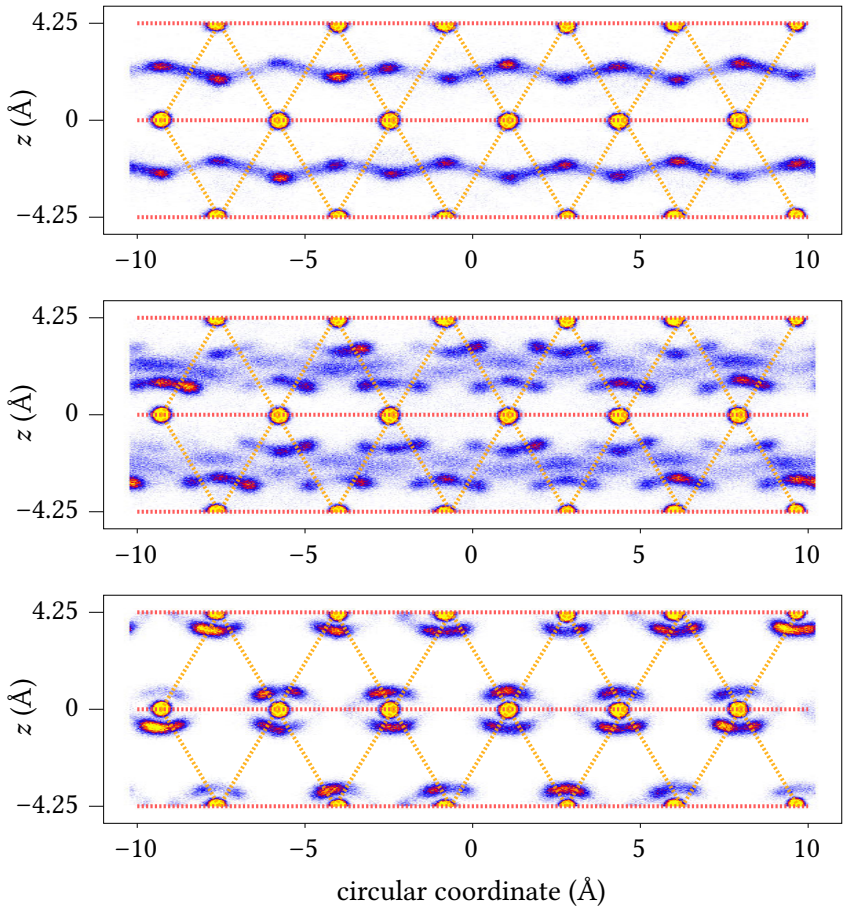


Figure 5.18 – Density profiles on the flattened hydrated nanotube planes for water oxygen (top), water hydrogen (middle), and silanol hydrogen (bottom). The circular coordinate corresponds to a curvilinear abscissa that draws a circle of radius R = 6.5 Å *centered on the axis z. On all the graphs, internal oxygen atoms appear as yellow dots.*

A final visualization can be obtained by a density plot of O and H atoms, where the density is mapped against z and a circular coordinate, as if a slice of the nanotube had been cut and unrolled. These densities are presented in figure 5.18. Red dotted lines show the SiOH rings, while orange dotted lines show the triangular mesh described above. The adsorption sites on top of each triangle are clearly visible. We further see that there is a strong anisotropy of the system, where there is possible circular movement of the water molecules in-between sites in the same z plane, while there is no observed density crossing the silanol rings, as already observed by Creton et al.[198]. This shows that there is surface diffusion of water molecules, but only in the xy plane.

5.2.4 Hydrogen bonding patterns

The localization of the first layer of adsorbed water on the internal surface of the imogolite nanotube, and the strong hydrogen bonds that are created with the tube's silanol groups make the first layer of water strongly ordered. We can see on figure 5.16 (bottom panel) that this includes rotational ordering of the water molecules, with three marked preferential positions for their hydrogen atoms at room temperature. Two of those, equivalent by symmetry, corresponding to the formation of a hydrogen bond donated by the water molecule to the silanol group. The third position corresponds to a hydrogen atom dangling toward the inside of the nanotube. Because the water molecules have 2 hydrogen atoms that can occupy these positions, there are 3 possible geometries, summarized on figure 5.19: a water molecule can be donating two hydrogen bonds (case A); accepting one and donating one (case B); or accepting two hydrogen bonds from neighboring silanol groups (case C).

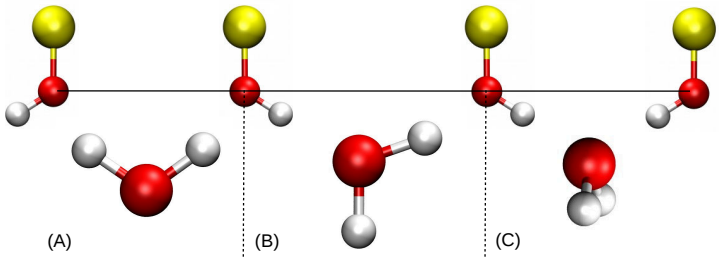


Figure 5.19 – Three possible geometries of water adsorbed in the first layer of the imogolite internal surface.

The average number of hydrogen bonds formed during the trajectory is presented in table 5.6. The change in silanol–silanol hydrogen bonding confirms the structural change upon adsorption: when the nanotube is empty, approximately half of the silanol groups are linked by a hydrogen bond within silicon rings (as detected per our rather strict criteria). When hydrated, only 2% form a hydrogen bond with another silanol group. Regarding the hydrogen bonds between a water molecule and the surface, 94% of the silanol groups donate a hydrogen bond to a water molecule and each silanol group receives on average 0.96 hydrogen bonds from water. Since a silanol group can donate only 1 hydrogen bond and can receive 1 or 2 hydrogen bonds, the bonding possibilities of the surface are extensively used. Within the nanotube, on average 58% of water molecules receive or donate at least one hydrogen bond from or to a silanol group, *i.e.* are in the first adsorbed layer.

Table 5.6 – Average number of hydrogen bonds (± the standard deviation) for given donor–acceptor pairs along a 1 ns trajectory of empty and hydrated nanotubes. Total number of silanol groups: 72; total number of water molecules: 98.

Donor	Acceptor	Empty Nanotube	Hydrated Nanotube
Silanol	Silanol	35.4 ± 2.3	2.0 ± 1.3
Water	Silanol		69.0 ± 2.1
Silanol	Water		68.7 ± 1.7
Water	Water		112.4 ± 3.7

To further characterize the local hydrogen bond network, we considered all 4 singular hydrogen bonding patterns possible for a single water molecule: donating to a silanol or a water molecule, or accepting from either. Each of these singular behaviors can appear zero, once, or twice. The patterns that represent more than 5% of the occurrences found over the whole trajectory are described in table 5.7.

Table 5.7 – Occurrence of the most recurrent patterns of hydrogen bonding characterized by the number of bonds given to silanol and water and received from silanol and water. Percentages are given with respect to the total number of water molecules.

Pattern	Occurrence	Donating to silanol	Accepting from silanol	Donating to water	Accepting from water
1	22.4%	0	0	2	2
2	14.9%	1	2	1	0
3	9.3%	2	1	0	1
4	7.8%	0	0	2	1
5	7.3%	1	1	1	1
6	6.3%	1	1	1	0

The six patterns highlighted by this analysis can be divided into two groups: (i) water molecules that do not form any hydrogen bond with a silanol group (patterns 1 and 4), that can have either 4 hydrogen bonds (tetragonal arrangement) or 3 hydrogen bonds, similarly to bulk water; and (ii) molecules whose adsorption on the internal surface of the nanotube involves at least one hydrogen bond (types 2, 3, 5 and 6). It is worth noting that the mean number of hydrogen bonds for a water molecule not bonded to the surface is 3.65, slightly higher than the bulk value for f-SPC water (3.60 hydrogen bonds per water molecule). The most common patterns are represented in figure 5.20.

For adsorbed molecules, the most common pattern is an "upright" water molecule (pattern 2 in figure 5.20), in a plane bisecting the triangular adsorption site. This water molecule receives two hydrogen bonds from two silanol groups and donates one hydrogen bond to the third silanol. A water hydroxyl group is left pointing toward the center of the nanotube, free to donate a hydrogen bond to a water molecule of the second layer. This creates a structuration of the adsorbed water beyond the first layer, which is clearly visible in the density maps (figures 5.15 and 5.16), with a favorable position for water oxygen atoms across that pending hydroxyl group. It also defines an equilibrium distance

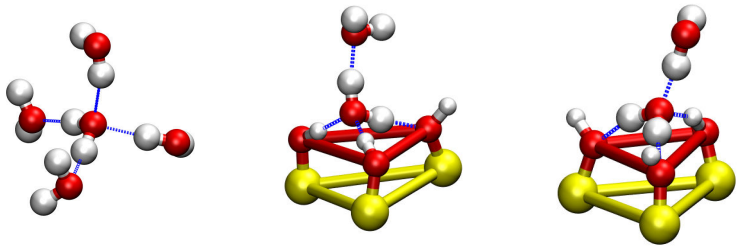


Figure 5.20 – Illustration of the most common hydrogen bonding patterns for water in an imogolite nanotube. From left to right these are patterns 1, 2 and 3. For the numbering of patterns, see table 5.7.

between the first and second layer of adsorption, which corresponds to water–water hydrogen bond length.

The second most common pattern for the first layer of water is a water molecule lying "flat" on the internal imogolite surface, donating two hydrogen bonds to neighboring silanol groups and receiving one from the third silanol neighbor. This allows the formation of a fourth hydrogen bond from a water molecule outside of the first layer. In this case, the direction of the hydrogen bond is less constrained, explaining the "diffuse" part of the oxygen density in the second water layer, seen in figure 5.16.

The large possibilities of organization of the first layer of adsorbed water molecules, with orientational disorder, is reminiscent of the Bernal–Fowler theory of ice rules[209]. From the relative energy of each individual configurations, as well as the frustration arising from constraints (no more than two hydrogen bonds can be donated or received, and no two neighboring adsorption sites can be populated at the same time), we believe that a model of the orientational disorder in that strongly bound layer of adsorbed water could be constructed, possibly explaining why the adsorption sites are only occupied at nearly 38% at saturation uptake.

5.2.5 Water dynamics

Diffusion

From the structural analysis of the water density and the hydrogen bonds in the adsorbed water, there are clear hints at the dynamic nature of the system and the lability of the water molecules. We have seen that hydrogen bond patterns are statistically distributed among several configurations, and that there is anisotropic surface diffusion in the first layer of adsorbed water (see figure 5.18 and accompanying text).

We extracted diffusion coefficients from mean square displacement (figure 5.21) of water molecules in the nanotube using Einstein diffusion relation between the mean square displacement (MSD) and the diffusion coefficient D:

$$\mathrm{MSD}(t) = 2\delta\ \mathrm{D}\ t. \tag{5.6}$$

In this relation, δ is the number of spatial dimensions of the movement. In our case, it will take the value of 1 for diffusion along the z axis; and 2 for diffusion in the xy plane.

We find a diffusion coefficient in the z axis of $3.5 \times 10^{-6}\,\mathrm{cm}^2/\mathrm{s}$, which is three times smaller than the coefficient calculated by Zang et al.[210] for a similar but lower water

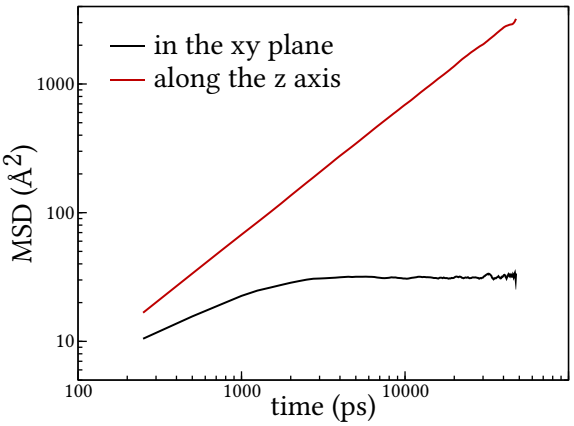


Figure 5.21 – Mean square displacement as function of time decomposed along the z axis and in the xy plane.

density. Moreover, we do not see a long-time diffusive behavior in the *xy* plane because of the confinement. This surface diffusion occurs without breaking all the hydrogen bonds involved. By breaking a single hydrogen bond, an adsorbed water molecule can flip around the axis formed by the remaining silanol groups, and thereby diffuses to the adjacent site, then reforms a hydrogen bond with another SiOH group. Since the triangular sites are isosceles and the surface is curved, the flip is easier and requires less configurational change within the furrows, than crossing the silicon rings. This explains the preferential circular diffusion as opposed to the axial diffusion in fully loaded nanotube.

Hydrogen bonds dynamics

In this section, we go further to quantify the dynamics of the water inside the imogolite nanotube using — as in the study of water intrusion in ZIF-8 — hydrogen bonds formation. We calculated autocorrelation functions of the presence of hydrogen bonds between water molecules, or water molecules and silanol groups, both within the first adsorption layer and in the less-ordered "core" of the tube. Due to the relatively slow dynamics of the strongly-confined water, these quantities were calculated from 50 ns classical trajectories. For all three possible donor–acceptor hydrogen bond types, we computed autocorrelation decay, represented in figure 5.22. From these autocorrelation functions, the lifetime of each hydrogen bond type was extracted as the time for the autocorrelation function to decay to 50%, and are presented in table 5.8. The autocorrelation decay is not strictly exponential, because the hydrogen bonds form multiple populations (first and second layer, etc.) with different decays.

Donor	Acceptor	Characteristic time τ
Water	Water	4.43 ps
Water	Silanol	57.43 ps
Silanol	Water	123.5 ps

Table 5.8 – Characteristic lifetime for hydrogen bonds calculated from the time length for the autocorrelation function to decay to 50%.

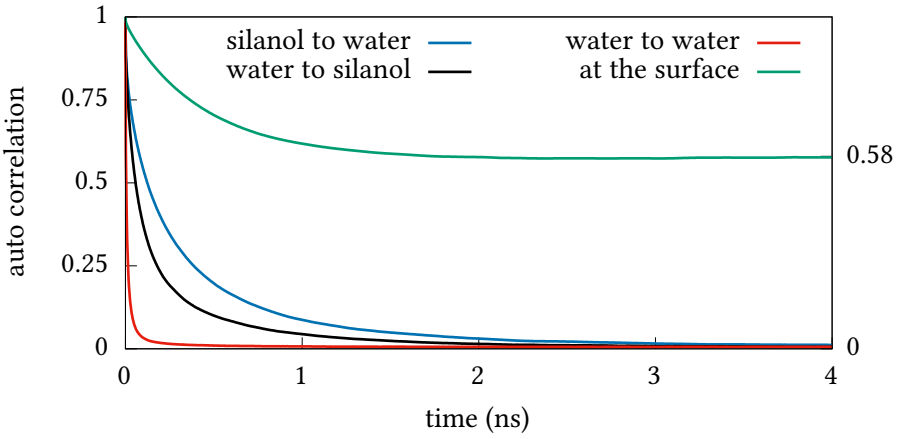


Figure 5.22 – Autocorrelation decay of the hydrogen bond existence for three types of donor–acceptor couples: water–silanol (black), silanol–water (blue) and water–water (red); as well as hydrogen bonds formed by molecules in the first adsorption layer (green);

First, looking at the water–water hydrogen bonds, we see that their lifetime is roughly 2 times higher in the imogolite nanotube than in bulk water (calculated at 2.67 ps from a simulation of pure f-SPC water). This slow down of the water dynamics under confinement has been well studied and explained in previous work, and is due in part to the stronger organization of the water, and in part to excluded volume effects: water molecules finding fewer partners to switch a hydrogen bond to[172, 211]. Secondly, we see that hydrogen bonds between water and the silanol-rich surface have a lifetime more than an order of magnitude greater than water–water hydrogen bonds. A study on hydrogen bonding of water on the α-quartz interface — whose (001) surface is chemically similar to imogolite, in that it presents geminal silanol groups — calculated persistence times of SiOH–water hydrogen bonds between 160 and 170 fs[212]. The extremely large difference, and the very strong bonding of water on the imogolite internal surface can be explained by the cooperative effect of the formation of three hydrogen bonds, on top of which is present the generic confinement slow-down.

We also investigated the residence time of water molecules in the first adsorbed layer, inside the imogolite nanotube. There are two ways to define the "first layer", by using geometrical radius, or by considering molecules that form at least one hydrogen bond with the surface. We chose the latter option, which is more chemically robust. The time autocorrelation function deriving from it is plotted in figure 5.22. It shows an exponential decay, and reaches a plateau at 58%, *i.e.* the average percentage of water molecules that belong to that first layer. The decay time for the autocorrelation function is found to be 410 ps. This is two or four times larger than the indicative lifetimes of hydrogen bonds with the surface (donated or received by the silanol group, respectively). This confirms the existence of an "internal surface diffusion" mechanism that allows diffusion between adsorption sites without exiting the layer itself. We can also compare this lifetime to residence times in protein adsorption sites. For example, for myoglobin[213], residence times span from 10 to 450 ps where the unusually high residence times (> 80 ps) correspond to water trapped either in the cavities inside the protein or in the grooves and concave regions.

Conclusions

This study of the properties of water adsorbed inside hydrophilic imogolite nanotubes was motivated by the need for a better understanding of the specific behavior of water due to its tight confinement, with numerous silanol groups on the internal surface of the nanotubes. Because of the strong interactions between water and imogolite, and the slow dynamics of the confined water, we relied on a classical force field in order to perform molecular dynamics studies at long time scales.

We then developed a model for the structure of adsorbed water, based on the calculated atomic densities and statistical analysis of the hydrogen bonding patterns observed between water molecules and silanol groups. We show that adsorption in the first layer vicinal to the imogolite tube can be described by an anisotropic triangular lattice of adsorption sites, which are not fully populated. We showed that interactions between the water molecules adsorbed at neighboring sites are complex, based on (i) exclusion of some nearest neighbor pairs (where the sites are too close to be occupied at the same time), and (ii) the number of hydrogen bonds donated (and accepted) by the water molecules to (and from) neighboring silanol groups. Maximizing the number of hydrogen bonds created, while adhering to these rules, creates frustration in the system and leads to the emergence of a heavily disordered state. This situation is very similar to the Bernal–Fowler rules describing the orientation of water molecules in ice[209]. Finally, we also characterized the dynamics of the confined water molecules, as well as the hydrogen bonds. We find that in addition to the generic effect of slow-down of confined water, the strong interactions of water molecules with the silanol groups giving rise to lifetimes that compare to typical residence times of water in protein adsorption sites.

6

IMPLEMENTATION OF MOLECULAR SIMULATION SOFTWARE

Part of my PhD project was to work on the implementation of molecular simulation software as part of the Domino software library. In this chapter, I will describe the context of this project and my contributions to it, as well as some advanced simulation techniques I worked to incorporate in Domino. The first of these advanced methods is the Hybrid Monte Carlo simulation technique, which joins molecular dynamics and Monte Carlo to exploit the advantages of both. I will present the original method and some of the most recent developments and improvements to it. I also explored methods for dealing with the treatment of electrostatic interactions in the context of periodic boundary conditions. I will first describe the Ewald summation technique, and some tricks for optimizing its implementation in the Domino code base. Then I will discuss the Wolf summation method as a possible alternative to Ewald summation that only relies on a sum of pairwise terms.

6.1 Domino: extensible molecular simulations library

Domino is a C++ library designed for classical atomistic molecular simulations. In this section, I will present its design, the code base and my contributions to it, as well as some of the project goals and software architecture choices made to reach these goals.

Domino is written in C++, a computer programming language created in 1983 by Bjarne Stroustrup as an extension of the C language. C++ mainly adds object-oriented programming, functions overloading and generic programming to C. We chose C++ because it is a general-purpose language, which we can use to implement complex algorithms while retaining the ability to manage memory allocations, locality, and layout. These two last points are what allows C++ to run faster than *managed* languages, where the users don't have explicit control over memory management. Using C++ does not necessarily makes software faster, but it enables developers to optimize the code more than other languages. Domino uses the C++ 11 version of C++ standard, which brought major changes to the language, making it easier to use correctly and more expressive.

In this section, I will assume some basic programming, C++ and object-oriented knowledge from the readers. There are a lot of great online resources on all of these subjects, as well as books such as *C++ Primer*[214], or *Effective Modern C++*[215]. Concerning the algorithms I implemented, I found the book by Frenkel and Smit[92] to be a great help.

6.1.1 Goals and architecture

The main goal of Domino is to be an extensible classical molecular simulation library. A software library is a collection of functions and classes working together to provide a given functionality. One specificity of Domino is that it provides facilities to run either classical Monte Carlo, molecular dynamics or energy minimization simulations, where most other simulation software only offer either Monte Carlo or molecular dynamics as *first-class citizen*. Notable exceptions that provide both Monte Carlo and molecular dynamics include RASPA[216] and Sire[217] software. We also try to make Domino easy to use, and easy to extend, meaning it can be used as a basis for the development of new molecular simulations methods. These goals translate into some of the architectural choices: in particular, the code needs to be simple to understand in order to be simple to extend. This means I restricted myself to a simple subset of C++, not using complex template-based meta-programming or deep class hierarchies.

The behavior and capabilities of Domino can be extended using two different mechanisms. Classes that provide central behavior, such as potentials or molecular dynamics integrators are manipulated through pointers to the corresponding pure virtual base class — or *interface* — throughout the code. By creating a new class inheriting from this interface, it is possible to add behavior to Domino without having to modify any of its internal source code. For example, the source code listing 1 shows the interface used for potentials and the implementation for Lennard-Jones potentials. Adding a new potential to Domino is as simple as creating a new class and implementing the corresponding required functions: `energy` and `force`. Other extensible parts of Domino include Monte Carlo moves, molecular dynamics thermostat and integrators, and simulation propagators.

```cpp
/// Abstract base class for all energy and forces computations. All
/// functions take an abstract parameter 'x' that will be the distance for
/// pair potentials and the angle for angles or dihedral angles potentials.
class Potential {
public:
    Potential() = default;
    virtual ~Potential() = default;

    /// Get the energy for the parameter `x`
    virtual double energy(double x) const = 0;
    /// Get the force factor for the parameter `x`
    virtual double force(double x) const = 0;
};

class LennardJones final: public Potential {
public:
    LennardJones(double epsilon, double sigma): sigma_(sigma), epsilon_(epsilon) {}

    double energy(double r) const override {
        auto sr = sigma_ / r;
        auto sr3 = sr * sr * sr;
        auto sr6 = sr3 * sr3;
        return 4 * epsilon_ * sr6 * (sr6 - 1);
    }

    double force(double r) const override {
        auto sr = sigma_ / r;
        auto sr3 = sr * sr * sr;
        auto sr6 = sr3 * sr3;
        return -24 * epsilon_ * sr6 * (1 - 2 * sr6) / r;
    }

private:
    double sigma_ = 0;
    double epsilon_ = 0;
};
```

Listing 1 – Extract of the definition of the `Potential` *interface in Domino, and implementation for Lennard-Jones potential.*

It is also possible to extend Domino by directly adding code before, inside or after the main simulation loop. As is it a library, Domino does not impose a structure on the simulation program. It is thus possible to customize the simulation flow, adding *on-the-fly* analysis, or merging multiple molecular dynamics simulation into a parallel tempering one. See the source code listing 2 for a simple example of constant pressure Monte Carlo simulation, showing how a simple simulation can be created by calling the high-level functions of the Domino library.

```cpp
#include "domino.hpp"
using namespace domino;

int main(int argc, char *argv[]) {
    domino::initialization();

    // Read the system
    auto system = Trajectory("initial.pdb").read();
    // Read the interaction potential
    domino::InputFile("potential.yml").read_to(system);

    // Setup the simulation: Monte Carlo at 300 K with three moves
    auto mc = MonteCarlo(units::from(300, "K"));
    mc.add_move(Translate(units::from(0.5, "A"), 50));
    mc.add_move(Rotate(units::from(20, "deg"), 50));
    mc.add_move(Resize(units::from(500, "bar"), 1));
    mc.setup(system);

    // Add code before the simulation loop
    // ...

    // Simulation loop
    for (size_t i=0; i<1e6; i++) {
        // This function call will propagate the simulation for one step
        mc.propagate(system);

        // Add code inside the simulation loop
        if (i % 100 == 0) {
            std::cout << i << " " << system.energy() << "\n";
        }
    }

    // Add code after the loop
    std::cout << mc.summary() << std::endl;

    domino::finalization();
    return 0;
}
```

Listing 2 – Example of a constant pressure Monte Carlo simulation using Domino.

Contributions to Domino

The Domino project was started in 2009 by François-Xavier Coudert and Anne Boutin. Before I started working on it in 2015, it had contributions by Jean-Marie Teuler, Julien Germond and David Bousquet. Since then, I have been the sole contributor to the code, and rewrote most of it, creating 480 git commits (self-contained modifications) out of the 800 of the repository. My contributions were related to two principal areas of the code: code quality improvement and simulation algorithms implementation.

On the code quality improvements side, I ported Domino to C++ 11, and cleaned up multiple previous experiments, such as a Python binding, atomic selection language and file input and output (for which I developed the standalone chemfiles library, presented

in appendix A). I added tests with respect to NIST simulation reference data[218], as well as an input file system to specify interaction potentials. I improved the simulated system in-memory representation and canonicalization: dealing with normalization of bond and angles indices and representation of the bonded distance matrix.

On the side of algorithms implementation, I added the interface-based extensibility of Domino. I also implemented grand canonical and hybrid Monte Carlo moves, as well as an anisotropic Berendsen barostat, a Monte Carlo barostat and the CSVR thermostat for molecular dynamics. I improved the support for anisotropic simulations by implementing anisotropic stress and pressure calculations for both Monte Carlo and molecular dynamics. I implemented both Ewald and Wolf summation methods for electrostatic energy and forces calculations. Finally, I improved the caching of energy in Monte Carlo simulations, adding support for more types of move and electrostatic interactions.

6.1.2 Challenges encountered

I encountered a few challenges while working on Domino, some of which are generic to the field of molecular simulation, others which were specific to the fact that Domino provides both Monte Carlo and molecular dynamics. I will present here some of these challenges, as well as the technical solutions I developed to overcome them. Several of these challenges were related to the efficiency of computing the energy and forces under periodic boundary conditions, which are the most expensive steps of simulations.

A performance model of modern computers

At their basis, computers are made of two main components: the central processing unit (CPU) or processor executes the instructions, and the memory — usually in the form of random access memory (RAM) — stores data used by the processor. On the first personal computers in the 80s, CPU executed instructions at 1 MHz, and RAM accesses were made at roughly the same speed. Optimizing code then mainly meant reducing the number of instructions executed by the CPU. Since then, CPU speed increased dramatically to around 4 GHz, while the RAM access speed only increased to 10 MHz. This means that data access patterns are much more important to consider when optimizing software for low run time. In order to speed up data access, modern CPUs come with multiple levels of cache, from the fast but small first level (L1) cache, to the bigger but slower L2 and L3 caches. Upon the first access, data is copied from RAM to the cache, and on subsequent read or write operations, the data from the cache is used. Later on, data is copied back to RAM. Typically latency for data access on caches and RAM are represented on figure 6.1.

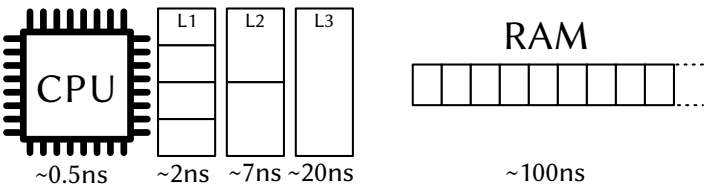


Figure 6.1 – Latency for data access from the CPU to three level of CPU cache (L1, L2 and L3) and the RAM of a modern computer. The numbers are for Intel Corporation's i7 and Xeon CPUs[219].

Because the CPU can spend hundred of cycles waiting for data which is not yet in the caches, it is important to make sure data in the RAM is accessed in a predictible way. When some data is explicit fetched from the RAM, another mechanism called *pre-fetching* kicks in, fetching additional data close to the one requested and storing it in the cache. If the code accesses data in a linear way, the next data fetch will be faster, as everything required is already in the caches. Software behaving this way is said to have high *memory locality*. Improving memory locality is nowadays an additional important part of software optimization.

Mixed molecular dynamics and Monte Carlo

Monte Carlo, and in particular Grand Canonical Monte Carlo, requires a different set of operations on the system that molecular dynamics, and has a different optimal data representation. Specifically, Grand Canonical Monte Carlo requires the possibility to dynamically change the number of molecules, and prevents the use of atomic index as atomic identifiers, as the indices will change during the simulation. At the same time, molecular dynamics can benefit from distributed memory parallelism (using for example the *Message Passing Interface* MPI) by using a domain decomposition algorithm, but this kind of parallelism is less trivial to integrate efficiently with Monte Carlo. To both allow GCMC simulations and future implementation of distributed memory parallelism for molecular dynamics, the storage of atoms (and the associated properties) uses dynamic allocation in growable containers. Furthermore, atoms from the same molecules are kept contiguous in memory, improving memory locality when computing intra-molecular interactions and allowing the implementation of domain decomposition.

Moreover, molecular dynamics needs information on the forces acting on the system, while Monte Carlo only needs to compute the energy. When computing the forces, adding the energy is a minimal additional runtime cost — periodic boundary conditions have already been applied — and no additional memory cost. On the opposite, when computing energy, adding forces increases both the memory and runtime cost: for a small system containing 50 water molecules, computing only the energy with Domino takes 1.3 ms, and computing the energy and the forces takes 2.6 ms. Domino thus provides two different code paths, one computing only the energy for Monte Carlo and another one computing both the energy and the forces for molecular dynamics.

Another possible improvement to the speed of simulations comes from the fact that most Monte Carlo moves only update the positions of a few atoms. This means that most of the energy components do not change before and after the move. Instead of recomputing all the energy components when evaluating the energy difference ΔU in the Metropolis criterion, Domino uses a cache for the energies, storing the energy associated with a pair in a bi-dimensional array, and only updating the energies associated with the modified pairs. This strategy is illustrated in figure 6.2. This can result in multiple orders of magnitude of improvement when compared to a full energy evaluation.

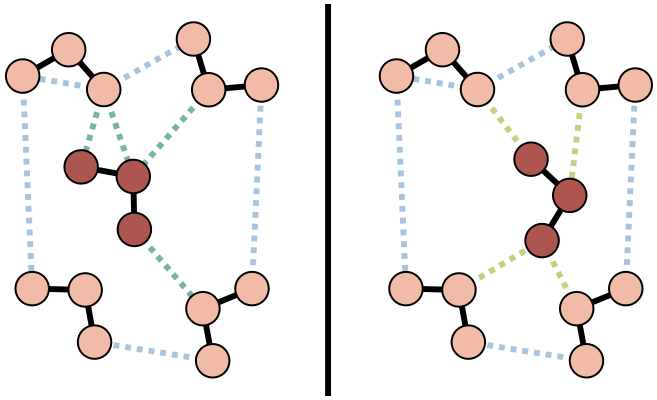


Figure 6.2 – Illustration of the use of an energy cache with Monte Carlo simulation. Before the move, all the interaction contributions are computed and stored. After the move, only the interactions in green needs to be computed to get the energy difference. Not all pair contributions are represented on the figure.

Computation of the stress tensor and pressure

Another difference between Monte Carlo and molecular dynamics concerns the computation of virial pressure with rigid or flexible molecules. It is often interesting to use a rigid model for small molecules in classical molecular simulation. If we are not interested in the small vibrational movements around equilibrium of the bonds and angles, making the approximation that molecules are rigid restricts the size of the phase space we have to sample. Representing the molecules as rigid bodies also allow us to use larger timesteps in molecular dynamics simulations. Molecular dynamics simulations need to use specific algorithms to simulate rigid molecules, either by using rigid body dynamics, or by adding corrections to the positions, velocities, and forces. The latter case is the realm of the SHAKE and RATTLE family of algorithms[220, 221]. Monte Carlo simulations have the ability to run explicitly model rigid molecules by simply omitting moves that would deform molecules: we can restrict translations to entire molecules and include whole molecules rotation moves.

This difference comes into play when trying to evaluate the instantaneous pressure of the system. In the kinetic theory of gases, the pressure arise from shocks of molecules with the containing walls. In molecular simulations, there are usually no walls, and we want to compute an instantaneous pressure instead of a time average. We use instead the thermodynamic definition of the pressure:

$$\mathrm{P} = -\left.\frac{\partial \mathrm{F}}{\partial \mathrm{V}}\right|_{\mathrm{T,N}} \tag{6.1}$$

For a given potential energy surface $\mathrm{U}(\boldsymbol{r})$, this equation gives the definition for the *virial pressure*:

$$\mathrm{P} = \frac{n_f k_\mathrm{B} \mathrm{T}}{3\mathrm{V}} + \frac{1}{3\mathrm{V}}\left[\sum_i \boldsymbol{r}_i \cdot \boldsymbol{f}_i - 3\mathrm{V}\frac{\partial \mathrm{U}}{\partial \mathrm{V}}\right], \tag{6.2}$$

where V is the simulation volume, T the instantaneous simulation temperature, n_f the number of degrees of freedom, $\boldsymbol{r}_i$ is the position of atom i and $\boldsymbol{f}_i$ the force acting on it. Unfortunately, the sum is not trivial to compute in presence of periodic boundary

conditions, as there are multiple equivalent values for $\boldsymbol{r}_i$. Instead, we typically use in molecular simulation software origin-independent formulations. For fully flexible molecules and potentials without explicit dependency on the volume, it is possible to rewrite equation (6.2) so that is only depends on pairs distances and forces. This gives us the more standard formulation of the virial pressure:

$$\mathrm{P} = \frac{n_f k_\mathrm{B} \mathrm{T}}{3\mathrm{V}} + \frac{1}{3\mathrm{V}} \sum_i \sum_{j>i} \boldsymbol{r}_{ij} \cdot \boldsymbol{f}_{ij}. \tag{6.3}$$

Only the forces arising from non-bonded or bonded pair potentials appear in the double sum, as the sum cancels out for three or four body potentials that only depend on the geometrical angle or dihedral angle[222]. The stress tensor — for anisotropic simulation — is computed in a similar way:

$$\underline{\sigma} = \frac{n_f k_\mathrm{B} \mathrm{T}}{3\mathrm{V}} \mathbb{1} + \frac{1}{3\mathrm{V}} \sum_i \sum_{j>i} \boldsymbol{r}_{ij} \otimes \boldsymbol{f}_{ij}; \tag{6.4}$$

where $\otimes$ denotes the tensorial product: $(\boldsymbol{r} \otimes \boldsymbol{f})_{\alpha\beta} = r_\alpha \times f_\beta$. The pressure is simply the trace of the stress tensor: $\mathrm{P} = \mathrm{Tr}(\underline{\sigma})$.

For rigid molecules, the above expression is impossible to use, as we don't know the forces acting between atoms inside the same molecule. Instead, we can use the fact that molecules are rigid to reformulate equation (6.2) into a different expression:

$$\underline{\sigma} = \frac{n_f k_\mathrm{B} \mathrm{T}}{3\mathrm{V}} \mathbb{1} + \frac{1}{3\mathrm{V}} \sum_a \sum_{b>a} \sum_{i \in a} \sum_{j \in b} \boldsymbol{r}_{ij} \otimes \boldsymbol{f}_{ij} \frac{\boldsymbol{r}_{ab} \cdot \boldsymbol{r}_{ij}}{\|r_{ij}\|^2}. \tag{6.5}$$

Here, the sums over a and b run over different molecules, and the sums over i and j run over the atoms in the molecules. In both cases, if the potential has an explicit dependency on the volume, the corresponding terms needs to be included. This is in particular the case for the electrostatic potential when computed with Ewald summation.

Domino uses either equation (6.4) or (6.5) depending on the propagator used (molecular dynamics or Monte Carlo), as well as the set of moves for Monte Carlo simulations. Each move registers how it changes the system when generating a trial configuration, and Domino figures out the number of simulated degrees of freedom.

6.2 Hybrid Monte Carlo

Hybrid Monte Carlo, also called *Hamiltonian Monte Carlo* (HMC) is an improvement on standard Metropolis Monte Carlo that allows simulating complex systems more efficiently. Considering a Monte Carlo simulation in the NVT ensemble, the usual way to generate new conformation in the Markov chain is to randomly pick and translate (and additionally rotate for rigid molecules) a particle in the system. The amplitude of the translation is limited by the acceptance rate of the move: while higher amplitudes allow sampling the phase space more efficiently, they are correlated with a much lower acceptance rate; and thus increase the amount of work done by the simulation to generate a new conformation.

This is even more relevant when the system contains large flexible molecules, spanning long distances, such as proteins, polymers or here, metal–organic frameworks. The typical movements and conformation changes of molecules are by nature collective, multiple atoms moving together in the same direction. The standard Monte Carlo moves have difficulties to sample this kind of collective behavior, as they would require multiple single atom moves in the same direction. Multiple improved Monte Carlo moves have been proposed to overcome these limitations. For small molecules displaying intra-molecular flexibility, it is possible to directly sample rotations around all the bonds using configuration bias Monte Carlo based on Rosenbluth sampling (see reference [92] for more details on the algorithm), where the molecule is fully regrown from a starting atom, randomly choosing the orientation of the next backbone bond at each step. This procedure is illustrated in figure 6.3.

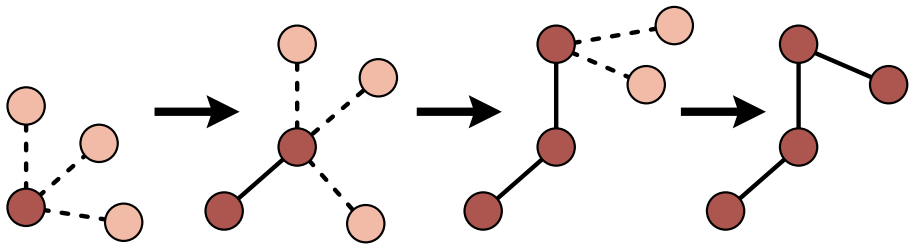


Figure 6.3 – Illustration of a single configurational bias Monte Carlo move. A linear molecule is regrown step by step. At each step, the position for the next atom is picked at random, depending on the position of all the previous atoms.

The main issue with configuration bias Monte Carlo is that it requires knowledge of all the separated intra-molecular interaction terms to be able to generate the position of the next atom while maintaining detailed balance; and not only the global energy change as for simpler moves. It is also harder to use even with the simplest branched molecules.

Flexible nanoporous materials add another level of difficulty for efficient Monte Carlo simulations. Some of them display collective behavior linked to global deformations of the simulation cell, such as breathing or gate-opening. Here, the volume of the simulation cell changes as the linkers rotate and move, thus needing Monte Carlo moves able to sample both the collective rotation of linkers and the changes in the unit cell shape and volume. It is possible to use molecular dynamics instead of Monte Carlo to explore the

response of the material to external stress, as molecular dynamics simulations are very efficient when sampling collective behaviors. But when the stress is created by adsorbed molecules in a variable number (for simulations with fixed chemical potential), we need to use grand canonical Monte Carlo (molecular dynamics is not suited for simulations in the grand canonical ensemble, see section 3.4.2). Hybrid Monte Carlo is a technique that can bridge the gap between traditional Monte Carlo and molecular dynamics, bringing the ability to sample such collective motions to a Monte Carlo simulation[223].

In addition to being able to improve the sampling efficiency of Monte Carlo simulations, Hybrid Monte Carlo can bring the power of non-physical moves in simulations usually relying on molecular dynamics. For example, the study of large bio-molecules — the typical example is the simulation of protein folding — is often limited by the time scale at which the simulation can produce new conformations, decorrelated from the previous ones[224]. Monte Carlo can help reduce this time by allowing jumps from one conformation to another, and incorporate domain-specific knowledge (which part of the protein can rotate, which parts will move together) to improve simulation efficiency.

Another area that can see improvements by incorporating non-physical moves is the simulation of diluted aqueous environments, such as the salt and pH environment around proteins. The pH of human blood is constant around 7.4, meaning that both HO^- and H^+ ions are only present at a concentration around 10^{-7} mol/L. If we want to simulate a realistic pH environment, we need to simulate more than 500 million water molecule for each HO^- or H^+ ion. In addition to that, because the surface of proteins can carry non-neutral charges, the local ion concentration can depart from the measured, global concentration in blood plasma or cell's cytoplasm. Grand Canonical and semi-Grand Canonical Monte Carlo moves used together with Hybrid Monte Carlo can bring realistic salt and pH condition to these simulations[225].

6.2.1 Mixing molecular dynamics and Monte Carlo

Hybrid Monte Carlo was first devised in 1987 by Duane et al. for calculations in lattice quantum chromodynamics [226]. It was then adapted to condensed matter molecular simulation by Mehlig et al. in 1992[227]. The central idea is to use a short molecular dynamics simulation (around 10 steps) to generate a new conformation for the Markov chain. Once the molecular dynamics simulation finished, the final step is considered as a trial conformation, and accepted or rejected with the adapted Metropolis criterion.

One global move in *configuration space* consists in propagating the system through *phase space* for a fixed number of steps using some integration scheme $\psi_{\delta t}$ of Hamilton's equations. $\psi_{\delta t}$ depends on the integration time step δt and the Hamiltonian of the system $\mathcal{H}$; and maps the initial configuration $(\boldsymbol{r}, \boldsymbol{v})$ in phase space to the final one $(\boldsymbol{r}', \boldsymbol{v}')$:

$$\begin{array}{rccc} \psi_{\delta t}: & \mathbb{R}^{6N} & \longrightarrow & \mathbb{R}^{6N} \\ & (\boldsymbol{r}, \boldsymbol{v}) & \longmapsto & \psi_{\delta t}(\boldsymbol{r}, \boldsymbol{v}) \equiv (\boldsymbol{r}', \boldsymbol{v}') \end{array} \tag{6.6}$$

Since the usual Monte Carlo scheme does not use (or propagate) the atomic velocities, we need to generate new velocities before starting the molecular dynamics simulation. We

choose to generate them randomly according to the the canonical ensemble distribution at temperature T:

$$\mathcal{P}_{\mathrm{T}}(\boldsymbol{v}) \propto \exp\left(-\beta \sum_i \frac{1}{2} m_i \boldsymbol{v}_i^2\right) \tag{6.7}$$

Following the same notations as in chapter 3, because the time integration is deterministic, the probability $\alpha(\boldsymbol{r} \to \boldsymbol{r}')$ to generate a given conformation $\boldsymbol{r}'$ starting from $\boldsymbol{r}$ is the same as the probability to generate a specific set of initial velocities $\mathcal{P}_{\mathrm{T}}(\boldsymbol{v})$:

$$\alpha(\boldsymbol{r} \to \boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}' = \mathcal{P}_{\mathrm{T}}(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} \tag{6.8}$$

The error in energy conservation made by the propagator $\psi_{\delta t}$ is called the discretization error $\delta\mathcal{H}$. This value is associated with the numeric integration scheme used by the molecular simulation, and depends only on δt and the number of steps used to propagate the system with molecular dynamics.

$$\delta\mathcal{H} = \mathcal{H}(\boldsymbol{r}', \boldsymbol{v}') - \mathcal{H}(\boldsymbol{r}, \boldsymbol{v}) \tag{6.9}$$

If we use an acceptation probability that depends on the discretization error $\delta\mathcal{H}$:

$$\mathrm{acc}\left[(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}', \boldsymbol{v}')\right] = \min\left(1, e^{-\beta\delta\mathcal{H}}\right), \tag{6.10}$$

we can show that the resulting Monte Carlo move respects the detailed balance, provided that the integration scheme $\psi_{\delta t}$ is both *time reversible* and *symplectic*.

$$\begin{aligned}
\mathcal{P}(\boldsymbol{r})\, \pi(\boldsymbol{r} \to \boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{v} &= \mathcal{P}(\boldsymbol{r})\ \mathcal{P}_{\mathrm{T}}(\boldsymbol{v})\ \mathrm{acc}\big((\boldsymbol{r}, \boldsymbol{v}) \to \psi_{\delta t}(\boldsymbol{r}, \boldsymbol{v})\big)\ \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{v} \\
\textit{see below} &= \mathcal{P}(\boldsymbol{r}')\ \mathcal{P}_{\mathrm{T}}(\boldsymbol{v}')\ \mathrm{acc}\big(\psi_{\delta t}(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}, \boldsymbol{v})\big)\ \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{v} \\
\textit{time reversible} &= \mathcal{P}(\boldsymbol{r}')\ \mathcal{P}_{\mathrm{T}}(\boldsymbol{v}')\ \mathrm{acc}\big((\boldsymbol{r}', \boldsymbol{v}') \to \psi_{-\delta t}(\boldsymbol{r}', \boldsymbol{v}')\big)\ \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{v} \\
\textit{symplectic} &= \mathcal{P}(\boldsymbol{r}')\ \mathcal{P}_{\mathrm{T}}(\boldsymbol{v}')\ \mathrm{acc}\big((\boldsymbol{r}', \boldsymbol{v}') \to \psi_{-\delta t}(\boldsymbol{r}', \boldsymbol{v}')\big)\ \mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{v}' \\
&= \mathcal{P}(\boldsymbol{r}')\ \pi(\boldsymbol{r}' \to \boldsymbol{r})\, \mathrm{d}\boldsymbol{r}'\mathrm{d}\boldsymbol{v}'
\end{aligned} \tag{6.11}$$

The first step in the above demonstration comes from the mathematical identity

$$e^{-\beta\mathcal{H}(\boldsymbol{r},\boldsymbol{v})} \min\left(1, e^{-\beta\delta\mathcal{H}}\right) = e^{-\beta\mathcal{H}(\boldsymbol{r}',\boldsymbol{v}')} \min\left(e^{\beta\delta\mathcal{H}}, 1\right). \tag{6.12}$$

It is worth noting that even though the molecular dynamics simulation evolves in the microcanonical NVE ensemble, the overall HMC simulation is sampling the canonical NVT ensemble. This allows using HMC simulations as a rigorous way to sample the NVT ensemble using molecular dynamics.

Constant pressure simulations

There are two ways to use Hybrid Monte Carlo simulations to sample the isobaric-isothermal NPT ensemble. The first one is to use separate moves to change the volume and the particles' positions — the former using standard Monte Carlo moves and the latter relying on hybrid moves. This is the usual way of running NPT simulations with Monte Carlo, with some added hybrid moves to sample the configuration space at different fixed volumes. While doing so is simpler from an implementation point of view, it can be less efficient when the changes in volume are coupled to local deformations of the system, which is often the case in flexible nanoporous materials.

Another possibility is to use the hybrid moves to sample both the volume changes and the particles' displacements. The same construction described above can be used to create an hybrid move that sample the NPT ensemble. The molecular dynamics integrator now maps the initial position in phase space including the volume V $(\mathrm{V}, \boldsymbol{r}, \boldsymbol{v})$ to a new position in phase space $(\mathrm{V}', \boldsymbol{r}', \boldsymbol{v}')$. If the integrator is deterministic, the probability to generate a specific new state from the initial one, $\alpha\left[(\mathrm{V}, \boldsymbol{r}) \to (\mathrm{V}', \boldsymbol{r}')\right]$, is still the same as the probability to generate the initial velocities $\mathcal{P}_{\mathrm{T}}(\boldsymbol{v})$. Provided the integrator is time reversible and symplectic, the same demonstration as in equation (6.11) applies by adapting the acceptance criterion to

$$\operatorname{acc}\left[(\mathrm{V}, \boldsymbol{r}, \boldsymbol{v}) \to (\mathrm{V}', \boldsymbol{r}', \boldsymbol{v}')\right] = \min\left(1, e^{-\beta\delta\mathcal{H} - \beta\ \mathrm{P}\Delta\mathrm{V}}\right) \tag{6.13}$$

If an extended Lagrangian integrator such as the one proposed by Andersen (see reference [228]) is used, the momentum of the fictitious external piston should also be included in both the probability of creating a new state $\alpha((\mathrm{V}, \boldsymbol{r}) \to (\mathrm{V}', \boldsymbol{r}'))$ and in the Hamiltonian part of the acceptance criterion[229, 230].

Again, we should note that the molecular dynamics simulation only samples the isenthalpic-isobaric NPH ensemble, and the Monte Carlo acceptance criterion ensures sampling of the isobaric-isothermal NPT ensemble. Actually, the molecular dynamics does not even need to sample an accurate NPH ensemble, only to generate new states with different volume and following the above properties of being deterministic, time-reversible and symplectic. The Metropolis acceptance criterion ensures that the correct ensemble will be sampled. However, care must still be taken to minimize the enthalpy drift $\delta\mathcal{H} + \mathrm{P}\Delta\mathrm{V}$ during the simulation to maintain a high acceptance rate for hybrid moves.

Osmotic simulations

Once we are able to use hybrid Monte Carlo simulations to sample the NPT ensemble, moving to the osmotic $\mathrm{N}_{\mathrm{host}}\,\mu\,\mathrm{PT}$ ensemble is accomplished by adding insertion/deletion moves to allow the number of adsorbed particles to vary. Earlier this year, Rogge et al.[223] used such hybrid Monte Carlo simulations to study the adsorption of noble gases, CO_2 and CH_4 in MIL-53(Al) using the osmotic ensemble. They were able to reach very good agreement with experimental measurements of adsorption isotherms and predict the bi-stability and breathing of MIL-53(Al) upon adsorption.

Because the correctness of the sampled ensemble is validated by the Metropolis criterion and the detailed balance, the short molecular dynamics simulations used in hybrid moves are not required to produce the statistically correct ensemble. This means that it should be possible to use an adapted Grand Canonical Molecular Dynamics scheme to generate new trial conformation to be accepted with the Metropolis criterion. To my knowledge, this as not been attempted yet, but could improve insertion rate for osmotic simulations in dense phases — such as the simulation of intrusion in porous solids. Traditional insertion/deletion moves suffer from a very low acceptance rate in dense phases such as liquids, because molecules are already densely packed and there is not enough space to add new molecules. Grand canonical molecular dynamics can help by progressively scaling the *presence* of the new molecule up, allowing its surroundings to relax.

Even if we have a simulation method able to sample the osmotic ensemble, we will still need force fields able to accurately describe both the flexibility of the materials and the interactions between the material and the fluids inside. Developing such force field will require additional efforts, especially considering the variability of structures of hybrid metal–organic materials, and the inherent complexity of describing coordination bonds. Force field parametrization methods based on machine learning — such as the one I presented in section 4.3 — could be very valuable for this endeavor.

Related algorithms and methods

In the last decade, some improvements to the simple hybrid Monte Carlo method presented above have been proposed. Shadow Hybrid Monte Carlo[224] relies on the existence and computability of a *shadow* Hamiltonian being propagated exactly (without any propagation error) by $\psi_{\delta t}$ to improve the acceptance rate of hybrid moves, and reconstruct *a posteriori* the probabilities of each generated conformation. Generalized Hybrid Monte Carlo[231–233] uses configurations in the phase space instead of conformations in the Markov chain, keeping the velocities from one Monte Carlo move to another. The different steps are still accepted or rejected by a Metropolis criterion, and velocities are partially updated with new random values at each step. The combination of these two approaches is called Generalized Shadow Hybrid Monte Carlo, and implemented in the GROMACS simulation software[230]. Finally, tentatives to lift the symplectic requirement on the integrator are at the origin of Compressible Generalized Hybrid Monte Carlo[234, 235]. The core idea is to include the Jacobian $\mathcal{J}_{\psi_{\delta t}}$ of the propagator in the probability ratio of the acceptance test:

$$\text{acc}\left[(\boldsymbol{r}, \boldsymbol{v}) \to (\boldsymbol{r}', \boldsymbol{v}')\right] = \min\left(1, \frac{\mathcal{P}(\boldsymbol{r}', \boldsymbol{v}')}{\mathcal{P}(\boldsymbol{r}, \boldsymbol{v})} \left|\mathcal{J}_{\psi_{\delta t}}\right|\right) \tag{6.14}$$

where $\mathcal{J}_{\psi_{\delta t}} = \det\left[\partial \psi_{\delta t} / \partial(\boldsymbol{r}, \boldsymbol{v})\right]$. For a symplectic integrator, $|\mathcal{J}_{\psi_{\delta t}}| = 1$, and we get back the standard hybrid Monte Carlo scheme.

6.2.2 Hybrid simulations in osmotic ensemble

I used my implementation of hybrid Monte Carlo in Domino in a preliminary study of the adsorption of methane CH_4 in MOF-5 at 300 K. I used a simplified model of both MOF-5 and CH_4 as Domino did not support computing electrostatic interactions when I first implemented these algorithms. CH_4 molecules were approximated by Lennard-Jones spheres, using $\sigma = 3.737$ Å and $\varepsilon = 1.247$ kJ/mol. I used the Lennard-Jones and intra-molecular terms from QuickFF[141] to describe the MOF, ignoring the atomic charges. While the resulting model is not an accurate representation of MOF-5, it is still an interesting test case for the use of hybrid Monte Carlo simulations in adsorption.

To obtain a full isotherm, I ran 8 simulations at different CH_4 pressures. Each simulation used two Monte Carlo moves: insertion/deletion of CH_4 taking the current pressure as the gas fugacity; and hybrid moves. The hybrid moves used a Velocity-Verlet integrator with a timestep of 1 fs, and a Berendsen barostat setting the external pressure at the same value as the CH_4 fugacity and a time step of 5 ps. I should point out that I used the Berendsen barostat even if it is not symplectic nor time reversible, as it was the only barostat implemented in Domino. Despite of its flaws, this barostat is very common in the literature as it is the simplest possible choice. Again, while the resulting simulation might not sample the adequate ensemble, it is still an interesting check. Each simulation

was propagated for a million of Monte Carlo moves, using the first 250 000 moves as the equilibration period. The resulting isotherm is shown in figure 6.4 together with the changes in unit cell volume as the pressure increases.

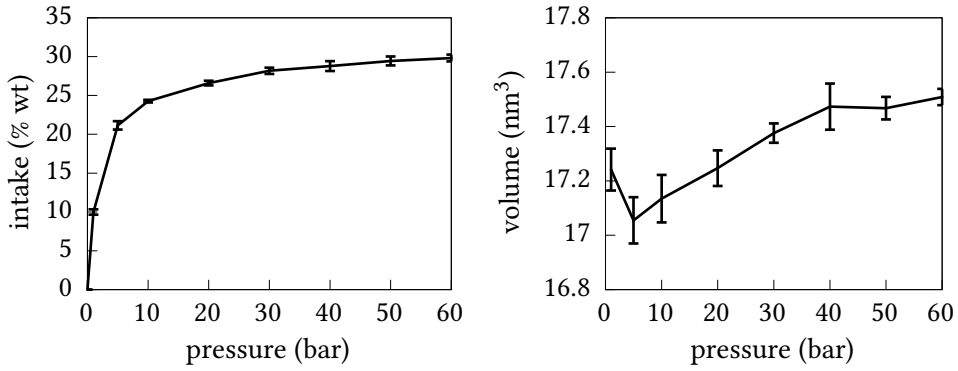


Figure 6.4 – Hybrid Grand Canonical Monte Carlo simulations results for the adsorption of methane in a simplified MOF-5 model. (left) adsorption isotherm at 300 K*, (right) volume change during adsorption.*

The resulting isotherm is a simple type I isotherm, as expected for the adsorption of methane in MOF-5. Contrary to standard GCMC simulations, this isotherm incorporates flexibility effects. More interesting is the non-monotonous behavior of the curve of volume deformations as a function of pressure. We first see a small contraction of the unit cell at low loading, before the expected increase at higher pressures. This contract–expand behavior is reminiscent of sorption-induced deformation in other porous materials[208, 236]: the presence of few molecules inside the pores induces a *softening* and a contraction of the whole system. Another way to look at this phenomenon is to envision the molecules inside the pore *pulling* on the pores' wall.

It is remarkable to see that such a simplistic model of the MOF is able to reproduce this relatively complex behavior. In order to improve the model and predictive possibilities of hybrid Monte Carlo simulations, I needed to implement a way to compute electrostatic interactions. In the next section, I will discuss two methods one can use to compute these interactions in the presence of periodic boundary conditions.

6.3 Computation of electrostatic energy in classical simulations

Classical force fields represent the energy of an atomic system as a sum of multiple contributions, including dispersion and short-range repulsion, intra-molecular interactions and electrostatic interactions. In the last case, atoms are often identified to point charges, interacting through a Coulombic potential. For two atoms i and j carrying charges q_i and q_j, at a distance r from one another, this potential reads:

$$\mathrm{V}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r}. \tag{6.15}$$

In the remaining of this chapter, I will be using units such that $4\pi\varepsilon_0 = 1$.

6.3.1 The problem

As presented in section 3.2.3, we are limited in the size of systems we can simulate with classical simulation methods. To remove the size and surface effects arising from the relatively small number of simulated particles, most if not all classical simulations use periodic boundary conditions. This means that we are, in effect, simulating an infinite system, extending in all three dimensions of space. All the inter-molecular interaction potentials decay to zero as the distance between atoms goes to infinity. It is thus customary to use a cutoff radius r_c when computing the energy of a system: any interaction between atoms further apart than this radius is supposed negligible, and set to zero. This allows to speed up the calculation of energies in the simulations by ignoring contributions between atoms that are too far apart. The error ε arising from the use of a cutoff radius can be quantified — using $\mathrm{V}(r)$ for the pair potential and $g(r)$ for the radial distribution function:

$$\varepsilon = \int_{r_c}^{\infty} r^2\, \mathrm{V}(r)\, g(r)\, \mathrm{d}r. \tag{6.16}$$

Assuming isotropic density, and for a cutoff radius big enough so that $g(r) \simeq 1$, we can compute this value — also called *long-range* or *tail* correction — and use it to correct after the fact the energy and pressure computed from a simulation.

Unfortunately, the above integral only converges if $\mathrm{V}(r)$ goes to zero faster than $1/r^3$, which is not the case for the electrostatic potential. In the following, I will be describing two methods one can use to compute the electrostatic interactions accurately in classical molecular simulations: Ewald summation and Wolf summation. I implemented both in the Domino project, and I will give some software implementation techniques I used to speed up energy and forces computations.

6.3.2 Ewald summation

Ewald summation was proposed by Paul Peter Ewald in 1921[237]. The core idea is to decompose the electrostatic potential using the identity:

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}. \tag{6.17}$$

By choosing the right $f(r)$, one can decompose $1/r$ potential into a short-range, rapidly converging potential that can be computed with a cutoff radius; and a long-range, smooth potential that will be computed in Fourier space. While I am focusing here on electrostatic

interaction, the Ewald summation is a more general technique that can be used with other potentials. The usual decomposition uses the error function erf and the complementary error function erfc:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \mathrm{d}t \qquad \operatorname{erfc}(x) = 1 - \operatorname{erf}(x) \tag{6.18}$$

The three functions are represented in figure 6.5, were we can see that $\operatorname{erfc}(r)/r$ goes to zero faster than $1/r$, and that $\operatorname{erf}(r)/r$ is very smooth, and can be prolonged to a finite value on $r = 0$.

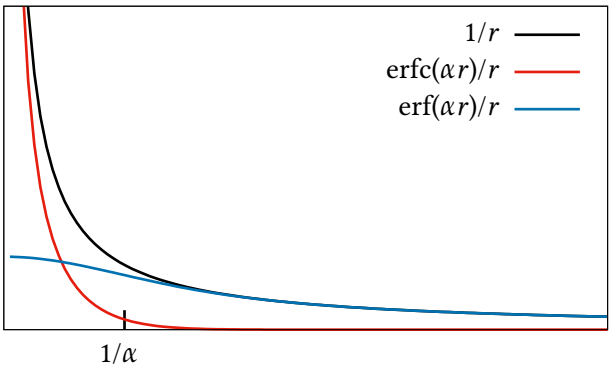


Figure 6.5 – Shape of the functions involved in Ewald summation: $1/r$, $\operatorname{erfc}(\alpha r)/r$, *and* $\operatorname{erf}(\alpha r)/r$.

In practice, a damping parameter α is used to tune the distance at which $\operatorname{erfc}(\alpha r)/r$ becomes dominant over $\operatorname{erf}(\alpha r)/r$. The potential to be computed is given by:

$$V(r) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r)}{r} + \frac{\operatorname{erf}(\alpha r)}{r} \right]. \tag{6.19}$$

The total energy of the system is a sum over all possible pairs, and can be decomposed as a short-range contribution U_{short} (containing the erfc terms) and a long-range contribution U_{long} (containing the erf terms):

$$U = \frac{1}{2} \sum_{i \neq j}^{\infty} q_i q_j \frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} + \frac{1}{2} \sum_{i \neq j}^{\infty} q_i q_j \frac{\operatorname{erf}(\alpha r_{ij})}{r_{ij}}; \tag{6.20}$$

$$U = U_{\text{short}} + U_{\text{long}} \tag{6.21}$$

The short-range contribution sum can be evaluated directly (in *real space*) using a cutoff radius r_c, as it decays faster than $1/r^3$:

$$U_{\text{short}} = \frac{1}{2} \sum_{\substack{i \neq j \\ r_{ij} < r_c}} q_i q_j \frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}}. \tag{6.22}$$

Concerning the long-range contribution, we can use the periodicity introduced by periodic boundary conditions to our advantage: as the potential $\operatorname{erf}(\alpha r)/r$ is relatively smooth, its Fourier transform will only contain a small number of harmonics. This means it will converge rapidly in *Fourier space*.

First we make the sum fully periodic by adding the term for $i = j$, and extending $\text{erf}(\alpha r)/r$ in zero to $\alpha/\sqrt{\pi}$:

$$\text{U}_{\text{long}} = \frac{1}{2}\sum_{i,j}^{\infty} q_i q_j \frac{\text{erf}(\alpha r_{ij})}{r_{ij}} - \sum_{i}^{\text{N}} \frac{\alpha}{\sqrt{\pi}} q_i^2. \tag{6.23}$$

By taking the Fourier transform of the sum over all reciprocal lattice translations $\boldsymbol{k}$, and after a bit of algebra, we get a simpler expression:

$$\text{U}_{\text{long}} = \frac{\pi}{\text{V}} \int_0^{\alpha} \frac{1}{u^3} \text{d}u \sum_{i,j} q_i q_j + \frac{2\pi}{\text{V}} \sum_{\boldsymbol{k}\neq 0}^{\infty} \frac{e^{-k^2/4\alpha^2}}{k^2} \left| \sum_{i}^{\text{N}} q_i \, e^{j\boldsymbol{k}\cdot\boldsymbol{r}_i} \right|^2 - \sum_{i}^{\text{N}} \frac{\alpha}{\sqrt{\pi}} q_i^2, \tag{6.24}$$

where j is the imaginary number ($j^2 = -1$). The first integral in equation (6.24) comes from the $\boldsymbol{k} = 0$ case, and is infinite. This means that the energy is infinite unless $\sum_{i,j} q_i q_j$ is null, *i.e.* unless the simulated system is neutral. This makes sense, as an infinite periodic charged system would have an infinite total charge. If this is the case, we need to add an uniform background charge to ensure the neutrality of the system.

The sum over $\boldsymbol{k}$ converges quickly, and is usually computed over a small number of $\boldsymbol{k}$ vectors, using a cutoff in Fourier space k_c:

$$\text{U}_{\text{long}} = \frac{2\pi}{\text{V}} \sum_{\substack{\boldsymbol{k}\neq 0 \\ \boldsymbol{k}^2 < k_c^2}} \frac{e^{-k^2/4\alpha^2}}{k^2} \left|\rho(\boldsymbol{k})\right|^2 - \sum_{i}^{\text{N}} \frac{\alpha}{\sqrt{\pi}} q_i^2. \tag{6.25}$$

$$\rho(\boldsymbol{k}) = \sum_{i}^{\text{N}} q_i \, e^{j\boldsymbol{k}\cdot\boldsymbol{r}_i} \tag{6.26}$$

The total energy in the Ewald formulation can include additional terms not derived here, such as a correction for charged systems, and a correction for intra-molecular interactions:

$$\begin{aligned}
&\text{short-range interaction} && \text{U} = \frac{1}{2} \sum_{i\neq j,\ r_{ij}<r_c}^{\ddagger} q_i q_j \frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} \\
&\text{long-range interaction} && + \frac{2\pi}{\text{V}} \sum_{\boldsymbol{k}\neq 0,\ \boldsymbol{k}^2<k_c^2} \frac{e^{-k^2/4\alpha^2}}{k^2} \left|\rho(\boldsymbol{k})\right|^2 \\
&\text{self interaction} && - \sum_{i}^{\text{N}} \frac{\alpha}{\sqrt{\pi}} q_i^2 \\
&\text{molecular correction} && - \frac{1}{2} \sum_{i\neq j}^{\ddagger -1} q_i q_j \frac{\text{erf}(\alpha r_{ij})}{r_{ij}} \\
&\text{charge correction} && - \frac{\pi}{2\alpha^2 \text{V}} \left| \sum_i q_i \right|^2
\end{aligned} \tag{6.27}$$

The molecular correction comes from another trick used when fitting the parameters of classical force fields. Electrostatic interactions between bonded atoms are ignored, allowing the use of less stiff bonded potential. They are typically completely ignored for 1-2 and 1-3 atoms, *i.e.* atoms directly bonded and atoms both bonded to a common central atom. 1-4 interaction can be included, scaled down or ignored depending on the force field. In the above expression, the $\sum^{\ddagger}$ symbol on the short-range term indicate that

intra-molecular contributions should be skipped or scaled down as necessary, and $\sum^{\ddagger -1}$ indicate that the sum run over interactions that should be excluded. For the long-range term, we only want to remove interaction between atoms in the same molecule and in the same simulation box image. Instead of changing the sum over the $\boldsymbol{k}$ vectors, we explicitly remove the corresponding contribution — leading to the molecular correction term in equation (6.27).

Following the same procedure, one can compute the force $\boldsymbol{f}_i$ acting on a given atom at position $\boldsymbol{r}_i$, and the stress tensor contribution from electrostatic interactions $\underline{\sigma}$.

$$\begin{aligned}
\boldsymbol{f}_i &= q_i \sum_{j=1,\ r_{ij}<r_c}^{\mathrm{N}\ \ddagger} q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 r_{ij}^2}\right] \frac{\boldsymbol{r}_{ij}}{r_{ij}^2} \\
&+ \frac{4\pi}{\mathrm{V}} \sum_{\boldsymbol{k}\neq 0,\ \boldsymbol{k}^2<k_c^2} \frac{e^{-k^2/4\alpha^2}}{k^2} \mathcal{I}m\left(e^{j\boldsymbol{k}\cdot\boldsymbol{r}_i}\rho(\boldsymbol{k})\right)\boldsymbol{k} \\
&+ \frac{1}{2} q_i \sum_{j}^{\ddagger -1} q_j \left[\frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 r_{ij}^2} - \frac{\operatorname{erf}(\alpha r_{ij})}{r_{ij}}\right] \frac{\boldsymbol{r}_{ij}}{r_{ij}}
\end{aligned} \tag{6.28}$$

$$\begin{aligned}
3\mathrm{V}\,\underline{\sigma} &= \frac{1}{2} \sum_{i\neq j,\ r_{ij}<r_c}^{\mathrm{N}\ \ddagger} q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} + \frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 r_{ij}^2}\right] \frac{\boldsymbol{r}_{ij}\otimes\boldsymbol{r}_{ij}}{r_{ij}^2} \\
&+ \frac{4\pi}{\mathrm{V}} \sum_{\boldsymbol{k}\neq 0,\ \boldsymbol{k}^2<k_c^2} \frac{e^{-k^2/4\alpha^2}}{k^2} \left|\rho(\boldsymbol{k})\right|^2 \left[\mathbb{1} - \frac{2}{1/k^2 + 1/4\alpha^2}\boldsymbol{k}\otimes\boldsymbol{k}\right] \\
&+ \frac{1}{2} \sum_{i\neq j}^{\ddagger -1} q_i q_j \left[\frac{2\alpha}{\sqrt{\pi}} e^{-\alpha^2 r_{ij}^2} - \frac{\operatorname{erf}(\alpha r_{ij})}{r_{ij}}\right] \frac{\boldsymbol{r}_{ij}\otimes\boldsymbol{r}_{ij}}{r_{ij}}
\end{aligned} \tag{6.29}$$

Here, $\mathcal{I}m$ is the imaginary part of a complex number. We note that the self interaction and charge correction do not contribute to the forces or stress.

Implementation tricks

When using the Ewald method for electrostatic potential computation, the short-range contribution is the easiest to compute, as it behaves like a standard pair potential. The long-range interaction is usually harder, and needs more fine-tuning. A first easy improvement is to compute separately the factors only depending on $\boldsymbol{k}$ vectors (such as $\exp(-k^2/4\alpha^2)/k^2$ and $2/(1/k^2 + 1/4\alpha^2)$), and the Fourier transform of charge density $\rho(\boldsymbol{k})$. This allows re-using the factors as long as the unit cell does not change, and only recompute the density. This also helps to improve efficiency for Monte Carlo simulations, using the knowledge that only some atoms moved to update $\rho(\boldsymbol{k})$ (see figure 6.2 and accompanying text for more information on this strategy).

The choice of storage for the multiplicative factors and $\rho(\boldsymbol{k})$ also has a large impact on performance. As these quantities depend on $\boldsymbol{k}$ vectors, it is natural to store them in three-dimensional arrays, corresponding to the three components of the vector. But this strategy induces a lot of jumps when accessing memory (called cache misses), lowering performance (see section 6.1.2 for a performance model of modern computers). A better strategy consists in using a first linear array to store the $\boldsymbol{k}$ vectors, each vector being

associated with an index. Then, the multiplicative factors and $\rho(\boldsymbol{k})$ can also be stored in linear arrays, the corresponding $\boldsymbol{k}$ vector for each entry being found by the entry index. This drastically improves memory locality and performance of the calculation.

6.3.3 Wolf summation

The Wolf summation technique, proposed by the eponym author in 1999[238], is based on the observation that the effective electrostatic potential created by a multipole decays faster than $1/r$ at long distances. For example, dipolar potential decays as $1/r^2$, quadrupolar moment decays as $1/r^3$, and Wolf showed that complex multipoles such as crystals create a potential that decays as $1/r^5$. It should thus be possible to use a direct pair sum to evaluate it. The authors showed that the direct pair sum does not converge when using spherical cutoff because the system included inside the cutoff is not charge-neutral. They proposed to include a corrective term for the remaining charge, effectively replacing the electrostatic potential by a shifted potential:

$$V_{SP}(r_{ij}) = q_i q_j \left(\frac{1}{r_{ij}} - \frac{1}{r_c} \right) \tag{6.30}$$

The energy resulting from this potential is a good approximation of the true electrostatic potential, while being a lot easier to compute. It gets closer to the real energy as the cutoff distance is increased.

Because the naive shifted potential presented above requires relatively large cutoff radius to offer a good approximation of the energy, the authors also proposed to *damp* the potential using the complementary error function. This allow the use of smaller cutoff radius (10 – 20 Å), while still being a good approximation of the true electrostatic potential. This *damped shifted potential* (DSP) — and the corresponding force expression — are given by:

$$V_{DSP}(r_{ij}) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\operatorname{erfc}(\alpha r_c)}{r_c} \right] \tag{6.31}$$

$$\boldsymbol{f}_{DSP}(\boldsymbol{r}_{ij}) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}^2} + \frac{2\alpha}{\sqrt{\pi}} \frac{\exp(-\alpha^2 r_{ij}^2)}{r_{ij}} \right] \frac{\boldsymbol{r}_{ij}}{r_{ij}}, \tag{6.32}$$

where α is the damping parameter. This expression for the forces presents a discontinuity at $r_{ij} = r_c$, which can introduce artifacts in molecular dynamics simulation as particles enter or leave the cutoff sphere.

In 2006, Fennell et al.[239] proposed to remove this discontinuity by changing the potential to:

$$V_{DSF}(r_{ij}) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\operatorname{erfc}(\alpha r_c)}{r_c} + (r_{ij} - r_c) \frac{\partial V_{DSP}(r_{ij})}{\partial r_{ij}}(r_c) \right]. \tag{6.33}$$

This approach gives us the *damped shifted force* (DSF) formulation of Wolf potential:

$$V_{DSF}(r_{ij}) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\operatorname{erfc}(\alpha r_c)}{r_c} + \left(\frac{\operatorname{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\sqrt{\pi}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} \right) (r_{ij} - r_c) \right] \tag{6.34}$$

$$\boldsymbol{f}_{\text{DSF}}(\boldsymbol{r}_{ij}) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r_{ij})}{r_{ij}^2} + \frac{2\alpha}{\sqrt{\pi}} \frac{\exp(-\alpha^2 r_{ij}^2)}{r_{ij}} - \frac{\operatorname{erfc}(\alpha r_c)}{r_c^2} - \frac{2\alpha}{\sqrt{\pi}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} \right] \frac{\boldsymbol{r}_{ij}}{r_{ij}} \quad (6.35)$$

Fennell et al. also compared the these approaches on a large set of systems, and concluded that the DSF approach is able to reproduce the forces computed by Ewald summation very well with a damping parameter $\alpha = 0.2\,\text{Å}^{-1}$; while the DSP approach was able to reproduce energies using the same damping parameter. They suggest using the DSP approach for Monte Carlo simulation, and the DSF approach for molecular dynamics.

6.3.4 Comparing Ewald and Wolf summations

I implemented both the DSP and the DSF formulation of Wolf summation in Domino. As they are based on sums of pair terms, they were much easier to implement and integrate with Monte Carlo energy caching than Ewald summation. Being simpler, the computation time for both energy and forces is smaller for Wolf summation than for Ewald: for a small system containing 50 water molecules, forces computation takes 755 µs with Wolf, and 1810 µs with Ewald (respectively 755 µs and 860 µs for the energy).

To check that my implementation of both Ewald and Wolf was correct, and to compare their behavior in simple cases, I ran simulations in the NVT ensemble of NaCl in crystalline form at 300 K and molten at 2000 K using both Domino and LAMMPS. I ran both Monte Carlo simulations, and molecular dynamics simulations with Domino. I also ran molecular dynamics with LAMMPS as a reference implementation. The molecular dynamics used a time step of 1 fs and a Berendsen thermostat to fix the temperature, with a thermostat time constant of 100 fs. Both Ewald and Wolf simulations used a cutoff radius of 11 Å and a damping parameter α of $0.29\,\text{Å}^{-1}$. The Ewald simulation considered $\boldsymbol{k}$ vectors up to 10 translations in reciprocal space.

The system contained 256 Na^+ and 256 Cl^- atoms, placed on a face-centered cubic lattice, in a cubic simulation cell of 22.56 Å. The molten NaCl simulations were started from the same conformation, placed into a bigger unit cell of 25.38 Å. The atoms interacted through both electrostatic and Lennard-Jones potentials, using parameters from reference [240]. From these simulations, I extracted the average pressure and potential energy, which are presented in table 6.1.

First, we can see that my implementation of Ewald gives the same average values for energy and pressure than LAMMPS, both for crystalline and molten NaCl; and both for molecular dynamics and Monte Carlo simulations. My implementation of Wolf also gives the same results with molecular dynamics and Monte Carlo, and the same pressures as LAMMPS. Strangely, it does not give the same average energies as LAMMPS, maybe because LAMMPS chooses a different energy origin. This does not mean that my implementation is incorrect, as energies are only defined up to an additive constant, but the origin of this difference should be investigated further.

Another proof that my implementation is correct comes from the radial distribution functions for all the simulation presented in figure 6.6. All simulations methods and electrostatic summations gives the exact same curves, except for rounding errors and statistical noise. This is also another evidence that Wolf summation is able to reproduce properties predicted by Ewald summation.

Table 6.1 – Energy and pressure (± standard deviation) for multiple simulations comparing Ewald and Wolf on crystalline and molten NaCl. MC denotes Monte Carlo simulations, and MD molecular dynamics.

Software	System	Potential	Simulation	Energy (kJ/mol)	Pressure (bar)
LAMMPS	Crystal	Ewald	MD	-203419 ± 45	13494 ± 302
		Wolf	MD	-203428 ± 45	13724 ± 297
	Molten	Ewald	MD	-182283 ± 345	15971 ± 1812
		Wolf	MD	-182424 ± 355	17977 ± 1770
Domino	Crystal	Ewald	MD	-203470 ± 53	13431 ± 229
		Ewald	MC	-203511 ± 62	13064 ± 383
		Wolf	MD	-86714 ± 69	13909 ± 355
		Wolf	MC	-86723 ± 69	13658 ± 455
	Molten	Ewald	MD	-182310 ± 278	15527 ± 1676
		Ewald	MC	-182818 ± 367	15090 ± 1620
		Wolf	MD	-65803 ± 306	18440 ± 1306
		Wolf	MC	-65785 ± 359	17686 ± 2054

In conclusion, Wolf summation seems to be able to replace Ewald summation, at least for simple cases. It can also be twice as fast as Ewald, and simpler to implement in simulation software. I did not have the time to use it in conjunction with hybrid simulations, but this would be a natural extension of this work.

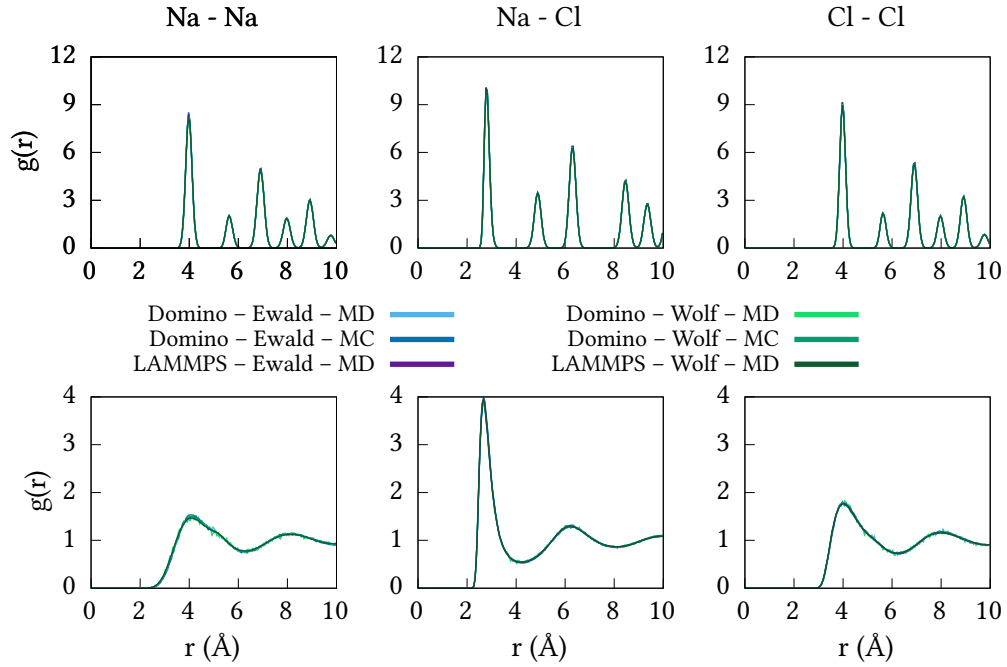


Figure 6.6 – Radial distribution function for three pairs (from left to right Na – Na, Na – Cl, and Cl – Cl) on two system: a solid NaCl crystal at 300 K *on top, and melted NaCl at* 2000 K *on bottom.*

GENERAL CONCLUSIONS

The work presented in this thesis is related to the study of adsorption and intrusion in nanoporous flexible materials, the deformation of these materials and the coupling between the two phenomena. Confining a fluid inside a porous network has significant effects on its thermodynamics properties, due to the competition between pore size and pore shape effects, and interface interactions. This competition generates specific new behaviors, such as new fluid phases and phases transitions, and is especially acute in nanoporous materials, where the typical width of the pore and the range of the interactions are of the same order of magnitude. On the other hand, the presence of a confined fluid can also have strong effects on the surrounding solid, creating the opportunity for new phases to exist and shifting the balance between multiple meta-stable phases. This is particularly poignant in the case of *flexible* nanoporous material, such as many metal–organic frameworks.

Because these materials are relatively recent, their flexibility has often been overlooked, and it was only in the recent years that the scientific community started to take it into account. An example of such shift is presented in the second chapter of this manuscript, with the incorporation of the osmotic thermodynamic ensemble into the Ideal Adsorbed Solution Theory (IAST) for the study of co-adsorption of gases, leading to the creation of the Osmotic Framework Adsorbed Solution Theory (OFAST). In the aforementioned chapter, I demonstrate that IAST is by construction invalid for the treatment of co-adsorption when the adsorbing host is not inert during adsorption. In particular, I show that IAST cannot be used for the prediction of co-adsorption of fluid mixtures in frameworks presenting a gate-opening behavior, and that it predicts non-physical selectivity, up to two orders of magnitude higher than OFAST. Even when IAST is not explicitly used to compute selectivity in flexible frameworks, researchers should be careful when comparing single-component isotherms of different guests in presence of flexibility. Differences in step pressure of stepped isotherms can lead to claims of strong selectivity, when applying concepts that are valid only for rigid host matrices.

We should also take care of not going too far the other way either, and attributing all observed behaviors to the flexibility of the material. In the third chapter of this document, I used *ab initio* molecular dynamics simulations to explain the origin of a stepped isotherm for the adsorption of nitrogen in ZIF-8(CH_3) and ZIF-8(Cl), and its absence in the similar framework of ZIF-8(Br). I showed that while the framework does deform during adsorption for both ZIF-8(CH_3) and ZIF-8(Cl), the deformations do not change the accessible volume and pore size distribution of these materials. Instead, the increase in uptake in the isotherm is linked to a reorganization of the fluid confined in

the pores, reorganization which does not happen in ZIF-8(Br) because of the difference in pore size. It is thus fundamental to account for both flexibility and confinement effects when studying adsorption in soft porous crystals.

The same is true of intrusion, adsorption's big brother. In chapter five, I used classical molecular simulations to study the confinement under high-pressure of water and aqueous electrolyte solutions in ZIF-8, and imogolite nanotubes. I observed confinement effects ranging from stronger spatial organization, changes in elastic properties, to water dynamics slowdown. Interestingly, the presence of ions at high concentrations can have the same effects on bulk water, structuring the hydrogen bonds network and slowing down dynamics. The intrusion of aqueous solution in hydrophobic material is a promising way to store and dissipate mechanical energy. Adding electrolyte in the water at varying concentration allows tuning the behavior and even switching from energy storage to dissipation. I looked at the impact of the ions on the intrusion behavior using umbrella sampling simulations to extract the free energy profile of entry inside ZIF-8, showing that different ions have dissimilar barriers when traversing ZIF-8 windows. This study is one of the first on the subject of intrusion of electrolytes in metal–organic frameworks, and allowed to shed some light on the complex behaviors emerging in these systems.

The need to simultaneously account for adsorption and deformations was a recurrent theme of all these studies. But current simulations methods are only able to address one dimension of the problem: molecular dynamics simulations can describe deformations, but modeling open systems and thus adsorption is not possible. Metropolis Monte Carlo simulations can be used for open systems, but have difficulties efficiently sampling collective deformations. Hybrid Monte Carlo is a possible answer to this dilemma, combining the efficiency of molecular dynamics with the versatility of Monte Carlo simulations (in particular the possibility of sampling open and extended ensembles). My last chapter presents the hybrid Monte Carlo simulation method and how to use it for direct simulations in the osmotic ensemble. Increasing the reach of such novel methods within the scientific community requires them to be readily available in generic, easy to use, and efficient software. I explored the design space for such software with the Domino project, described in the last chapter of this document.

There is another condition to fulfill before being able to widely use osmotic ensemble simulations for the study of adsorption and intrusion in soft porous crystals. We need to be able to predict the energy changes related to the flexibility of the frameworks and their interactions with fluids. On one hand, first-principle or *ab initio* methods (such as the density functional theory) enable to accurately compute the energy of arbitrary atomistic systems. On the other hand, they require large amount of computational power, which prevents them from being routinely used on large systems. When faced with such large systems — large in the number of atoms, the timescale of processes, or for high-throughput screening — we therefore often turn to classical force fields instead.

Classical force fields are either accurate, *i.e.* reproducing well the actual potential energy surface; or transferable, *i.e.* usable with multiple different systems. Current transferable force fields are not well suited to describe the flexibility arising from coordination bonds, so we need to create new force fields for these systems. Historically, the parametrization of new force fields has been a rather long and tedious process. In recent years, new

machine learning-based techniques for the consistent and fast derivation of accurate force fields have been devised. I presented one of these techniques in the fourth chapter of this document, and used it to derive force fields for ZIF-8 and some of its derivatives from *ab initio* data. Such automated methods are especially crucial for the study of metal–organic frameworks given the huge diversity of their structures. I hope that the availability of both accurate force fields and hybrid Monte Carlo simulations capable software will make it easier to use molecular simulations to engineer new materials tailored for specific applications.

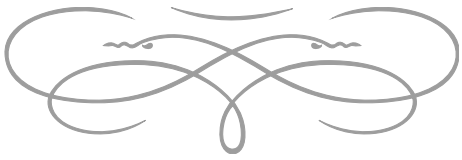
Overall, this PhD presents multiple molecular modeling methods that can be used for the study of adsorption and intrusion in flexible nanoporous materials. From *ab initio* simulations to remove the need for force field parametrization; classical simulations using molecular dynamics to describe flexibility; Monte Carlo and particularly Grand Canonical Monte Carlo for adsorption; free energy methods such as umbrella sampling; to hybrid Monte Carlo for the direct simulation of collective behaviors in open systems; all the way to macroscopic modeling methods, based on classical thermodynamics.

This work opens perspectives for improvements in various directions. Concerning molecular simulation methods, hybrid Monte Carlo seems like a very powerful technique, that can be used for a wide variety of systems outside of the problem of adsorption in soft porous crystals. First, hybrid Monte Carlo, being a Metropolis Monte Carlo method, will always converge to and sample the phase space distribution of the correct statistical thermodynamic ensemble. This is in opposition with molecular dynamics, which samples the micro-canonical ensemble by default, and relies on thermostats and barostats to sample other ensembles. These thermostats and especially barostats are not all equal, and only some algorithms are able to precisely generate the correct ensemble. At the same time, hybrid moves greatly improve the efficiency of Monte Carlo simulations by taking into account the local curvature of the potential energy surface.

Whether the simulation of open systems is compatible with molecular dynamics is still an open research question. On the opposite, such simulations are routinely performed using Grand Canonical and Gibbs ensemble Monte Carlo. Hybrid Monte Carlo could thus be used for simulations of open ensembles and dilute systems, such as constant pH simulations, description of the ionic environment of proteins, or the simulation of defects in crystalline materials, improving efficiency compared to standard Monte Carlo, and giving access to such open systems to molecular dynamics users. Thanks to the Metropolis Monte Carlo acceptance scheme, there is no need for the short molecular dynamics run used inside hybrid move to sample any actual thermodynamic ensemble or any physical Hamiltonian. This property could be exploited to create even more efficient Monte Carlo simulations, for example gradually inserting new molecules in a system while simultaneously relaxing its environment. Compressible generalized hybrid Monte Carlo is particularly promising, as it enables using a custom Hamiltonian tailored for the problem at hand.

Another perspective for future work concerns classical force fields, and their ability to accurately reproduce potential energy surfaces. The traditional approach to force fields has been to decompose the energy as a sum of terms depending on simple scalar values

with physical meaning: distances, bond length, bond angles, torsion dihedral angles, *etc.*. These scalar values are then combined with simple mathematical expressions, such as power or exponential functions. This approach prevents the potential energy surface from including multi-body effects, and accurately reproducing the shape of *ab initio* potential energy surfaces used as references. Machine learning tools, in particular neural network and Gaussian processes, can improve both areas. First, neural-networks ability to reproduce arbitrary functions from $\mathbb{R}^n$ to $\mathbb{R}$ can reduce the differences between the potential energy surfaces shapes. For example, instead of imposing the Lennard-Jones functional form, neural networks can reproduce the actual function. Secondly, machine learning algorithms can be coupled with better descriptors of the atomic structure, accounting for many-body effects. In recent years, multiple independent scientific teams worked to design, train and evaluate such machine learning force fields and the associated descriptors. To my knowledge, they remain to be used for the simulation of nanoporous flexible materials.



A

CHEMFILES: A LIBRARY FOR CHEMISTRY INPUT/OUTPUT

During my PhD, I also wrote chemfiles, an open source library allowing to read and write multiple file formats used in computational chemistry. I will present here the software, its goals and some software architecture choices I made to reach these goals. Chemfiles is available online at `https://chemfiles.org` and `https://github.com/chemfiles`.

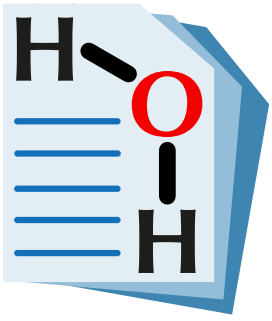


Figure A.1 – Chemfiles' project logo

A.1 The computational chemistry formats zoo

A recurrent pain point for anyone working in the theoretical and computational chemistry field is the multiplication of file formats. Every simulation software comes with its own set of files formats. And it is up to the user to adapt, and learn how to work with the set of formats used by specific software. These formats are used to store data generated by simulation software, before analyzing it either through visualization of the atoms and their individual motions; or by computing properties of the system from the positions and velocities, using the framework of statistical thermodynamics presented in section 3.1.

At the same time, every format contains the same kind of information: atomic names, positions, velocities, forces, and topological information (bonds, residues, *etc.*). Different formats exist because, in addition to storing these basic data, they have different use cases. For example, the XYZ format is very simple to read an write, both by hand and when creating new software. Binary formats, such as NetCDF or TNG are able to store huge trajectories efficiently, both in term of disk size and reading speed. The PBD, mmCIF and MMTF format are optimized to store proteins and other bio-molecules.

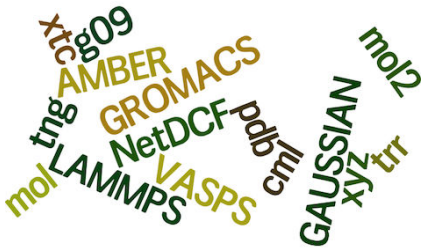


Figure A.2 – Some of the existing formats used in computational chemistry.

This multiplication of formats means that there is less interoperability between the different software used in computational chemistry. Analysis and visualization software each support specific formats, which are not always the same as the ones produced by simulation software. To run specific analysis on the results of a simulation we have to either use readily available algorithms, or learn how to read the format of our current software, and implement the analysis ourselves. This friction can prevent switching between simulation software or using specific analysis methods, and gets in the way of using the tools of molecular simulation to improve our understanding of the world.

Chemfiles is an attempt to provide an unified and simple interface for programmers to work with all these file formats. In order to do so, it defines a representation for the data that can exist in various formats, and transparently reads and writes files from and to this representation. Which means that software developers only need to learn how to use chemfiles, and can then use it to read or write every supported file format. Chemfiles is a software library; a collection of functions and classes meant to be integrated into other software. It can be used in simulation software (for example Domino uses chemfiles for reading the initial configuration and writing the trajectory); visualization or analysis software.

A.2 An overview of chemfiles

A.2.1 Goals and non-goals

The main goal of chemfiles is to be useful to the research community at large. To be useful, chemfiles must first be usable, and different researchers and software developers work with different computing environments. Chemfiles is portable between different platforms, operating systems (Windows, GNU/Linux, macOS), word sizes (32 or 64-bits environments) and CPU architecture (Intel, ARM, PowerPC, …). It is also usable from multiple languages, to allow any computation chemistry project to use it, regardless of the implementation language the authors choose. The core of the library is written in C++, and it offers interfaces to C, Python, Fortran, Julia, and Rust.

Computational chemistry is used with a wide range of systems, and each kind of system has specific requirements, which chemfiles tries to support. For example, simulations of bio-molecules such as proteins or nucleic acid strands use complex topologies, grouping some atoms in residues or monomers. Oftentimes, this topology is stored in a separated file. While bio-molecules simulation mostly only needs orthorhombic unit cells, material science and crystallographic data need to describe triclinic unit cells, where some of the angles are not 90°. Finally, when working with simulations in the grand canonical ensemble such as GCMC, the number of atoms in the system changes along the trajectory. Support for such simulations where the number of atoms is not constant is often missing from existing software.

Chemfiles also has explicit non-goals: features that should not be part of chemfiles itself, but could instead be built on the top of it. For example, trajectory analysis, energy minimization or simulation algorithms would add too much complexity to the code. Chemfiles is also actively trying **not** to create a new format. Instead, it focuses on providing interoperability between existing formats.

A.2.2 Existing alternatives

OpenBabel

OpenBabel[241] is C++ library that provides read and write capabilities to other software. It is a well-established software supporting more than 110 different file formats. Unfortunately, it only supports text-based formats, and no binary formats such as Amber NetCDF or Gromacs TNG. Such binary formats are used to store big trajectories, are faster and take less space than equivalent text formats.

Two other caveats of OpenBabel for me were the complexity of the programming interface, and its license. OpenBabel is distributed under the Gnu Public License version 2 (GPL-v2), which requires any project using it to be released under the same license. This is due to the project's history, which was created from the OELib project, itself distributed under the GPL license. This requirement makes it harder to use OpenBabel in projects which use another license, either another open source license or proprietary software developed by a private company.

Concerning the programming interface, OpenBabel is written with the C++ 98 standard, and a lot of member functions return raw pointers to internal data directly, instead of the more modern references or smart pointers introduced with C++ 11. This type of interface

means that the programmer using the library needs to manage memory manually, which in turn introduces more memory bugs in the applications. Furthermore, I find the overall interface to be complex and hard to learn, with many different functionalities not directly related to file reading and writing. For example, it contains functions to search low energy conformers, to create graphical images of molecules or to manage chemical reactions. All of these functionalities are useful for cheminformatics applications, but can clutter other uses of OpenBabel.

Other alternatives

Table A.1 – Summary of existing popular software library providing read/write capabilities for chemistry file formats.

Project	language	C++ compatible	GCMC	license
OpenBabel[241]	C++	✓	✗	GPL-2
VMD[242]	C/C++	✓	✗	BSD
MDAnalysis[243]	Python	✗	✗	GPL-2
cclib[244]	Python	✗	✗	BSD
ASE[245]	Python	✗	✗	LGPL
CDK[246]	Java	✗	✗	LGPL

Multiple other alternatives to OpenBabel also exist, with different set of goals and implementations; I summarize the most popular ones in table A.1. Unfortunately, none of these alternatives support GCMC simulations, with a number of molecules which can vary during the simulation. Other than OpenBabel, only VMD molfile plugins (written in C and C++) are compatible with C++ simulation software.

I could not find an existing library that would cover all of my needs, which is why I started working on chemfiles. With this project, I hope to provide new options to developers of chemical scientific software, and tools for users of these software, improving interoperability. Chemfiles is open source, distributed under the BSD license at `https://github.com/chemfiles/chemfiles`. This license roughly means that anyone can download, use, modify and distribute the code — even together with software released under different licenses — provided that give attribution the original authors.

A.2.3 Some code statistics

I started working on chemfiles in December 2014, and released the version 0.1 the 16th of May 2015. Since then, I released 8 new versions, the last one being version 0.9, which was released on the 31 of March 2019. During this period, they have been 1500 git commits (self-contained modifications) in the core C++ chemfiles library, leading to a library containing 23 000 lines of code, as well as 11 000 lines of test code. Bindings to chemfiles in other languages (Python, Fortran, Rust, Julia) collectively contains 800 commits, 12 000 lines of code, and 5 000 lines of tests.

A few people helped me making chemfiles what it is today, I would like to thank them all here again: Jonathan Fine from Purdue University (USA); Patricio Germán Barletta from Universidad Nacional de Quilmes (Argentina); Laura Scalfi and Elsa Perrin from École Normale Supérieure (France).

A.3 Architecture and functionalities

In this section, I will describe the general architecture of chemfiles, and then dive into some specific features which I believe to be interesting in this context. In particular, I will describe how one can store additional metadata in a statically compiled language; the implementation and specificities of chemfiles' atomic selection language; and how to use a software library from multiple different programming languages.

A.3.1 Public classes

The full interface to chemfiles is contained into eight public classes, which are represented in figure A.3. The starting point is the `Trajectory` class, which provides functions to read or write a `Frame`, as well as managing metadata associated with the trajectory. The `Trajectory` does not do the actual work of parsing and formatting data, which is delegated to the `Format` classes, presented in section A.3.2.

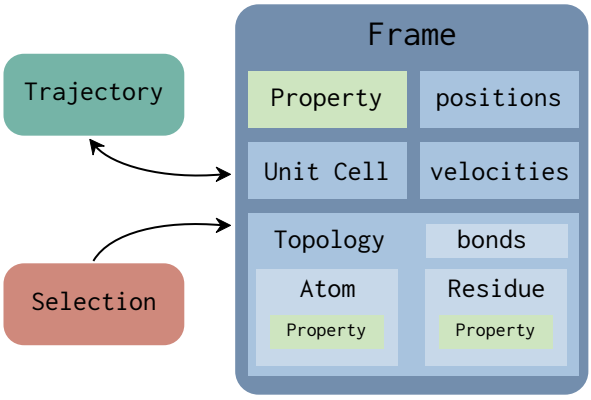


Figure A.3 – Relation between chemfiles' public classes.

The `Frame` class is the central class of chemfiles, it stores all the data associated with a simulation step: positions, velocities, `UnitCell` and `Topology`, and various properties. The `UnitCell` class describes the unit cell of the simulation, and provides calculation of distances that accounts for periodic boundary conditions. The `Topology` contains the atoms and connectivity elements of the system: bonds, angles, dihedral angles, and improper dihedral angles; but also residues. `Atoms` store basic data associated with each atom: its name and type as character strings, its mass and charge as floating point values. For some formats, the atomic name and type are different: the atomic type will identify an atom or an extended particle ("He", "Ow", "CH4"); while the atomic name might be a unique identifier of this specific atom ("C_a1", "Ow22"). `Residue` stores additional atomic grouping, such as molecules, amino- or nucleic-acids, *etc.*

The `Property` class can be associated with either a frame, an atom or a residue. It stores additional metadata that only occurs in some specific formats, such as atomic hybridization state, molecule name or chain identifier in bio-molecules. Finally, the `Selection` class implements chemfiles' atomic selection language, which can be used to get a list of atoms matching some constraints, such as `name == O and x < 22` or `pairs: name(#1) O and name(#2) F and is_bonded(#1, name C)`. I will describe this functionality with more details in section A.3.4.

A.3.2 Files and formats

As mentioned above, chemfiles' trajectory class does not read or write frames by itself. Instead, it uses a subclass of the `Format` pure virtual abstract base class. Each supported format is a specialization of this base class, implementing at least one of three operations: read the next step of the trajectory, read an arbitrary step of the trajectory, or write a single frame to the trajectory. Implemented formats as of the 0.9 version of chemfiles are listed in table A.2.

Table A.2 – List of format supported by chemfiles as of the 0.9 release, with associated file extension, classification as text or binary format and read or write support.

Format	extension	text based	can read	can write
Amber NetCDF[247]	.nc	✗	✓	✓
CSSR	.cssr	✓	✓	✓
DCD	.dcd	✗	✓	✗
GRO	.gro	✓	✓	✓
LAMMPS Data		✓	✓	✓
LAMMPS	.lammpstrj	✓	✓	✗
mmCIF[248]	.mmcif	✓	✓	✓
MMTF[249]	.mmtf	✗	✓	✓
MOL2	.mol2	✓	✓	✓
Molden	.molen	✓	✓	✗
PDB[250]	.pdb	✓	✓	✓
SDF	.sdf	✓	✓	✓
TNG[251]	.tng	✗	✓	✗
TRJ	.trj	✗	✓	✗
TRR	.trr	✗	✓	✗
Tinker	.arc	✓	✓	✓
XTC	.xtc	✗	✓	✗
XYZ	.xyz	✓	✓	✓

There are two main types of formats: text-based formats store data in a text file, which must be parsed to extract the data. Some notable examples of text-based formats include the XYZ, PDB and CIF formats. In all text formats, floating point data such as positions will be stored as the list of characters representing their digits in base 10, which must be parsed to reconstruct the data.

The other class of formats is binary formats, including Amber NetCDF, GROMACS TNG or XTC. In these formats, data is directly stored as a set of bytes in the file, which can directly be interpreted as the corresponding floating point type without additional transformations (except for managing the order in which bits are written inside a byte, also called *endianess*). Using a binary format allows to use less space to store the same data: for example the number 1.23456789 requires 10 bytes when represented as text using UTF-8 encoding, and only 4 as single precision IEEE-754 floating point number. Some formats such as XTC uses additional molecular simulation specific encoding (such as reduced decimal precision up to 10^{-3} Å) to reduce storage size even further. This is particularly important for long simulations of large systems, where the trajectory size

can easily grow to multiple hundreds of gigabytes. All these features combine to make binary formats more space- and time-efficient than text-based formats: I found Amber NetCDF trajectories to take roughly five times less space than the corresponding XYZ file, and be five times faster to read.

Most binary formats come with a C or C++ software library that should be used to read or write them, or use another generic binary format with conventions specific to molecular simulations. This is the case of Amber NetCDF, using the NetCDF format[247], MMTF using the MessagePack format[249], or H5MD using HDF5 data storage[252]. This is their main shortcoming when compared to text-based formats, which does not require special software to create or edit a file: any text editor can be used.

In chemfiles, the file being read or written is managed by a subclass of either the `TextFile` or `BinaryFile` classes. `BinaryFile` mainly provide C++ style memory management with constructors and destructors on the top of the format library data structures. Text-based formats use `TextFile` which works slightly differently: the base class defines an interface (based on C++ streams), which is then implemented by multiple child classes. The formats can use any implementation to get access to the data. In addition to the obvious basic file reader, chemfiles uses this to transparently read and write compressed files with either the gzip (.gz files) or lzma (.xz files) compression algorithms. All text-based format thus directly support reading and writing compressed files without any specific code inside the format class. In the future, this functionality could be expanded to transparently use memory-mapped input/output which can be faster for some files; and to download remote files from the internet.

A.3.3 Storing additional data: properties

Most formats store the same basic subset of data from a molecular simulation: positions, velocities, atomic names, and connectivity. But they also store additional data or metadata, which are specific to one or few file format. Being able to read, store and possibly write back this data is important for two reasons: it increases the amount of data that can be converted from one format to another through chemfiles, assuming both format support storing it. Secondly, it allows users of chemfiles working with a specific format to still access all the data defined in the format.

Adding new attributes to atoms, frames or residues for every possible property would not be tractable. As most metadata are only defined by one or two formats, adding a new data member in the classes would be a waste of memory. It would also increase the complexity of the code, as implementing a new format would require adding new data members not only to the core C++ classes, but also to the corresponding objects in the other languages bindings. Adding new data member also means that it would be impossible to maintain a stable binary interface between releases of chemfiles, as the size of classes would change.

Chemfiles solution is to introduce a dynamically typed property, which can be attached to frames, atoms or residues in associative maps. The associative map links each property with a property name, which can be used to check if the property is defined on a given frame, atom or residue. The `Property` class is a tagged union, containing a type tag which can change at run-time and the associated data. It can contain either a character string, a

Boolean value (*true* or *false*), a number stored a double precision IEEE-754 floating point or a three-dimensional vector as three double precision floating point numbers. These four types cover the vast majority of data found in molecular simulation: for example, atomic forces can be stored as three-dimensional vector atomic properties, and the user who created a particular file can be stored as character string associated with the frame. This strategy trades a small performance penalty at runtime — the code must check if a property exists, and if it has the right type – for a huge flexibility when storing additional data and metadata.

A.3.4 Atomic selection language

Chemfiles implements a rich selection language that allows finding the atoms in a frame that match a set of constraints. For example, `atom: name == H and x > 15` would select all atoms with a name equal to "H" and x cartesian coordinate bigger than 15. Here, `name == H` and `x > 15` are individual constraints, and they are combined with `and`, meaning both of them must be true for an atom to match to full selection.

Multi-atoms selections

Chemfiles atomic selection language differs from other implementations such as the one in VMD, MDAnalysis, or ASE (the Python Atomistic Simulation Environment) by the fact that it is possible to formulate constraints not only on single atoms, but also on pairs, triplets, and quadruplets of atoms. For example, `angles: name(#2) O and mass(#3) < 1.5` will select all sets of three bonded atoms forming an angle such that the name of the second atom is O and the mass of the third atom is less than 1.5. Here, the first atom is left unconstrained. Where evaluating simple selections yields a list of matching atomic indexes, evaluating triplet selection will return a list of triplets of atomic indexes (and correspondingly for pairs and quadruplets).

The number of atoms to select together is indicated in chemfiles by a context, separated from the main selection by a colon. Seven contexts are available: **`atoms`** is the default context, matching single atoms. **`two`**, **`three`**, and **`four`** match arbitrary pairs, triplets and quadruplets respectively. **`bonds`**, **`angles`**, and **`dihedrals`** match pairs, triplets and quadruplets of atoms bonded together to form the corresponding connectivity element.

The selections are built by assembling simple constraints with Boolean operators: **`and`**, **`or`** and **`not`**. They follow the usual interpretation of logic: `A and B` will be true only if both A and B are true, `A or B` will be true if one of A or B is, and `not A` will be true is A is false. The constraints can check text values, such as **`name`** (atomic name), **`type`** (atomic type), **`resname`** (residue name), with a fixed string (`name(#1) == Fe`) or another text value (`type(#1) != type(#3)`). There are also constraints on numeric values, for example **`x`**, **`y`**, **`z`** for the position of the atoms, **`mass`** for the atomic mass, **`index`** for the index of the atom in the frame, *etc.* Numeric values can be combined with the usual mathematical operations: `x^2 + y^2 + z^2 < 10^2` will check for atoms inside the sphere with a radius of 10 Å centered on the origin. Finally, Boolean constraints such as `is_bonded(i, j)`, `is_angle(i, j, k)` or `is_dihedral(i, j, k, m)` are used to query the connectivity of atoms. Here, the arguments to `is_bonded` and the others can either be one of the atoms currently being matched (`#1`, `#2`, `#3`, `#4`) or a sub-selection, and every atom matched by the sub-selection will be checked. This makes for example `is_bonded(#1, name O)` select all atoms bonded to an oxygen.

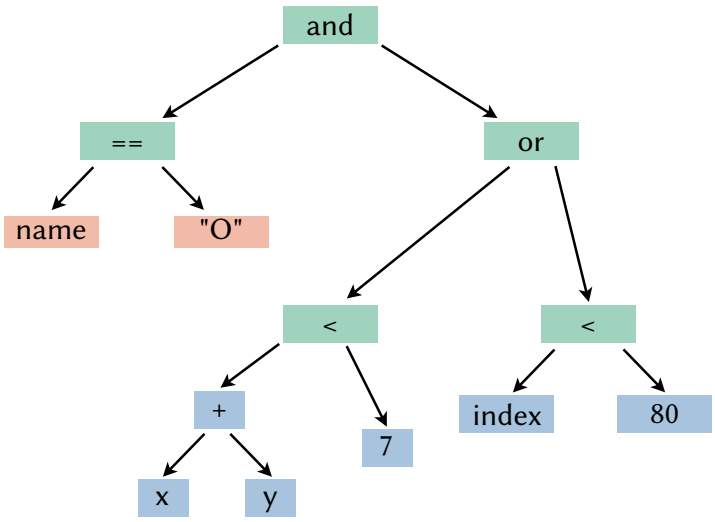


Figure A.4 – Abstract syntax tree for the selection `name O and (x + y < 7 or index < 80)`*. Nodes in green return a Boolean value, nodes in red a character string and nodes in blue deal with numbers.*

Implementation strategy

The transformation from a character string describing the constraints in a domain specific language to a `Selection` object able to find the atoms matching these constraints go through three main steps. First, the string is transformed into a stream of simple tokens during an operation called *lexing*. For example, a single parenthesis **`(`** is a token, each mathematical operator (**`+`**, **`-`**, **`*`**, **`/`**, **`^`**) is its own token, keywords such as **`and`** and **`or`** are tokens, and single words such as **`name`** or **`index`** are tokens.

Then, the tokens are checked for specific patterns during *parsing*. For example, if the tokens **`name`**, **`==`** and **`"O"`** are found in this order, we know we have found a constraint stating that the atom name must be equal to "O". Multiple different algorithms are available to implement the parsing step, and chemfiles uses a recursive descent parser. This kind of parser is built around multiple functions that can call each other (hence the *recursive* part of recursive descent), each function being able to parse a small subset of the possible expressions. There is a function that recognize additions and subtractions ($\cdots$ **`+|-`** $\cdots$), another one that deals with multiplication and divisions ($\cdots$ **`*|/`** $\cdots$), *etc.* By combining these function, it is then possible to parse complex expressions with the right precedence. An example of recursive calling is the parsing of `3 + 4 * (5 + 2)`. First the function for additions is called, and it calls the one for multiplication to parse `4 * (5 + 2)`, which itself calls the function for addition to parse `5 + 2`.

Finally, an *Abstract Syntax Tree* (AST) is built from the recognized constraints. Such tree is illustrated in figure A.4 for the selection `name O and (x + y < 7 or index < 80)`. In this tree, each node knows how to evaluate itself, by first asking its left and right children to evaluate themselves and combining the results. AST in chemfiles deal with three types of values: Boolean nodes create or combine Boolean values, string nodes create character strings, and numeric nodes create or combine numbers.

To get the list of atoms matching the selection, the AST is walked from top to bottom, evaluating individual components and combining them to build the full answer to the question "does this set of atoms fulfill these constraints?".

Conclusions

In this appendix, I presented chemfiles, an open source C++ software library that can be used to read and write a variety of file format produced by computational chemistry software. Chemfiles currently supports 18 different file formats, and I plan to add support for more popular formats, such as the Crystallographic Interchange Format (CIF) and Protein Structure Files (PSF). It also supports reading and writing compressed files for all text-based format.

The main goal of the chemfiles project is to produce a library simple enough that scientist can learn it quickly and use it for their own needs. To do so, it has a simple structure, with only seven public classes. Additional data and metadata are represented as dynamically typed properties, keeping the overall size of the programming interface small.

Chemfiles also features a rich and innovative atomic selection language, able to use constraints on multiples atoms at once. While these selections are not strictly central to file input and output, they are a central building block for analysis and visualization software. The complexity associated with the implementation and performance of selections justify their inclusion in a core, shared library.

I used chemfiles in the `cfiles` command line tool to implement analysis algorithms for molecular simulation. `cfiles` can compute radial distribution functions; angles and dihedral angles distributions, density profiles along a variety of axes (cartesian and radial) and density maps, elastic constants from unit cell fluctuations in NPT simulations[253], hydrogen bonds and hydrogen bond autocorrelation, and rotational autocorrelation. The autocorrelation of time series is computed with a Fast Fourier Transform (FFT) approach, as the FFT algorithm is $\mathcal{O}(n \log n)$ in the size of the input instead of $\mathcal{O}(n^2)$ for the naive version[254]. `cfiles` also provides some tools to manage trajectory data, converting from a format to another and merging multiple trajectories together.

LIST OF PUBLICATIONS

Publications related to this work

1. Jack D. Evans, Guillaume Fraux, Romain Gaillac, Daniela Kohen, Fabien Trousselet, Jean-Mathieu Vanson, and François-Xavier Coudert. "Computational Chemistry Methods for Nanoporous Materials". In: *Chemistry of Materials* 29.1 (Sept. 2016), pp. 199–212. DOI: 10.1021/acs.chemmater.6b02994.

2. Guillaume Fraux and François-Xavier Coudert. "Recent advances in the computational chemistry of soft porous crystals". In: *Chemical Communications* 53.53 (2017), pp. 7211–7221. DOI: 10.1039/c7cc03306k.

3. Guillaume Fraux, François-Xavier Coudert, Anne Boutin, and Alain H. Fuchs. "Forced intrusion of water and aqueous solutions in microporous materials: from fundamental thermodynamics to energy storage devices". In: *Chemical Society Reviews* 46.23 (2017), pp. 7421–7437. DOI: 10.1039/c7cs00478h.

4. Gérald Chaplais, Guillaume Fraux, Jean-Louis Paillaud, Claire Marichal, Habiba Nouali, Alain H. Fuchs, François-Xavier Coudert, and Joël Patarin. "Impacts of the Imidazolate Linker Substitution (CH3, Cl, or Br) on the Structural and Adsorptive Properties of ZIF-8". In: *The Journal of Physical Chemistry C* 122.47 (Oct. 2018), pp. 26945–26955. DOI: 10.1021/acs.jpcc.8b08706.

5. Guillaume Fraux, Anne Boutin, Alain H. Fuchs, and François-Xavier Coudert. "On the use of the IAST method for gas separation studies in porous materials with gate-opening behavior". In: *Adsorption* 24.3 (Mar. 2018), pp. 233–241. DOI: 10.1007/s10450-018-9942-5.

6. Laura Scalfi, Guillaume Fraux, Anne Boutin, and François-Xavier Coudert. "Structure and Dynamics of Water Confined in Imogolite Nanotubes". In: *Langmuir* 34.23 (May 2018), pp. 6748–6756. DOI: 10.1021/acs.langmuir.8b01115.

7. Johannes P. Dürholt, Guillaume Fraux, François-Xavier Coudert, and Rochus Schmid. "Ab Initio Derived Force Fields for Zeolitic Imidazolate Frameworks: MOF-FF for ZIFs". In: *Journal of Chemical Theory and Computation* 15.4 (Mar. 2019), pp. 2420–2432. DOI: 10.1021/acs.jctc.8b01041.

8. Guillaume Fraux, Anne Boutin, Alain H. Fuchs, and Franccois-Xavier Coudert. "Structure, Dynamics, and Thermodynamics of Intruded Electrolytes in ZIF-8". In: *The Journal of Physical Chemistry C* (June 2019). DOI: 10.1021/acs.jpcc.9b02718.

9. Guillaume Fraux, Siwar Chibani, and François-Xavier Coudert. “Modeling of framework materials at multiple scales: current practices and open questions”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 377.2149 (July 2019), p. 20180220. DOI: 10.1098/rsta.2018.0220.

Publications linked to previous work

10. Guillaume Fraux and Jonathan P. K. Doye. “Note: Heterogeneous ice nucleation on silver-iodide-like surfaces”. In: *The Journal of Chemical Physics* 141.21 (Dec. 2014), p. 216101. DOI: 10.1063/1.4902382.

11. Giovanni Marco Nocera, Kalthoum Ben M’Barek, Dario Giovanni Bazzoli, Guillaume Fraux, Morgane Bontems-Van Heijenoort, Jeannette Chokki, Sonia Georgeault, Yong Chen, and Jacques Fattaccioli. “Fluorescent microparticles fabricated through chemical coating of O/W emulsion droplets with a thin metallic film”. In: *RSC Advances* 4.23 (2014), p. 11564. DOI: 10.1039/c3ra47063f.

12. Marie Basire, Félix Mouhat, Guillaume Fraux, Amélie Bordage, Jean-Louis Hazemann, Marion Louvel, Riccardo Spezia, Sara Bonella, and Rodolphe Vuilleumier. “Fermi resonance in CO2: Mode assignment and quantum nuclear effects from first principles molecular dynamics”. In: *The Journal of Chemical Physics* 146.13 (Apr. 2017), p. 134102. DOI: 10.1063/1.4979199.

BIBLIOGRAPHY

[1] O. K. Farha, I. Eryazici, N. C. Jeong, B. G. Hauser, C. E. Wilmer, A. A. Sarjeant, R. Q. Snurr, S. T. Nguyen, A. Ö. Yazaydın, and J. T. Hupp. "Metal–Organic Framework Materials with Ultrahigh Surface Areas: Is the Sky the Limit?" In: *Journal of the American Chemical Society* 134.36 (Aug. 2012), pp. 15016–15021. DOI: 10.1021/ja3055639.

[2] J. Rouquerol, D. Avnir, C. W. Fairbridge, D. H. Everett, J. M. Haynes, N. Pernicone, J. D. F. Ramsay, K. S. W. Sing, and K. K. Unger. "Recommendations for the characterization of porous solids (Technical Report)". In: *Pure and Applied Chemistry* 66.8 (Jan. 1994), pp. 1739–1758. DOI: 10.1351/pac199466081739.

[3] P. Levitz. "Slow dynamics in colloidal glasses and porous media as probed by NMR relaxometry: assessment of solvent levy statistics in the strong adsorption regime". In: *Magnetic Resonance Imaging* 21.3-4 (Apr. 2003), pp. 177–184. DOI: 10.1016/s0730-725x(03)00122-x.

[4] D. Das, D. P. Samal, and M. BC. "Preparation of Activated Carbon from Green Coconut Shell and its Characterization". In: *Journal of Chemical Engineering & Process Technology* 06.05 (2015). DOI: 10.4172/2157-7048.1000248.

[5] G. Férey. "Microporous Solids: From Organically Templated Inorganic Skeletons to Hybrid Frameworks ...Ecumenism in Chemistry". In: *Chemistry of Materials* 13.10 (Oct. 2001), pp. 3084–3098. DOI: 10.1021/cm011070n.

[6] I. Z. Association. *Database of Zeolite Structures*. Apr. 2019. URL: http://www.iza-structure.org/databases/.

[7] M. D. Foster and M. M. Treacy. *Atlas of Prospective Zeolite Structures*. Apr. 2019. URL: http://www.hypotheticalzeolites.net/.

[8] B. F. Abrahams, B. F. Hoskins, and R. Robson. "A new type of infinite 3D polymeric network containing 4-connected, peripherally-linked metalloporphyrin building blocks". In: *Journal of the American Chemical Society* 113.9 (Apr. 1991), pp. 3606–3607. DOI: 10.1021/ja00009a065.

[9] R. Robson. "Design and its limitations in the construction of bi- and poly-nuclear coordination complexes and coordination polymers (aka MOFs): a personal view". In: *Dalton Transactions* 38 (2008), p. 5113. DOI: 10.1039/b805617j.

[10] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi. "Design and synthesis of an exceptionally stable and highly porous metal–organic framework". In: *Nature* 402.6759 (Nov. 1999), pp. 276–279. DOI: 10.1038/46248.

[11] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi. "The Chemistry and Applications of Metal–Organic Frameworks". In: *Science* 341.6149 (Aug. 2013), pp. 1230444–1230444. DOI: 10.1126/science.1230444.

[12] S. Horike, S. Shimomura, and S. Kitagawa. “Soft porous crystals”. In: *Nature Chemistry* 1.9 (Nov. 2009), pp. 695–704. DOI: 10.1038/nchem.444.

[13] S. Kitagawa and K. Uemura. “Dynamic porous properties of coordination polymers inspired by hydrogen bonds”. In: *Chemical Society Reviews* 34.2 (2005), p. 109. DOI: 10.1039/b313997m.

[14] D. Bradshaw, J. B. Claridge, E. J. Cussen, T. J. Prior, and M. J. Rosseinsky. “Design, Chirality, and Flexibility in Nanoporous Molecule-Based Materials”. In: *ChemInform* 36.34 (Aug. 2005). DOI: 10.1002/chin.200534317.

[15] F.-X. Coudert. “Responsive Metal–Organic Frameworks and Framework Materials: Under Pressure, Taking the Heat, in the Spotlight, with Friends”. In: *Chemistry of Materials* 27.6 (Feb. 2015), pp. 1905–1916. DOI: 10.1021/acs.chemmater.5b00046.

[16] T. Tian, Z. Zeng, D. Vulpe, M. E. Casco, G. Divitini, P. A. Midgley, J. Silvestre-Albero, J.-C. Tan, P. Z. Moghadam, and D. Fairen-Jimenez. “A sol–gel monolithic metal–organic framework with enhanced methane uptake”. In: *Nature Materials* 17.2 (Dec. 2017), pp. 174–179. DOI: 10.1038/nmat5050.

[17] H. Oh, S. Maurer, R. Balderas-Xicohtencatl, L. Arnold, O. V. Magdysyuk, G. Schütz, U. Müller, and M. Hirscher. “Efficient synthesis for large-scale production and characterization for hydrogen storage of ligand exchanged MOF-74/174/184-M (M = Mg^{2+}, Ni^{2+})”. In: *International Journal of Hydrogen Energy* 42.2 (Jan. 2017), pp. 1027–1035. DOI: 10.1016/j.ijhydene.2016.08.153.

[18] J. L. C. Rowsell, A. R. Millward, K. S. Park, and O. M. Yaghi. “Hydrogen Sorption in Functionalized Metal–Organic Frameworks”. In: *Journal of the American Chemical Society* 126.18 (May 2004), pp. 5666–5667. DOI: 10.1021/ja049408c.

[19] K. S. Park, Z. Ni, A. P. Cote, J. Y. Choi, R. Huang, F. J. Uribe-Romo, H. K. Chae, M. O’Keeffe, and O. M. Yaghi. “Exceptional chemical and thermal stability of zeolitic imidazolate frameworks”. In: *Proceedings of the National Academy of Sciences* 103.27 (June 2006), pp. 10186–10191. DOI: 10.1073/pnas.0602439103.

[20] F.-X. Coudert, A. Boutin, M. Jeffroy, C. Mellot-Draznieks, and A. H. Fuchs. “Thermodynamic Methods and Models to Study Flexible Metal–Organic Frameworks”. In: *ChemPhysChem* 12.2 (Jan. 2011), pp. 247–258. DOI: 10.1002/cphc.201000590.

[21] D. Dubbeldam, K. S. Walton, D. E. Ellis, and R. Q. Snurr. “Exceptional Negative Thermal Expansion in Isoreticular Metal–Organic Frameworks”. In: *Angewandte Chemie International Edition* 46.24 (June 2007), pp. 4496–4499. DOI: 10.1002/anie.200700218.

[22] C. Serre, C. Mellot-Draznieks, S. Surble, N. Audebrand, Y. Filinchuk, and G. Ferey. “Role of Solvent-Host Interactions That Lead to Very Large Swelling of Hybrid Frameworks”. In: *Science* 315.5820 (Mar. 2007), pp. 1828–1831. DOI: 10.1126/science.1137975.

[23] C. Serre, F. Millange, C. Thouvenot, M. Noguès, G. Marsolier, D. Louër, and G. Férey. “Very Large Breathing Effect in the First Nanoporous Chromium(III)-Based Solids: MIL-53 or CrIII(OH)[O2C-C6H4-CO2][HO2C-C6H4-CO2H]x H2Oy”. In: *Journal of the American Chemical Society* 124.45 (Nov. 2002), pp. 13519–13526. DOI: 10.1021/ja0276974.

[24] S. U. Rege and R. T. Yang. “Limits for Air Separation by Adsorption with LiX Zeolite”. In: *Industrial & Engineering Chemistry Research* 36.12 (Dec. 1997), pp. 5358–5365. DOI: 10.1021/ie9705214.

[25] A. Primo and H. Garcia. “Zeolites as catalysts in oil refining”. In: *Chem. Soc. Rev.* 43.22 (Mar. 2014), pp. 7548–7561. DOI: 10.1039/c3cs60394f.

[26] L. Ćurković, S. Cerjan-Stefanović, and T. Filipan. “Metal ion exchange by natural and modified zeolites”. In: *Water Research* 31.6 (June 1997), pp. 1379–1382. DOI: 10.1016/s0043-1354(96)00411-3.

[27] E. Borai, R. Harjula, L. malinen, and A. Paajanen. “Efficient removal of cesium from low-level radioactive liquid waste using natural and impregnated zeolite minerals”. In: *Journal of Hazardous Materials* 172.1 (Dec. 2009), pp. 416–422. DOI: 10.1016/j.jhazmat.2009.07.033.

[28] E. M. Flanigen, J. M. Bennett, R. W. Grose, J. P. Cohen, R. L. Patton, R. M. Kirchner, and J. V. Smith. “Silicalite, a new hydrophobic crystalline silica molecular sieve”. In: *Nature* 271.5645 (Feb. 1978), pp. 512–516. DOI: 10.1038/271512a0.

[29] H. Karge and J. Weitkamp, eds. *Zeolites as Catalysts, Sorbents and Detergent Builders - Applications and Innovations, Proceedings of an International Symposium*. Elsevier, 1989. DOI: 10.1016/s0167-2991(08)x6020-3.

[30] D. S. Sholl and R. P. Lively. “Seven chemical separations to change the world”. In: *Nature* 532.7600 (Apr. 2016), pp. 435–437. DOI: 10.1038/532435a.

[31] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae, and J. R. Long. “Carbon Dioxide Capture in Metal–Organic Frameworks”. In: *Chemical Reviews* 112.2 (Dec. 2011), pp. 724–781. DOI: 10.1021/cr2003272.

[32] U. Mueller, M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndt, and J. Pastré. “Metal–organic frameworks: prospective industrial applications”. In: *J. Mater. Chem.* 16.7 (2006), pp. 626–636. DOI: 10.1039/b511962f.

[33] M. T. Kapelewski, T. Runcevski, J. D. Tarver, H. Z. H. Jiang, K. E. Hurst, P. A. Parilla, A. Ayala, T. Gennett, S. A. FitzGerald, C. M. Brown, and J. R. Long. “Record High Hydrogen Storage Capacity in the Metal–Organic Framework Ni2(m-dobdc) at Near-Ambient Temperatures”. In: *Chemistry of Materials* 30.22 (Nov. 2018), pp. 8179–8189. DOI: 10.1021/acs.chemmater.8b03276.

[34] O. Karagiaridi, M. B. Lalonde, W. Bury, A. A. Sarjeant, O. K. Farha, and J. T. Hupp. “Opening ZIF-8: A Catalytically Active Zeolitic Imidazolate Framework of Sodalite Topology with Unsubstituted Linkers”. In: *Journal of the American Chemical Society* 134.45 (Oct. 2012), pp. 18790–18796. DOI: 10.1021/ja308786r.

[35] N. Tshabang, G. P. Makgatle, S. A. Bourne, N. Kann, J. D. Evans, F.-X. Coudert, and L. Öhrström. “Conformational chiral polymorphism in cis-bis-triphenylphosphine complexes of transition metals”. In: *CrystEngComm* 20.35 (2018), pp. 5137–5142. DOI: 10.1039/c8ce00337h.

[36] P. Horcajada, S. Surblé, C. Serre, D.-Y. Hong, Y.-K. Seo, J.-S. Chang, J.-M. Grenèche, I. Margiolaki, and G. Férey. “Synthesis and catalytic properties of MIL-100(Fe), an iron(iii) carboxylate with large pores”. In: *Chem. Commun.* 27 (2007), pp. 2820–2822. DOI: 10.1039/b704325b.

[37] K. Schlichte, T. Kratzke, and S. Kaskel. “Improved synthesis, thermal stability and catalytic properties of the metal-organic framework compound Cu3(BTC)2”. In: *Microporous and Mesoporous Materials* 73.1-2 (Aug. 2004), pp. 81–88. DOI: 10.1016/j.micromeso.2003.12.027.

[38] J. Gascon, U. Aktay, M. D. Hernandez-Alonso, G. P. van Klink, and F. Kapteijn. “Amino-based metal-organic frameworks as stable, highly active basic catalysts”.

In: *Journal of Catalysis* 261.1 (Jan. 2009), pp. 75–87. DOI: 10.1016/j.jcat.2008.11.010.

[39] A. Lan, K. Li, H. Wu, D. H. Olson, T. J. Emge, W. Ki, M. Hong, and J. Li. "A Luminescent Microporous Metal-Organic Framework for the Fast and Reversible Detection of High Explosives". In: *Angewandte Chemie International Edition* 48.13 (Jan. 2009), pp. 2334–2338. DOI: 10.1002/anie.200804853.

[40] S. Pramanik, C. Zheng, X. Zhang, T. J. Emge, and J. Li. "New Microporous Metal-Organic Framework Demonstrating Unique Selectivity for Detection of High Explosives and Aromatic Compounds". In: *Journal of the American Chemical Society* 133.12 (Mar. 2011), pp. 4153–4155. DOI: 10.1021/ja106851d.

[41] D. M. Ruthven. *Principles of Adsorption and Adsorption Processes*. Elsevier, 1984. ISBN: 978-0-471-86606-0.

[42] R. T. Yang. *Gas Separation by Adsorption Processes*. Elsevier, 1987. DOI: 10.1016/c2013-0-04269-7.

[43] K. S. W. Sing. "Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (Recommendations 1984)". In: *Pure and Applied Chemistry* 57.4 (Jan. 1985), pp. 603–619. DOI: 10.1351/pac198557040603.

[44] J. Rouquerol, G. Baron, R. Denoyel, H. Giesche, J. Groen, P. Klobes, P. Levitz, A. V. Neimark, S. Rigby, R. Skudas, K. Sing, M. Thommes, and K. Unger. "Liquid intrusion and alternative methods for the characterization of macroporous materials (IUPAC Technical Report)". In: *Pure and Applied Chemistry* 84.1 (Dec. 2011), pp. 107–136. DOI: 10.1351/pac-rep-10-11-19.

[45] G. Fraux, F.-X. Coudert, A. Boutin, and A. H. Fuchs. "Forced intrusion of water and aqueous solutions in microporous materials: from fundamental thermodynamics to energy storage devices". In: *Chemical Society Reviews* 46.23 (2017), pp. 7421–7437. DOI: 10.1039/c7cs00478h.

[46] A. Y. Fadeev and V. A. Eroshenko. "Study of Penetration of Water into Hydrophobized Porous Silicas". In: *Journal of Colloid and Interface Science* 187.2 (Mar. 1997), pp. 275–282. DOI: 10.1006/jcis.1996.4495.

[47] V. Eroshenko, R.-C. Regis, M. Soulard, and J. Patarin. "Energetics: A New Field of Applications for Hydrophobic Zeolites". In: *Journal of the American Chemical Society* 123.33 (Aug. 2001), pp. 8129–8130. DOI: 10.1021/ja011011a.

[48] V. Eroshenko, R.-C. Regis, M. Soulard, and J. Patarin. "Les systèmes hétérogènes « eau–zéolithe hydrophobe » : de nouveaux ressorts moléculaires". In: *Comptes Rendus Physique* 3.1 (Jan. 2002), pp. 111–119. DOI: 10.1016/s1631-0705(02)01285-9.

[49] B. Lefevre, A. Saugey, J. L. Barrat, L. Bocquet, E. Charlaix, P. F. Gobin, and G. Vigier. "Intrusion and extrusion of water in hydrophobic mesopores". In: *The Journal of Chemical Physics* 120.10 (Mar. 2004), pp. 4927–4938. DOI: 10.1063/1.1643728.

[50] L. Liu, X. Chen, W. Lu, A. Han, and Y. Qiao. "Infiltration of Electrolytes in Molecular-Sized Nanopores". In: *Physical Review Letters* 102.18 (May 2009). DOI: 10.1103/physrevlett.102.184501.

[51] M. Michelin-Jamois, C. Picard, G. Vigier, and E. Charlaix. "Giant Osmotic Pressure in the Forced Wetting of Hydrophobic Nanopores". In: *Physical Review Letters* 115.3 (July 2015). DOI: 10.1103/physrevlett.115.036101.

[52] A. Ryzhikov, I. Khay, H. Nouali, T. J. Daou, and J. Patarin. "Drastic change of the intrusion-extrusion behavior of electrolyte solutions in pure silica ∗BEA-type zeolite". In: *Physical Chemistry Chemical Physics* 16.33 (2014), pp. 17893–17899. DOI: 10.1039/c4cp01862a.

[53] L. Tzanis, H. Nouali, T. J. Daou, M. Soulard, and J. Patarin. "Influence of the aqueous medium on the energetic performances of Silicalite-1". In: *Materials Letters* 115 (Jan. 2014), pp. 229–232. DOI: 10.1016/j.matlet.2013.10.063.

[54] I. Khay, T. J. Daou, H. Nouali, A. Ryzhikov, S. Rigolet, and J. Patarin. "High Pressure Intrusion–Extrusion of LiCl Aqueous Solutions in Silicalite-1 Zeolite: Influence on Energetic Performances". In: *The Journal of Physical Chemistry C* 118.8 (Feb. 2014), pp. 3935–3941. DOI: 10.1021/jp4105163.

[55] G. Ortiz, H. Nouali, C. Marichal, G. Chaplais, and J. Patarin. "Versatile Energetic Behavior of ZIF-8 upon High Pressure Intrusion–Extrusion of Aqueous Electrolyte Solutions". In: *The Journal of Physical Chemistry C* 118.14 (Mar. 2014), pp. 7321–7328. DOI: 10.1021/jp412354f.

[56] Z. Hu, Y. Chen, and J. Jiang. "Zeolitic imidazolate framework-8 as a reverse osmosis membrane for water desalination: Insight from molecular simulation". In: *The Journal of Chemical Physics* 134.13 (Apr. 2011), p. 134705. DOI: 10.1063/1.3573902.

[57] R. Arletti, L. Ronchi, S. Quartieri, G. Vezzalini, A. Ryzhikov, H. Nouali, T. J. Daou, and J. Patarin. "Intrusion-extrusion experiments of MgCl2 aqueous solution in pure silica ferrierite: Evidence of the nature of intruded liquid by in situ high pressure synchrotron X-ray powder diffraction". In: *Microporous and Mesoporous Materials* 235 (Nov. 2016), pp. 253–260. DOI: 10.1016/j.micromeso.2016.08.024.

[58] E. Braun, J. J. Chen, S. K. Schnell, L.-C. Lin, J. A. Reimer, and B. Smit. "Nanoporous Materials Can Tune the Critical Point of a Pure Substance". In: *Angewandte Chemie International Edition* 54.48 (Sept. 2015), pp. 14349–14352. DOI: 10.1002/anie.201506865.

[59] S. Krause, V. Bon, I. Senkovska, U. Stoeck, D. Wallacher, D. M. Többens, S. Zander, R. S. Pillai, G. Maurin, F.-X. Coudert, and S. Kaskel. "A pressure-amplifying framework material with negative gas adsorption transitions". In: *Nature* 532.7599 (Apr. 2016), pp. 348–352. DOI: 10.1038/nature17430.

[60] M. Mehta and D. A. Kofke. "Coexistence diagrams of mixtures by molecular simulation". In: *Chemical Engineering Science* 49.16 (Aug. 1994), pp. 2633–2645. DOI: 10.1016/0009-2509(94)e0078-5.

[61] F. A. Escobedo. "Novel pseudoensembles for simulation of multicomponent phase equilibria". In: *The Journal of Chemical Physics* 108.21 (June 1998), pp. 8761–8772. DOI: 10.1063/1.475396.

[62] A. L. Myers and J. M. Prausnitz. "Thermodynamics of mixed-gas adsorption". In: *AIChE Journal* 11.1 (Jan. 1965), pp. 121–127. DOI: 10.1002/aic.690110125.

[63] M. B. Sweatman and N. Quirke. "Predicting the Adsorption of Gas Mixtures: Adsorbed Solution Theory versus Classical Density Functional Theory". In: *Langmuir* 18.26 (Dec. 2002), pp. 10443–10454. DOI: 10.1021/la0200358.

[64] S. Suwanayuen and R. P. Danner. "A gas adsorption isotherm equation based on vacancy solution theory". In: *AIChE Journal* 26.1 (Jan. 1980), pp. 68–76. DOI: 10.1002/aic.690260112.

[65] O. Talu and I. Zwiebel. "Multicomponent adsorption equilibria of nonideal mixtures". In: *AIChE Journal* 32.8 (Aug. 1986), pp. 1263–1276. DOI: 10.1002/aic.690320805.

[66] R. Kitaura, K. Seki, G. Akiyama, and S. Kitagawa. "Porous Coordination Polymer Crystals with Gated Channels Specific for Supercritical Gases". In: *Angewandte Chemie International Edition* 42.4 (Jan. 2003), pp. 428–431. DOI: 10.1002/anie.200390130.

[67] D. Tanaka, K. Nakagawa, M. Higuchi, S. Horike, Y. Kubota, T. C. Kobayashi, M. Takata, and S. Kitagawa. "Kinetic Gate-Opening Process in a Flexible Porous Coordination Polymer". In: *Angewandte Chemie International Edition* 47.21 (May 2008), pp. 3914–3918. DOI: 10.1002/anie.200705822.

[68] L. Li, Y. Wang, J. Yang, X. Wang, and J. Li. "Targeted capture and pressure / temperature-responsive separation in flexible metal–organic frameworks". In: *Journal of Materials Chemistry A* 3.45 (2015), pp. 22574–22583. DOI: 10.1039/c5ta00679a.

[69] D. Banerjee, J. Liu, and P. K. Thallapally. "Separation of C2 Hydrocarbons by Porous Materials: Metal Organic Frameworks as Platform". In: *Comments on Inorganic Chemistry* 35.1 (Dec. 2014), pp. 18–38. DOI: 10.1080/02603594.2014.976704.

[70] S. Mukherjee, B. Joarder, A. V. Desai, B. Manna, R. Krishna, and S. K. Ghosh. "Exploiting Framework Flexibility of a Metal–Organic Framework for Selective Adsorption of Styrene over Ethylbenzene". In: *Inorganic Chemistry* 54.9 (Apr. 2015), pp. 4403–4408. DOI: 10.1021/acs.inorgchem.5b00206.

[71] M. L. Foo, R. Matsuda, Y. Hijikata, R. Krishna, H. Sato, S. Horike, A. Hori, J. Duan, Y. Sato, Y. Kubota, M. Takata, and S. Kitagawa. "An Adsorbate Discriminatory Gate Effect in a Flexible Porous Coordination Polymer for Selective Adsorption of CO2 over C2H2". In: *Journal of the American Chemical Society* 138.9 (Feb. 2016), pp. 3022–3030. DOI: 10.1021/jacs.5b10491.

[72] L. Li, R. Krishna, Y. Wang, J. Yang, X. Wang, and J. Li. "Exploiting the gate opening effect in a flexible MOF for selective adsorption of propyne from C1/C2/C3 hydrocarbons". In: *Journal of Materials Chemistry A* 4.3 (2016), pp. 751–755. DOI: 10.1039/c5ta09029f.

[73] C. Gücüyener, J. van den Bergh, J. Gascon, and F. Kapteijn. "Ethane / Ethene Separation Turned on Its Head: Selective Ethane Adsorption on the Metal–Organic Framework ZIF-7 through a Gate-Opening Mechanism". In: *Journal of the American Chemical Society* 132.50 (Dec. 2010), pp. 17704–17706. DOI: 10.1021/ja1089765.

[74] Y. Inubushi, S. Horike, T. Fukushima, G. Akiyama, R. Matsuda, and S. Kitagawa. "Modification of flexible part in $Cu2^+$ interdigitated framework for CH4/CO2 separation". In: *Chemical Communications* 46.48 (2010), p. 9229. DOI: 10.1039/c0cc01294g.

[75] N. Nijem, H. Wu, P. Canepa, A. Marti, K. J. Balkus, T. Thonhauser, J. Li, and Y. J. Chabal. "Tuning the Gate Opening Pressure of Metal–Organic Frameworks (MOFs) for the Selective Separation of Hydrocarbons". In: *Journal of the American Chemical Society* 134.37 (Sept. 2012), pp. 15201–15204. DOI: 10.1021/ja305754f.

[76] S. Sanda, S. Parshamoni, and S. Konar. "Third-Generation Breathing Metal–Organic Framework with Selective, Stepwise, Reversible, and Hysteretic Adsorp-

tion Properties". In: *Inorganic Chemistry* 52.22 (Oct. 2013), pp. 12866–12868. DOI: 10.1021/ic402095u.

[77] B. Joarder, S. Mukherjee, A. K. Chaudhari, A. V. Desai, B. Manna, and S. K. Ghosh. "Guest-Responsive Function of a Dynamic Metal–Organic Framework with a π Lewis Acidic Pore Surface". In: *Chemistry - A European Journal* 20.47 (Oct. 2014), pp. 15303–15308. DOI: 10.1002/chem.201402855.

[78] S. Mukherjee, B. Joarder, B. Manna, A. V. Desai, A. K. Chaudhari, and S. K. Ghosh. "Framework-Flexibility Driven Selective Sorption of p-Xylene over Other Isomers by a Dynamic Metal–Organic Framework". In: *Scientific Reports* 4.1 (July 2014). DOI: 10.1038/srep05761.

[79] F.-X. Coudert, C. Mellot-Draznieks, A. H. Fuchs, and A. Boutin. "Prediction of Breathing and Gate-Opening Transitions Upon Binary Mixture Adsorption in Metal–Organic Frameworks". In: *Journal of the American Chemical Society* 131.32 (Aug. 2009), pp. 11329–11331. DOI: 10.1021/ja904123f.

[80] F.-X. Coudert. "The osmotic framework adsorbed solution theory: predicting mixture coadsorption in flexible nanoporous materials". In: *Physical Chemistry Chemical Physics* 12.36 (2010), p. 10904. DOI: 10.1039/c003434g.

[81] A. U. Ortiz, M.-A. Springuel-Huet, F.-X. Coudert, A. H. Fuchs, and A. Boutin. "Predicting Mixture Coadsorption in Soft Porous Crystals: Experimental and Theoretical Study of CO2/CH4 in MIL-53(Al)". In: *Langmuir* 28.1 (Dec. 2011), pp. 494–498. DOI: 10.1021/la203925y.

[82] G. Fraux, A. Boutin, A. H. Fuchs, and F.-X. Coudert. "On the use of the IAST method for gas separation studies in porous materials with gate-opening behavior". In: *Adsorption* 24.3 (Mar. 2018), pp. 233–241. DOI: 10.1007/s10450-018-9942-5.

[83] A. L. Myers and P. A. Monson. "Physical adsorption of gases: the case for absolute adsorption as the basis for thermodynamic analysis". In: *Adsorption* 20.4 (Mar. 2014), pp. 591–622. DOI: 10.1007/s10450-014-9604-1.

[84] S. Brandani, E. Mangano, and L. Sarkisov. "Net, excess and absolute adsorption and adsorption of helium". In: *Adsorption* 22.2 (Feb. 2016), pp. 261–276. DOI: 10.1007/s10450-016-9766-0.

[85] F.-X. Coudert, M. Jeffroy, A. H. Fuchs, A. Boutin, and C. Mellot-Draznieks. "Thermodynamics of Guest-Induced Structural Transitions in Hybrid Organic-Inorganic Frameworks". In: *Journal of the American Chemical Society* 130.43 (Oct. 2008), pp. 14294–14302. DOI: 10.1021/ja805129c.

[86] H. C. Hoffmann, B. Assfour, F. Epperlein, N. Klein, S. Paasch, I. Senkovska, S. Kaskel, G. Seifert, and E. Brunner. "High-Pressure in Situ129Xe NMR Spectroscopy and Computer Simulations of Breathing Transitions in the Metal–Organic Framework Ni2(2, 6-ndc)2(dabco) (DUT-8(Ni))". In: *Journal of the American Chemical Society* 133.22 (June 2011), pp. 8681–8690. DOI: 10.1021/ja201951t.

[87] J. Zang, S. Nair, and D. S. Sholl. "Osmotic ensemble methods for predicting adsorption-induced structural transitions in nanoporous materials using molecular simulations". In: *The Journal of Chemical Physics* 134.18 (May 2011), p. 184103. DOI: 10.1063/1.3586807.

[88] G. Fraux, F.-X. Coudert, et al. *Citable data produced by the Coudert research group.* URL: https://github.com/fxcoudert/citable-data.

[89] C. M. Simon, B. Smit, and M. Haranczyk. "pyIAST: Ideal adsorbed solution theory (IAST) Python package". In: *Computer Physics Communications* 200 (Mar. 2016), pp. 364–380. DOI: 10.1016/j.cpc.2015.11.016.

[90] A. Lan, K. Li, H. Wu, L. Kong, N. Nijem, D. H. Olson, T. J. Emge, Y. J. Chabal, D. C. Langreth, M. Hong, and J. Li. "RPM3: A Multifunctional Microporous MOF with Recyclable Framework and High H_2 Binding Energy". In: *Inorganic Chemistry* 48.15 (Aug. 2009), pp. 7165–7173. DOI: 10.1021/ic9002115.

[91] M. E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation (Oxford Graduate Texts)*. Oxford University Press, 2010. ISBN: 978-0-198-52526-4.

[92] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier, 2002. DOI: 10.1016/b978-0-12-267351-1.x5000-7.

[93] F. London. "Zur Theorie und Systematik der Molekularkräfte". In: *Zeitschrift für Physik* 63.3-4 (Mar. 1930), pp. 245–279. DOI: 10.1007/bf01421741.

[94] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (June 1953), pp. 1087–1092. DOI: 10.1063/1.1699114.

[95] D. A. Kofke and E. D. Glandt. "Monte Carlo simulation of multicomponent equilibria in a semigrand canonical ensemble". In: *Molecular Physics* 64.6 (Aug. 1988), pp. 1105–1131. DOI: 10.1080/00268978800100743.

[96] I. R. Craig and D. E. Manolopoulos. "Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics". In: *The Journal of Chemical Physics* 121.8 (Aug. 2004), pp. 3368–3373. DOI: 10.1063/1.1777575.

[97] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reversible multiple time scale molecular dynamics". In: *The Journal of Chemical Physics* 97.3 (Aug. 1992), pp. 1990–2001. DOI: 10.1063/1.463137.

[98] L. Verlet. "Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Physical Review* 159.1 (July 1967), pp. 98–103. DOI: 10.1103/physrev.159.98.

[99] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". In: *The Journal of Chemical Physics* 81.8 (Oct. 1984), pp. 3684–3690. DOI: 10.1063/1.448118.

[100] E. Braun, S. M. Moosavi, and B. Smit. "Anomalous Effects of Velocity Rescaling Algorithms: The Flying Ice Cube Effect Revisited". In: *Journal of Chemical Theory and Computation* 14.10 (Aug. 2018), pp. 5262–5272. DOI: 10.1021/acs.jctc.8b00446.

[101] S. Nosé. "A unified formulation of the constant temperature molecular dynamics methods". In: *The Journal of Chemical Physics* 81.1 (July 1984), pp. 511–519. DOI: 10.1063/1.447334.

[102] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". In: *Physical Review A* 31.3 (Mar. 1985), pp. 1695–1697. DOI: 10.1103/physreva.31.1695.

[103] G. Bussi, D. Donadio, and M. Parrinello. "Canonical sampling through velocity rescaling". In: *The Journal of Chemical Physics* 126.1 (Jan. 2007), p. 014101. DOI: 10.1063/1.2408420.

[104] B. Hess. "Stochastic Concepts in Molecular Simulation". PhD thesis. University of Groningen, 2002.

[105] T. Çağin and B. M. Pettitt. "Molecular dynamics with a variable number of molecules". In: *Molecular Physics* 72.1 (Jan. 1991), pp. 169–175. DOI: 10.1080/00268979100100111.

[106] C. Lo and B. Palmer. "Alternative Hamiltonian for molecular dynamics simulations in the grand canonical ensemble". In: *The Journal of Chemical Physics* 102.2 (Jan. 1995), pp. 925–931. DOI: 10.1063/1.469159.

[107] H. Eslami and F. Müller-Plathe. "Molecular dynamics simulation in the grand canonical ensemble". In: *Journal of Computational Chemistry* 28.10 (2007), pp. 1763–1773. DOI: 10.1002/jcc.20689.

[108] G. S. Heffelfinger and F. van Swol. "Diffusion in Lennard-Jones fluids using dual control volume grand canonical molecular dynamics simulation (DCV-GCMD)". In: *The Journal of Chemical Physics* 100.10 (May 1994), pp. 7548–7552. DOI: 10.1063/1.466849.

[109] R. F. Cracknell, D. Nicholson, and N. Quirke. "Direct Molecular Dynamics Simulation of Flow Down a Chemical Potential Gradient in a Slit-Shaped Micropore". In: *Physical Review Letters* 74.13 (Mar. 1995), pp. 2463–2466. DOI: 10.1103/physrevlett.74.2463.

[110] S. Boinepalli and P. Attard. "Grand canonical molecular dynamics". In: *The Journal of Chemical Physics* 119.24 (Dec. 2003), pp. 12769–12775. DOI: 10.1063/1.1629079.

[111] L. D. Site. "Formulation of Liouville's theorem for grand ensemble molecular simulations". In: *Physical Review E* 93.2 (Feb. 2016). DOI: 10.1103/physreve.93.022130.

[112] A. Laio and M. Parrinello. "Escaping free-energy minima". In: *Proceedings of the National Academy of Sciences* 99.20 (Sept. 2002), pp. 12562–12566. DOI: 10.1073/pnas.202427399.

[113] G. Ciccotti and M. Ferrario. "Blue Moon Approach to Rare Events". In: *Molecular Simulation* 30.11-12 (Sept. 2004), pp. 787–793. DOI: 10.1080/0892702042000270214.

[114] E. Darve, D. Rodríguez-Gómez, and A. Pohorille. "Adaptive biasing force method for scalar and vector free energy calculations". In: *The Journal of Chemical Physics* 128.14 (Apr. 2008), p. 144120. DOI: 10.1063/1.2829861.

[115] F. Wang and D. P. Landau. "Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States". In: *Physical Review Letters* 86.10 (Mar. 2001), pp. 2050–2053. DOI: 10.1103/physrevlett.86.2050.

[116] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. "Multidimensional free-energy calculations using the weighted histogram analysis method". In: *Journal of Computational Chemistry* 16.11 (Nov. 1995), pp. 1339–1350. DOI: 10.1002/jcc.540161104.

[117] E. Schrödinger. "An Undulatory Theory of the Mechanics of Atoms and Molecules". In: *Physical Review* 28.6 (Dec. 1926), pp. 1049–1070. DOI: 10.1103/physrev.28.1049.

[118] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 389.20 (1927), pp. 457–484. DOI: 10.1002/andp.19273892002.

[119] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical Review* 136.3B (Nov. 1964), B864–B871. DOI: 10.1103/physrev.136.b864.

[120] W. Kohn and L. J. Sham. "Self-Consistent Equations Including Exchange and Correlation Effects". In: *Physical Review* 140.4A (Nov. 1965), A1133–A1138. DOI: 10.1103/physrev.140.a1133.

[121] J. P. Perdew. "Jacob's ladder of density functional approximations for the exchange-correlation energy". In: *AIP Conference Proceedings*. AIP, 2001. DOI: 10.1063/1.1390175.

[122] S. Grimme. "Semiempirical GGA-type density functional constructed with a long-range dispersion correction". In: *Journal of Computational Chemistry* 27.15 (2006), pp. 1787–1799. DOI: 10.1002/jcc.20495.

[123] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu". In: *The Journal of Chemical Physics* 132.15 (Apr. 2010), p. 154104. DOI: 10.1063/1.3382344.

[124] G. Chaplais, G. Fraux, J.-L. Paillaud, C. Marichal, H. Nouali, A. H. Fuchs, F.-X. Coudert, and J. Patarin. "Impacts of the Imidazolate Linker Substitution (CH3, Cl, or Br) on the Structural and Adsorptive Properties of ZIF-8". In: *The Journal of Physical Chemistry C* 122.47 (Oct. 2018), pp. 26945–26955. DOI: 10.1021/acs.jpcc.8b08706.

[125] K. Li, D. H. Olson, J. Seidel, T. J. Emge, H. Gong, H. Zeng, and J. Li. "Zeolitic Imidazolate Frameworks for Kinetic Separation of Propane and Propene". In: *Journal of the American Chemical Society* 131.30 (Aug. 2009), pp. 10368–10369. DOI: 10.1021/ja9039983.

[126] H. Bux, A. Feldhoff, J. Cravillon, M. Wiebcke, Y.-S. Li, and J. Caro. "Oriented Zeolitic Imidazolate Framework-8 Membrane with Sharp H2/C3H8Molecular Sieve Separation". In: *Chemistry of Materials* 23.8 (Apr. 2011), pp. 2262–2269. DOI: 10.1021/cm200555s.

[127] R. J. Verploegh, S. Nair, and D. S. Sholl. "Temperature and Loading-Dependent Diffusion of Light Hydrocarbons in ZIF-8 as Predicted Through Fully Flexible Molecular Simulations". In: *Journal of the American Chemical Society* 137.50 (Dec. 2015), pp. 15760–15771. DOI: 10.1021/jacs.5b08746.

[128] S. A. Moggach, T. D. Bennett, and A. K. Cheetham. "The Effect of Pressure on ZIF-8: Increasing Pore Size with Pressure and the Formation of a High-Pressure Phase at 1.47 GPa". In: *Angewandte Chemie International Edition* 48.38 (Sept. 2009), pp. 7087–7089. DOI: 10.1002/anie.200902643.

[129] D. Fairen-Jimenez, S. A. Moggach, M. T. Wharmby, P. A. Wright, S. Parsons, and T. Düren. "Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations". In: *Journal of the American Chemical Society* 133.23 (June 2011), pp. 8900–8902. DOI: 10.1021/ja202154j.

[130] F.-X. Coudert. "Molecular Mechanism of Swing Effect in Zeolitic Imidazolate Framework ZIF-8: Continuous Deformation upon Adsorption". In: *ChemPhysChem* 18.19 (Aug. 2017), pp. 2732–2738. DOI: 10.1002/cphc.201700463.

[131] D. Fairen-Jimenez, R. Galvelis, A. Torrisi, A. D. Gellan, M. T. Wharmby, P. A. Wright, C. Mellot-Draznieks, and T. Düren. "Flexibility and swing effect on the adsorption of energy-related gases on ZIF-8: combined experimental and simulation study". In: *Dalton Transactions* 41.35 (2012), p. 10752. DOI: 10.1039/c2dt30774j.

[132] C. O. Ania, E. García-Pérez, M. Haro, J. J. Gutiérrez-Sevillano, T. Valdés-Solís, J. B. Parra, and S. Calero. "Understanding Gas-Induced Structural Deformation of ZIF-8". In: *The Journal of Physical Chemistry Letters* 3.9 (Apr. 2012), pp. 1159–1164. DOI: 10.1021/jz300292y.

[133] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. "PACKMOL: A package for building initial configurations for molecular dynamics simulations". In: *Journal of Computational Chemistry* 30.13 (Oct. 2009), pp. 2157–2164. DOI: 10.1002/jcc.21224.

[134] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach". In: *Computer Physics Communications* 167.2 (Apr. 2005), pp. 103–128. DOI: 10.1016/j.cpc.2004.12.014.

[135] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk. "Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials". In: *Microporous and Mesoporous Materials* 149.1 (Feb. 2012), pp. 134–141. DOI: 10.1016/j.micromeso.2011.08.020.

[136] J. P. Dürholt, G. Fraux, F.-X. Coudert, and R. Schmid. "Ab Initio Derived Force Fields for Zeolitic Imidazolate Frameworks: MOF-FF for ZIFs". In: *Journal of Chemical Theory and Computation* 15.4 (Mar. 2019), pp. 2420–2432. DOI: 10.1021/acs.jctc.8b01041.

[137] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff. "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations". In: *Journal of the American Chemical Society* 114.25 (Dec. 1992), pp. 10024–10035. DOI: 10.1021/ja00051a040.

[138] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. "Development and testing of a general amber force field". In: *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174. DOI: 10.1002/jcc.20035.

[139] M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine. "Extension of the Universal Force Field to Metal–Organic Frameworks". In: *Journal of Chemical Theory and Computation* 10.2 (Jan. 2014), pp. 880–891. DOI: 10.1021/ct400952t.

[140] D. E. Coupry, M. A. Addicoat, and T. Heine. "Extension of the Universal Force Field for Metal–Organic Frameworks". In: *Journal of Chemical Theory and Computation* 12.10 (Sept. 2016), pp. 5215–5225. DOI: 10.1021/acs.jctc.6b00664.

[141] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. V. Speybroeck. "QuickFF: A program for a quick and easy derivation of force fields for metal–organic frameworks fromab initioinput". In: *Journal of Computational Chemistry* 36.13 (Mar. 2015), pp. 1015–1027. DOI: 10.1002/jcc.23877.

[142] S. Bureekaew, S. Amirjalayer, M. Tafipolsky, C. Spickermann, T. K. Roy, and R. Schmid. "MOF-FF - A flexible first-principles derived force field for metal–organic frameworks". In: *physica status solidi (b)* 250.6 (Mar. 2013), pp. 1128–1141. DOI: 10.1002/pssb.201248460.

[143] N. L. Allinger, Y. H. Yuh, and J. H. Lii. "Molecular mechanics. The MM3 force field for hydrocarbons. 1". In: *Journal of the American Chemical Society* 111.23 (Nov. 1989), pp. 8551–8566. DOI: 10.1021/ja00205a001.

[144] N. L. Allinger, X. Zhou, and J. Bergsma. "Molecular mechanics parameters". In: *Journal of Molecular Structure: THEOCHEM* 312.1 (Jan. 1994), pp. 69–83. DOI: 10.1016/s0166-1280(09)80008-0.

[145] J.-C. Tan, B. Civalleri, C.-C. Lin, L. Valenzano, R. Galvelis, P.-F. Chen, T. D. Bennett, C. Mellot-Draznieks, C. M. Zicovich-Wilson, and A. K. Cheetham. "Exceptionally Low Shear Modulus in a Prototypical Imidazole-Based Metal–Organic Frame-

work". In: *Physical Review Letters* 108.9 (Feb. 2012). DOI: 10.1103/physrevlett.108.095502.

[146] B. Zheng, Y. Zhu, F. Fu, L. L. Wang, J. Wang, and H. Du. "Theoretical prediction of the mechanical properties of zeolitic imidazolate frameworks (ZIFs)". In: *RSC Advances* 7.66 (2017), pp. 41499–41503. DOI: 10.1039/c7ra07242b.

[147] M. Soulard, J. Patarin, V. Eroshenko, and R. Regis. "Molecular spring or bumper: A new application for hydrophobic zeolitic materials". In: *Recent Advances in the Science and Technology of Zeolites and Related Materials Part B, Proceedings of the 14th International Zeolite Conference*. Elsevier, 2004, pp. 1830–1837. DOI: 10.1016/s0167-2991(04)80716-x.

[148] G. Fraux and F.-X. Coudert. "Recent advances in the computational chemistry of soft porous crystals". In: *Chemical Communications* 53.53 (2017), pp. 7211–7221. DOI: 10.1039/c7cc03306k.

[149] M. A. Saada, M. Soulard, B. Marler, H. Gies, and J. Patarin. "High-Pressure Water Intrusion Investigation of Pure Silica RUB-41 and S-SOD Zeolite Materials". In: *The Journal of Physical Chemistry C* 115.2 (Dec. 2010), pp. 425–430. DOI: 10.1021/jp109064e.

[150] N. Desbiens, I. Demachy, A. H. Fuchs, H. Kirsch-Rodeschini, M. Soulard, and J. Patarin. "Water Condensation in Hydrophobic Nanopores". In: *Angewandte Chemie* 117.33 (Aug. 2005), pp. 5444–5447. DOI: 10.1002/ange.200501250.

[151] T. Humplik, R. Raj, S. C. Maroo, T. Laoui, and E. N. Wang. "Effect of Hydrophilic Defects on Water Transport in MFI Zeolites". In: *Langmuir* 30.22 (May 2014), pp. 6446–6453. DOI: 10.1021/la500939t.

[152] T. Humplik, R. Raj, S. C. Maroo, T. Laoui, and E. N. Wang. "Framework water capacity and infiltration pressure of MFI zeolites". In: *Microporous and Mesoporous Materials* 190 (May 2014), pp. 84–91. DOI: 10.1016/j.micromeso.2014.01.026.

[153] G. Ortiz, H. Nouali, C. Marichal, G. Chaplais, and J. Patarin. "Energetic performances of the metal–organic framework ZIF-8 obtained using high pressure water intrusion–extrusion experiments". In: *Physical Chemistry Chemical Physics* 15.14 (2013), p. 4888. DOI: 10.1039/c3cp00142c.

[154] Y. Grosu, S. Gomes, G. Renaudin, J.-P. E. Grolier, V. Eroshenko, and J.-M. Nedelec. "Stability of zeolitic imidazolate frameworks: effect of forced water intrusion and framework flexibility dynamics". In: *RSC Advances* 5.109 (2015), pp. 89498–89502. DOI: 10.1039/c5ra19879h.

[155] A. U. Ortiz, A. P. Freitas, A. Boutin, A. H. Fuchs, and F.-X. Coudert. "What makes zeolitic imidazolate frameworks hydrophobic or hydrophilic? The impact of geometry and functionalization on water adsorption". In: *Physical Chemistry Chemical Physics* 16.21 (2014), pp. 9940–9949. DOI: 10.1039/c3cp54292k.

[156] M. A. Camblor, A. Corma, and S. Valencia. "Spontaneous nucleation and growth of pure silica zeolite-β free of connectivity defects". In: *Chemical Communications* 20 (1996), p. 2365. DOI: 10.1039/cc9960002365.

[157] Y. Lee, J. A. Hriljac, T. Vogt, J. B. Parise, and G. Artioli. "First Structural Investigation of a Super-Hydrated Zeolite". In: *Journal of the American Chemical Society* 123.50 (Dec. 2001), pp. 12732–12733. DOI: 10.1021/ja017098h.

[158] D. Seoung, Y. Lee, C.-C. Kao, T. Vogt, and Y. Lee. "Super-Hydrated Zeolites: Pressure-Induced Hydration in Natrolites". In: *Chemistry - A European Journal* 19.33 (July 2013), pp. 10876–10883. DOI: 10.1002/chem.201300591.

[159] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer. "Flexible metal–organic frameworks". In: *Chem. Soc. Rev.* 43.16 (May 2014), pp. 6062–6096. DOI: 10.1039/c4cs00101j.

[160] V. Haigis, F.-X. Coudert, R. Vuilleumier, and A. Boutin. "Investigation of structure and dynamics of the hydrated metal–organic framework MIL-53(Cr) using first-principles molecular dynamics". In: *Physical Chemistry Chemical Physics* 15.43 (2013), p. 19049. DOI: 10.1039/c3cp53126k.

[161] A. Boutin, D. Bousquet, A. U. Ortiz, F.-X. Coudert, A. H. Fuchs, A. Ballandras, G. Weber, I. Bezverkhyy, J.-P. Bellat, G. Ortiz, G. Chaplais, J.-L. Paillaud, C. Marichal, H. Nouali, and J. Patarin. "Temperature-Induced Structural Transitions in the Gallium-Based MIL-53 Metal–Organic Framework". In: *The Journal of Physical Chemistry C* 117.16 (Apr. 2013), pp. 8180–8188. DOI: 10.1021/jp312179e.

[162] F.-X. Coudert, A. Boutin, and A. H. Fuchs. "A thermodynamic description of the adsorption-induced structural transitions in flexible MIL-53 Metal–Organic framework". In: *Molecular Physics* 112.9-10 (Feb. 2014), pp. 1257–1261. DOI: 10.1080/00268976.2014.889325.

[163] B. Mortada, G. Chaplais, V. Veremeienko, H. Nouali, C. Marichal, and J. Patarin. "Energetic Performances of ZIF-8 Derivatives: Impact of the Substitution (Me, Cl, or Br) on Imidazolate Linker". In: *The Journal of Physical Chemistry C* 122.7 (Feb. 2018), pp. 3846–3855. DOI: 10.1021/acs.jpcc.7b08999.

[164] G. Fraux, A. Boutin, A. H. Fuchs, and F.-X. Coudert. "Structure, Dynamics, and Thermodynamics of Intruded Electrolytes in ZIF-8". In: *The Journal of Physical Chemistry C* (June 2019). DOI: 10.1021/acs.jpcc.9b02718.

[165] S. Plimpton. *Fast parallel algorithms for short-range molecular dynamics*. Tech. rep. May 1993. DOI: 10.2172/10176421.

[166] G. Fiorin, M. L. Klein, and J. Hénin. "Using collective variables to drive molecular dynamics simulations". In: *Molecular Physics* 111.22-23 (Dec. 2013), pp. 3345–3362. DOI: 10.1080/00268976.2013.813594.

[167] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The missing term in effective pair potentials". In: *The Journal of Physical Chemistry* 91.24 (Nov. 1987), pp. 6269–6271. DOI: 10.1021/j100308a038.

[168] B. Zheng, M. Sant, P. Demontis, and G. B. Suffritti. "Force Field for Molecular Dynamics Computations in Flexible ZIF-8 Framework". In: *The Journal of Physical Chemistry C* 116.1 (Jan. 2012), pp. 933–938. DOI: 10.1021/jp209463a.

[169] S. Chowdhuri and A. Chandra. "Hydration structure and diffusion of ions in supercooled water: Ion size effects". In: *The Journal of Chemical Physics* 118.21 (June 2003), pp. 9719–9725. DOI: 10.1063/1.1570405.

[170] Y. Marcus. "Ionic radii in aqueous solutions". In: *Chemical Reviews* 88.8 (Dec. 1988), pp. 1475–1498. DOI: 10.1021/cr00090a003.

[171] A. Luzar and D. Chandler. "Hydrogen-bond kinetics in liquid water". In: *Nature* 379.6560 (Jan. 1996), pp. 55–57. DOI: 10.1038/379055a0.

[172] A. C. Fogarty, F.-X. Coudert, A. Boutin, and D. Laage. "Reorientational Dynamics of Water Confined in Zeolites". In: *ChemPhysChem* 15.3 (Jan. 2014), pp. 521–529. DOI: 10.1002/cphc.201300928.

[173] L. F. Scatena. "Water at Hydrophobic Surfaces: Weak Hydrogen Bonding and Strong Orientation Effects". In: *Science* 292.5518 (May 2001), pp. 908–912. DOI: 10.1126/science.1059514.

[174] S. Jeffery, P. M. Hoffmann, J. B. Pethica, C. Ramanujan, H. Ö. Özer, and A. Oral. "Direct measurement of molecular stiffness and damping in confined water layers". In: *Physical Review B* 70.5 (Aug. 2004). DOI: 10.1103/physrevb.70.054114.

[175] S. R.-V. Castrillón, N. Giovambattista, I. A. Aksay, and P. G. Debenedetti. "Evolution from Surface-Influenced to Bulk-Like Dynamics in Nanoscopically Confined Water". In: *The Journal of Physical Chemistry B* 113.23 (June 2009), pp. 7973–7976. DOI: 10.1021/jp9025392.

[176] L. Scalfi, G. Fraux, A. Boutin, and F.-X. Coudert. "Structure and Dynamics of Water Confined in Imogolite Nanotubes". In: *Langmuir* 34.23 (May 2018), pp. 6748–6756. DOI: 10.1021/acs.langmuir.8b01115.

[177] T. D. Bennett, P. Simoncic, S. A. Moggach, F. Gozzo, P. Macchi, D. A. Keen, J.-C. Tan, and A. K. Cheetham. "Reversible pressure-induced amorphization of a zeolitic imidazolate framework (ZIF-4)". In: *Chemical Communications* 47.28 (2011), p. 7983. DOI: 10.1039/c1cc11985k.

[178] S. Cao, T. D. Bennett, D. A. Keen, A. L. Goodwin, and A. K. Cheetham. "Amorphization of the prototypical zeolitic imidazolate framework ZIF-8 by ball-milling". In: *Chemical Communications* 48.63 (2012), p. 7805. DOI: 10.1039/c2cc33773h.

[179] A. U. Ortiz, A. Boutin, A. H. Fuchs, and F.-X. Coudert. "Investigating the Pressure-Induced Amorphization of Zeolitic Imidazolate Framework ZIF-8: Mechanical Instability Due to Shear Mode Softening". In: *The Journal of Physical Chemistry Letters* 4.11 (May 2013), pp. 1861–1865. DOI: 10.1021/jz400880p.

[180] E. H. Lanman and B. J. Mair. "The Compressibility of Aqueous Solutions". In: *Journal of the American Chemical Society* 56.2 (Feb. 1934), pp. 390–393. DOI: 10.1021/ja01317a033.

[181] T. Karbowiak, C. Paulin, A. Ballandras, G. Weber, and J.-P. Bellat. "Thermal Effects of Water Intrusion in Hydrophobic Nanoporous Materials". In: *Journal of the American Chemical Society* 131.29 (July 2009), pp. 9898–9899. DOI: 10.1021/ja903954h.

[182] T. Karbowiak, C. Paulin, and J.-P. Bellat. "Determination of water intrusion heat in hydrophobic microporous materials by high pressure calorimetry". In: *Microporous and Mesoporous Materials* 134.1-3 (Oct. 2010), pp. 8–15. DOI: 10.1016/j.micromeso.2010.05.001.

[183] A. Grossfield. *WHAM: the weighted histogram analysis method, version 2*. URL: http://membrane.urmc.rochester.edu/content/wham.

[184] N. Yoshinaga and S. Aomine. "Imogolite in some ando soils". In: *Soil Science and Plant Nutrition* 8.3 (May 1962), pp. 22–29. DOI: 10.1080/00380768.1962.10430993.

[185] M.-S. Amara, E. Paineau, M. Bacia-Verloop, M.-E. M. Krapf, P. Davidson, L. Belloni, C. Levard, J. Rose, P. Launois, and A. Thill. "Single-step formation of micron long (OH)3Al2O3Ge(OH) imogolite-like nanotubes". In: *Chemical Communications* 49.96 (2013), p. 11284. DOI: 10.1039/c3cc46839a.

[186] P. D. G. Cradwick, V. C. Farmer, J. D. Russell, C. R. Masson, K. Wada, and N. Yoshinaga. "Imogolite, a Hydrated Aluminium Silicate of Tubular Structure". In: *Nature Physical Science* 240.104 (Dec. 1972), pp. 187–189. DOI: 10.1038/physci240187a0.

[187] S. U. Lee, Y. C. Choi, S. G. Youm, and D. Sohn. "Origin of the Strain Energy Minimum in Imogolite Nanotubes". In: *The Journal of Physical Chemistry C* 115.13 (Mar. 2011), pp. 5226–5231. DOI: 10.1021/jp108629z.

[188] R. I. González, R. Ramírez, J. Rogan, J. A. Valdivia, F. Munoz, F. Valencia, M. Ramírez, and M. Kiwi. "Model for Self-Rolling of an Aluminosilicate Sheet into a Single-Walled Imogolite Nanotube". In: *The Journal of Physical Chemistry C* 118.48 (Nov. 2014), pp. 28227–28233. DOI: 10.1021/jp508637q.

[189] K. Tamura and K. Kawamura. "Molecular Dynamics Modeling of Tubular Aluminum Silicate: Imogolite". In: *The Journal of Physical Chemistry B* 106.2 (Jan. 2002), pp. 271–278. DOI: 10.1021/jp0124793.

[190] L. Guimarães, A. N. Enyashin, J. Frenzel, T. Heine, H. A. Duarte, and G. Seifert. "Imogolite Nanotubes: Stability, Electronic, and Mechanical Properties". In: *ACS Nano* 1.4 (Nov. 2007), pp. 362–368. DOI: 10.1021/nn700184k.

[191] M. Zhao, Y. Xia, and L. Mei. "Energetic Minimum Structures of Imogolite Nanotubes: A First-Principles Prediction". In: *The Journal of Physical Chemistry C* 113.33 (July 2009), pp. 14834–14837. DOI: 10.1021/jp9056169.

[192] R. Demichelis, Y. Noël, P. D'Arco, L. Maschio, R. Orlando, and R. Dovesi. "Structure and energetics of imogolite: a quantum mechanical ab initio study with B3LYP hybrid functional". In: *Journal of Materials Chemistry* 20.46 (2010), p. 10417. DOI: 10.1039/c0jm00771d.

[193] F. Alvarez-Ramírez. "Ab initiosimulation of the structural and electronic properties of aluminosilicate and aluminogermanate natotubes with imogolite-like structure". In: *Physical Review B* 76.12 (Sept. 2007). DOI: 10.1103/physrevb.76.125421.

[194] S. Konduri, H. M. Tong, S. Chempath, and S. Nair. "Water in Single-Walled Aluminosilicate Nanotubes: Diffusion and Adsorption Properties". In: *The Journal of Physical Chemistry C* 112.39 (Sept. 2008), pp. 15367–15374. DOI: 10.1021/jp8025144.

[195] R. I. González, J. Rogan, E. M. Bringa, and J. A. Valdivia. "Mechanical Response of Aluminosilicate Nanotubes under Compression". In: *The Journal of Physical Chemistry C* 120.26 (June 2016), pp. 14428–14434. DOI: 10.1021/acs.jpcc.6b04564.

[196] J. P. Gustafsson. "The Surface Chemistry of Imogolite". In: *Clays and Clay Minerals* 49.1 (2001), pp. 73–80. DOI: 10.1346/ccmn.2001.0490106.

[197] M. S. Amara, S. Rouzière, E. Paineau, M. Bacia-Verloop, A. Thill, and P. Launois. "Hexagonalization of Aluminogermanate Imogolite Nanotubes Organized into Closed-Packed Bundles". In: *The Journal of Physical Chemistry C* 118.17 (Apr. 2014), pp. 9299–9306. DOI: 10.1021/jp5029678.

[198] B. Creton, D. Bougeard, K. S. Smirnov, J. Guilment, and O. Poncelet. "Molecular dynamics study of hydrated imogolite : 2. Structure and dynamics of confined water". In: *Physical Chemistry Chemical Physics* 10.32 (2008), p. 4879. DOI: 10.1039/b803479f.

[199] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. "Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces". In: *Physical Review Letters* 100.13 (Apr. 2008). DOI: 10.1103/physrevlett.100.136406.

[200] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. D. L. Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat, and B. Kirtman. "CRYSTAL14: A program for the ab-initio investigation of crystalline solids". In: *International Journal of Quantum Chemistry* 114.19 (Mar. 2014), pp. 1287–1317. DOI: 10.1002/qua.24658.

[201] W. C. Ackerman, D. M. Smith, J. C. Huling, Y. W. Kim, J. K. Bailey, and C. J. Brinker. "Gas/vapor adsorption in imogolite: a microporous tubular aluminosilicate". In: *Langmuir* 9.4 (Apr. 1993), pp. 1051–1057. DOI: 10.1021/la00028a029.

[202] S. Mukherjee, V. M. Bartlow, and S. Nair. "Phenomenology of the Growth of Single-Walled Aluminosilicate and Aluminogermanate Nanotubes of Precise Dimensions". In: *Chemistry of Materials* 17.20 (Oct. 2005), pp. 4900–4909. DOI: 10.1021/cm0505852.

[203] J. Zang, S. Chempath, S. Konduri, S. Nair, and D. S. Sholl. "Flexibility of Ordered Surface Hydroxyls Influences the Adsorption of Molecules in Single-Walled Aluminosilicate Nanotubes". In: *The Journal of Physical Chemistry Letters* 1.8 (Mar. 2010), pp. 1235–1240. DOI: 10.1021/jz100219q.

[204] R. T. Cygan, J.-J. Liang, and A. G. Kalinichev. "Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field". In: *The Journal of Physical Chemistry B* 108.4 (Jan. 2004), pp. 1255–1266. DOI: 10.1021/jp0363287.

[205] O. Teleman, B. Jönsson, and S. Engström. "A molecular dynamics simulation of a water model with intramolecular degrees of freedom". In: *Molecular Physics* 60.1 (Jan. 1987), pp. 193–203. DOI: 10.1080/00268978700100141.

[206] G. Y. Gor and N. Bernstein. "Adsorption-Induced Surface Stresses of the Water/Quartz Interface: Ab Initio Molecular Dynamics Study". In: *Langmuir* 32.21 (May 2016), pp. 5259–5266. DOI: 10.1021/acs.langmuir.6b00923.

[207] A. V. Neimark, F.-X. Coudert, A. Boutin, and A. H. Fuchs. "Stress-Based Model for the Breathing of Metal–Organic Frameworks". In: *The Journal of Physical Chemistry Letters* 1.1 (Dec. 2009), pp. 445–449. DOI: 10.1021/jz9003087.

[208] F. Mouhat, D. Bousquet, A. Boutin, L. B. du Bourg, F.-X. Coudert, and A. H. Fuchs. "Softening upon Adsorption in Microporous Materials: A Counterintuitive Mechanical Response". In: *The Journal of Physical Chemistry Letters* 6.21 (Oct. 2015), pp. 4265–4269. DOI: 10.1021/acs.jpclett.5b01965.

[209] J. D. Bernal and R. H. Fowler. "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions". In: *The Journal of Chemical Physics* 1.8 (Aug. 1933), pp. 515–548. DOI: 10.1063/1.1749327.

[210] J. Zang, S. Konduri, S. Nair, and D. S. Sholl. "Self-Diffusion of Water and Simple Alcohols in Single-Walled Aluminosilicate Nanotubes". In: *ACS Nano* 3.6 (May 2009), pp. 1548–1556. DOI: 10.1021/nn9001837.

[211] A. C. Fogarty, E. Duboué-Dijon, D. Laage, and W. H. Thompson. "Origins of the non-exponential reorientation dynamics of nanoconfined water". In: *The Journal of Chemical Physics* 141.18 (Nov. 2014), p. 18C523. DOI: 10.1063/1.4896983.

[212] A. Ozkanlar and A. E. Clark. "ChemNetworks: A complex network analysis tool for chemical systems". In: *Journal of Computational Chemistry* 35.6 (Dec. 2013), pp. 495–505. DOI: 10.1002/jcc.23506.

[213] V. A. Makarov, B. K. Andrews, P. E. Smith, and B. M. Pettitt. "Residence Times of Water Molecules in the Hydration Sites of Myoglobin". In: *Biophysical Journal* 79.6 (Dec. 2000), pp. 2966–2974. DOI: 10.1016/s0006-3495(00)76533-7.

[214] S. B. Lippman, J. Lajoie, and B. E. Moo. *C++ Primer*. 5th ed. Addison Wesley, 2012. ISBN: 978-0-321-71411-4.

[215] S. Meyers. *Effective Modern C++*. O'Reilly, 2014. ISBN: 978-1-491-90399-5.

[216] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr. "RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials". In: *Molecular Simulation* 42.2 (Feb. 2015), pp. 81–101. DOI: 10.1080/08927022.2015.1010082.

[217] A. Mulholland, F. Manby, S. McIntosh-Smith, J. Michel, and C. Woods. *Sire: An advanced, multiscale, molecular simulation framework*. URL: http://siremol.org/.

[218] V. Shen. *Standard Reference Simulation Website, NIST Standard Reference Database 173*. 2006. URL: https://www.nist.gov/programs-projects/nist-standard-reference-simulation-website.

[219] *Approximate cost to access various caches and main memory*. 2019. URL: https://stackoverflow.com/q/4087280/4692076.

[220] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *Journal of Computational Physics* 23.3 (Mar. 1977), pp. 327–341. DOI: 10.1016/0021-9991(77)90098-5.

[221] H. C. Andersen. "Rattle: A velocity version of the shake algorithm for molecular dynamics calculations". In: *Journal of Computational Physics* 52.1 (Oct. 1983), pp. 24–34. DOI: 10.1016/0021-9991(83)90014-1.

[222] W. Smith. "Calculating the Pressure". In: *Information quarterly for computer simulation of condensed phases CCP5* 39 (Oct. 1993), pp. 14–20.

[223] S. M. J. Rogge, R. Goeminne, R. Demuynck, J. J. Gutiérrez-Sevillano, S. Vandenbrande, L. Vanduyfhuys, M. Waroquier, T. Verstraelen, and V. V. Speybroeck. "Modeling Gas Adsorption in Flexible Metal–Organic Frameworks via Hybrid Monte Carlo/Molecular Dynamics Schemes". In: *Advanced Theory and Simulations* 2.4 (Jan. 2019), p. 1800177. DOI: 10.1002/adts.201800177.

[224] J. A. Izaguirre and S. S. Hampton. "Shadow hybrid Monte Carlo: an efficient propagator in phase space of macromolecules". In: *Journal of Computational Physics* 200.2 (Nov. 2004), pp. 581–604. DOI: 10.1016/j.jcp.2004.04.016.

[225] G. A. Ross, A. S. Rustenburg, P. B. Grinaway, J. Fass, and J. D. Chodera. "Biomolecular Simulations under Realistic Macroscopic Salt Conditions". In: *The Journal of Physical Chemistry B* 122.21 (Apr. 2018), pp. 5466–5486. DOI: 10.1021/acs.jpcb.7b11734.

[226] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth. "Hybrid Monte Carlo". In: *Physics Letters B* 195.2 (Sept. 1987), pp. 216–222. DOI: 10.1016/0370-2693(87)91197-x.

[227] B. Mehlig, D. W. Heermann, and B. M. Forrest. "Hybrid Monte Carlo method for condensed-matter systems". In: *Physical Review B* 45.2 (Jan. 1992), pp. 679–685. DOI: 10.1103/physrevb.45.679.

[228] H. C. Andersen. “Molecular dynamics simulations at constant pressure and/or temperature”. In: *The Journal of Chemical Physics* 72.4 (Feb. 1980), pp. 2384–2393. DOI: 10.1063/1.439486.

[229] R. Faller and J. J. de Pablo. “Constant pressure hybrid Molecular Dynamics–Monte Carlo simulations”. In: *The Journal of Chemical Physics* 116.1 (2002), p. 55. DOI: 10.1063/1.1420460.

[230] M. Fernández-Pendás, B. Escribano, T. Radivojević, and E. Akhmatskaya. “Constant pressure hybrid Monte Carlo simulations in GROMACS”. In: *Journal of Molecular Modeling* 20.12 (Nov. 2014). DOI: 10.1007/s00894-014-2487-y.

[231] A. M. Horowitz. “A generalized guided Monte Carlo algorithm”. In: *Physics Letters B* 268.2 (Oct. 1991), pp. 247–252. DOI: 10.1016/0370-2693(91)90812-5.

[232] E. Akhmatskaya, N. Bou-Rabee, and S. Reich. “A comparison of generalized hybrid Monte Carlo methods with and without momentum flip”. In: *Journal of Computational Physics* 228.6 (Apr. 2009), pp. 2256–2265. DOI: 10.1016/j.jcp.2008.12.014.

[233] E. Akhmatskaya and S. Reich. “New Hybrid Monte Carlo Methods for Efficient Sampling:from Physics to Biology and Statistics”. In: *Progress in Nuclear Science and Technology* 2.0 (Oct. 2011), pp. 447–462. DOI: 10.15669/pnst.2.447.

[234] S. Lan, V. Stathopoulos, B. Shahbaba, and M. Girolami. “Markov Chain Monte Carlo from Lagrangian Dynamics”. In: *Journal of Computational and Graphical Statistics* 24.2 (Apr. 2015), pp. 357–378. DOI: 10.1080/10618600.2014.902764.

[235] Y. Fang, J. M. Sanz-Serna, and R. D. Skeel. “Compressible generalized hybrid Monte Carlo”. In: *The Journal of Chemical Physics* 140.17 (May 2014), p. 174108. DOI: 10.1063/1.4874000.

[236] C. Balzer, G. Reichenauer, and M. Wiener. “Sorption-Induced Deformation of Microporous Solids Studied by In-Situ Dilatometry”. In: *Poromechanics V*. American Society of Civil Engineers, June 2013. DOI: 10.1061/9780784412992.083.

[237] P. P. Ewald. “Die Berechnung optischer und elektrostatischer Gitterpotentiale”. In: *Annalen der Physik* 369.3 (1921), pp. 253–287. DOI: 10.1002/andp.19213690304.

[238] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht. “Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r-1 summation”. In: *The Journal of Chemical Physics* 110.17 (May 1999), pp. 8254–8282. DOI: 10.1063/1.478738.

[239] C. J. Fennell and J. D. Gezelter. “Is the Ewald summation still necessary? Pairwise alternatives to the accepted standard for long-range electrostatics”. In: *The Journal of Chemical Physics* 124.23 (June 2006), p. 234104. DOI: 10.1063/1.2206581.

[240] A. H. Mao and R. V. Pappu. “Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions”. In: *The Journal of Chemical Physics* 137.6 (Aug. 2012), p. 064104. DOI: 10.1063/1.4742068.

[241] N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. “Open Babel: An open chemical toolbox”. In: *Journal of Cheminformatics* 3.1 (2011), p. 33. DOI: 10.1186/1758-2946-3-33.

[242] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. In: *Journal of Molecular Graphics* 14.1 (Feb. 1996), pp. 33–38. DOI: 10.1016/0263-7855(96)00018-5.

[243] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations". In: *Journal of Computational Chemistry* 32.10 (Apr. 2011), pp. 2319–2327. DOI: 10.1002/jcc.21787.

[244] N. M. O'Boyle, A. L. Tenderholt, and K. M. Langner. "cclib: A library for package-independent computational chemistry algorithms". In: *Journal of Computational Chemistry* 29.5 (2008), pp. 839–845. DOI: 10.1002/jcc.20823.

[245] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen. "The atomic simulation environment — a Python library for working with atoms". In: *Journal of Physics: Condensed Matter* 29.27 (June 2017), p. 273002. DOI: 10.1088/1361-648x/aa680e.

[246] E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha, and C. Steinbeck. "The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching". In: *Journal of Cheminformatics* 9.1 (June 2017). DOI: 10.1186/s13321-017-0220-4.

[247] *AMBER NetCDF Trajectory/Restart ConventionVersion 1.0, Revision C*. URL: http://ambermd.org/netcdf/nctraj.xhtml.

[248] P. E. Bourne, H. M. Berman, B. McMahon, K. D. Watenpaugh, J. D. Westbrook, and P. M. Fitzgerald. "[30] Macromolecular crystallographic information file". In: *Methods in Enzymology*. Elsevier, 1997, pp. 571–590. DOI: 10.1016/s0076-6879(97)77032-0.

[249] A. R. Bradley, A. S. Rose, A. Pavelka, Y. Valasatava, J. M. Duarte, A. Prlić, and P. W. Rose. "MMTF–An efficient file format for the transmission, visualization, and analysis of macromolecular structures". In: *PLOS Computational Biology* 13.6 (June 2017). Ed. by D. Schneidman, e1005575. DOI: 10.1371/journal.pcbi.1005575.

[250] H. Berman, K. Henrick, and H. Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nature Structural & Molecular Biology* 10.12 (Dec. 2003), pp. 980–980. DOI: 10.1038/nsb1203-980.

[251] M. Lundborg, R. Apostolov, D. Spangberg, A. Gärdenäs, D. van der Spoel, and E. Lindahl. "An efficient and extensible format, library, and API for binary trajectory data from molecular simulations". In: *Journal of Computational Chemistry* 35.3 (Nov. 2013), pp. 260–269. DOI: 10.1002/jcc.23495.

[252] P. de Buyl, P. H. Colberg, and F. Höfling. "H5MD: A structured, efficient, and portable file format for molecular data". In: *Computer Physics Communications* 185.6 (June 2014), pp. 1546–1553. DOI: 10.1016/j.cpc.2014.01.018.

[253] G. Clavier, N. Desbiens, E. Bourasseau, V. Lachet, N. Brusselle-Dupend, and B. Rousseau. "Computation of elastic constants of solids using molecular simulation: comparison of constant volume and constant pressure ensemble methods". In: *Molecular Simulation* 43.17 (May 2017), pp. 1413–1422. DOI: 10.1080/08927022.2017.1313418.

[254] G. R. Kneller, V. Keiner, M. Kneller, and M. Schiller. “nMOLDYN: A program package for a neutron scattering oriented analysis of Molecular Dynamics simulations”. In: *Computer Physics Communications* 91.1-3 (Sept. 1995), pp. 191–214. DOI: 10.1016/0010-4655(95)00048-k.

RÉSUMÉ EN FRANÇAIS

Introduction

Les matériaux poreux sont des matériaux présentant une porosité structurelle avec des cavités appelées *pores* dans leur la structure tridimensionnelle. Ce réseau de pores peuvent varier en homogénéité et en régularité, créant ainsi une grande variété de matériaux poreux. Ils ont tous en commun une surface spécifique, à savoir la surface interne accessible par grammes de matériau, élevée — jusqu'à des milliers de mètres carrés par gramme de matériau[1] dans les cas les plus extrêmes. Cette très grande surface spécifique est exploitée dans nombre d'applications industrielles importantes, notamment dans les domaines de l'adsorption et de la catalyse. Par exemple, ils sont utilisés pour séparer dans des mélanges de gaz ou de liquide sous forme de tamis moléculaires ; pour filtrer et éliminer les métaux lourds dans l'eau ou comme catalyseurs hétérogènes dans les raffineries pétrolières lors du procédé de craquage.

L'union internationale de chimie pure et appliquée (IUPAC) recommande une classification des matériaux poreux en trois groupes, selon la taille des pores[2]. On trouve tout d'abord les solides *microporeux*, dont les pores ont un diamètre inférieur à 2 nm. Les solides *mésoporeux* ont des pores dont le diamètre est compris entre 2 et 50 nm. Enfin, les solides avec pores plus grands que 50 nm sont dit *macroporeux*. Les solides microporeux et mésoporeux sont souvent regroupés sous l'appellation de solides *nanoporeux*, dans lesquels la taille des pores ne dépasse pas 50 nm.

Deux familles de matériaux nanoporeux cristallins sont particulièrement intéressantes. Tout d'abord, les zéolithes sont des aluminosilicates poreux naturels et artificiels connus depuis 1756 et synthétisé artificiellement depuis les années 1940. Elles sont actuellement très utilisées industriellement, en particulier comme catalyseurs dans l'industrie pétrolière et comme adoucisseurs d'eau dans les lessives. Depuis les années 2000, une nouvelle famille

de matériaux nanoporeux cristallins hybride organiques–inorganiques appelés *Metal–Organic Frameworks* (MOF) ont été découvert et ont suscité l'intérêt de la communauté scientifique. Ces nouveaux matériaux sont construits à partir de centres métalliques, reliées entre eux par des ligands organiques. Cette méthode de construction permet un très grand éventail de structures différentes, et ouvre la voie à des matériaux conçu pour une application spécifique. En changeant la combinaison de ligands et métaux utilisés, il est possible de changer la taille, la forme et le comportement physico-chimique des pores, comme illustré sur la figure R1.

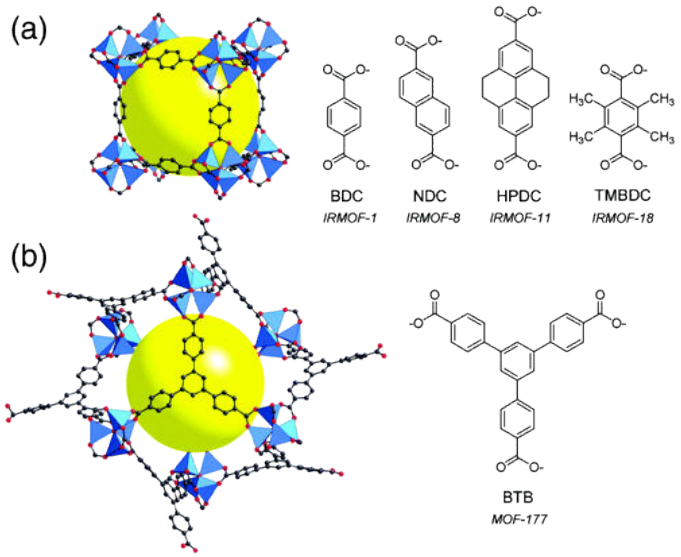


Figure R1 – Deux exemples de MOFs construit avec un centre métallique zinc. (a) Structure de la MOF-5, avec un ligand linéaire. (b) Structure de la MOF-177, avec un ligand trigonal. Reproduit avec permission de la référence [18], copyright (2004) American Chemical Society.

La faiblesse relative des liaisons de coordination entre les cations métalliques et les ligands organiques est à l'origine d'une flexibilité structurelle intrinsèque dans les MOFs, qui peut être locale ou étendue à l'ensemble du matériau. Certain MOF, regroupés sous l'appellation *"soft porous crystals"*, réagissent à des stimuli externes comme la température, la pression, l'adsorption de gaz ou même l'exposition à la lumière avec des modifications de grande envergure de leur structure. Les différents modes de flexibilité pouvant exister dans les MOFs sont représentés en figure R2. Si tous les MOFs peuvent présenter des déformations locales comme la rotation de ligands, seuls les *soft porous crystals* présente des déformations globale. Par exemple, les matériaux de la famille MIL-53 présentent pendant l'adsorption de gaz deux transitions de phase, allant d'une phase à pores ouverts à une phase à pores fermés, puis de nouveau à la phase à pores ouverts[23] lorsque l'on augmente de manière continue de la quantité de gaz adsorbé, donnant l'impression d'une *respiration* du matériau.

La plupart des applications des matériaux nanoporeux sont liées à l'entrée d'autres espèces chimiques (en phase liquide ou gazeuse) à l'intérieur des pores du matériau. Lorsque le fluide à l'extérieur est dans l'état gazeux, le processus est appelé adsorption, tandis que pour les liquides on parle d'intrusion. L'adsorption et l'intrusion ont tous deux un effet sur les propriétés physiques et chimiques du fluide confiné dans les nanopores. Les

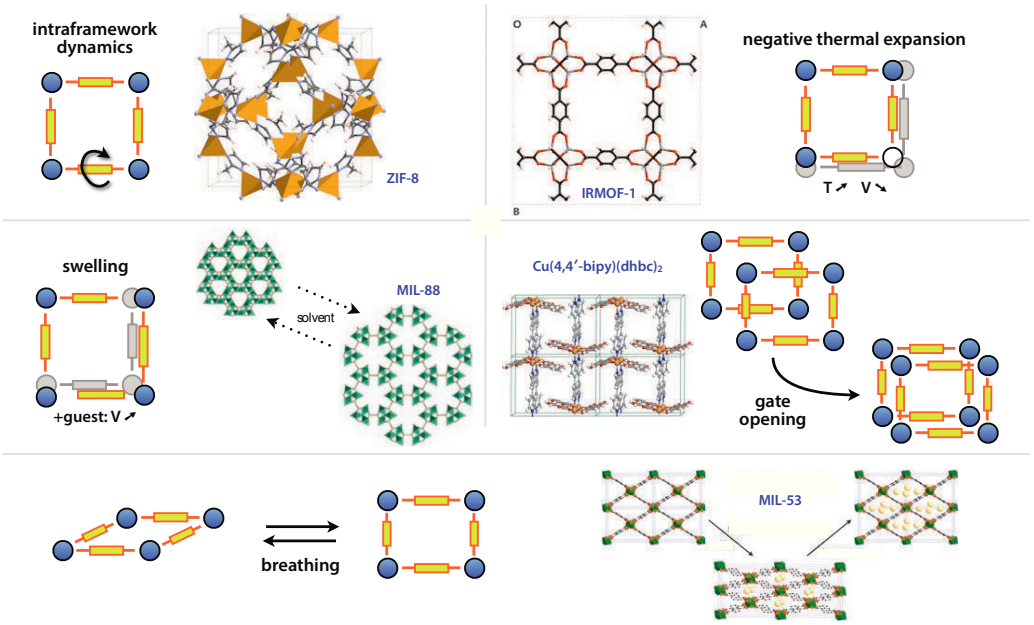


FIGURE R2 – Illustration des principaux modes de flexibilité des MOFs : rotation de ligands, expansion thermique, gonflement, ouverture de porte et respiration. Reproduit avec permission de la référence [20], copyright (2011) Wiley.

fluides confinés sont généralement organisés plus régulièrement, prenant ainsi l'aspect d'une phase solide tout en restant mobiles. L'inverse est également vrai : la présence du fluide à l'intérieur des pores peut modifier les propriétés et le comportement du matériau environnant. Les matériaux flexibles sont particulièrement touchés et peuvent subir des transitions de phase induites par l'adsorption, entraînant des phénomènes macroscopiques comme une *"ouverture de portes"* (*gate-opening*), la respiration ou l'adsorption négative de gaz. Le couplage entre l'adsorption ou l'intrusion et les changements de structure des matériaux nanoporeux flexibles est difficile à étudier, car il implique des équilibres de phases entre les fluides confinés et à l'état pur, ainsi que l'équilibre entre différentes phases du matériau.

Pendant ma thèse, je me suis intéressé à la simulation moléculaire de l'adsorption et de l'intrusion de fluides dans les matériaux nanoporeux flexibles. Les outils de simulation moléculaire peuvent accélérer le développement de nouveaux matériaux adaptés à des applications spécifiques. La prédiction des propriétés de ces matériaux avant leur synthèse réduit fortement le coût de création de nouveaux matériaux. Les propriétés qui peuvent être étudiées et la fiabilité des prédictions correspondantes dépendent des techniques utilisées pour modéliser les systèmes en question. Au cours de ma thèse, j'ai utilisé plusieurs techniques à différentes échelles de temps et de taille pour étudier l'adsorption et l'intrusion dans les matériaux poreux flexibles, et leurs effets tant sur le fluide confiné que les matériaux. J'ai ainsi utilisé des modèles basés sur la thermodynamique classique pour l'étude de la co-adsorption des gaz, des simulations *"umbrella sampling"* pour extraire des profils d'énergie libre d'intrusion de l'eau, la dynamique moléculaire classique et les simulations Monte-Carlo classiques pour l'intrusion d'eau et de solutions aqueuses dans des matériaux rigides et flexibles, et des simulations de dynamique moléculaire *ab initio* pour étudier l'organisation moléculaire des fluides dans des pores.

4 Séparation de gaz dans des matériaux flexibles

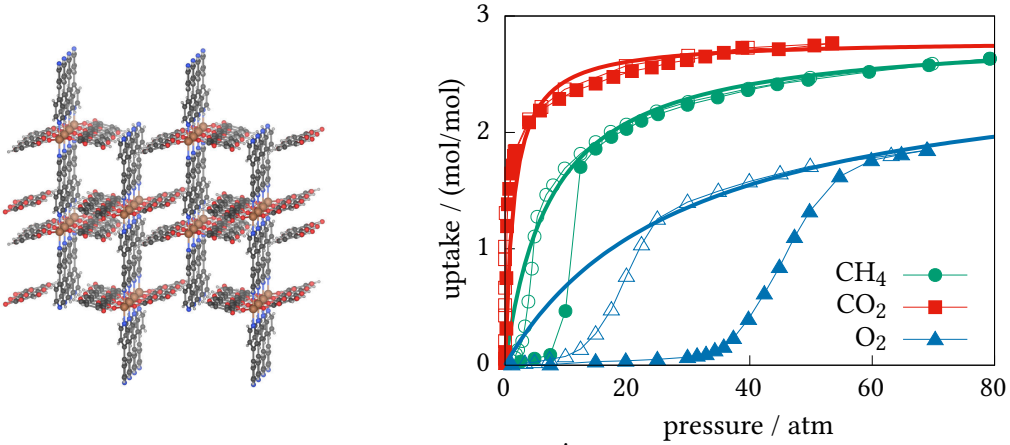


Figure R3 – (gauche) Structure de Cu(dhbc)$_2$(4,4′-bpy). (droite) Isothermes d'adsorption à 298 K dans Cu(dhbc)$_2$(4,4′-bpy) pour différents gaz. L'adsorption est représenté en symboles plein, la desorption en symboles creux. Les données expérimentale ont été publiées par Kitaura et al.[66]

La séparation de gaz est une étape importante dans de multiples procédés industriels, allant de séparation des hydrocarbures dans la chimie du pétrole, à la séparation et stockage du CO_2 ou l'extraction d'oxygène dans l'air. Les deux principales méthodes utilisées pour la séparation des gaz sont la distillation cryogénique, principalement utilisée pour la séparation de l'air, et l'adsorption différentielle. Les procédés de séparation des gaz basés sur l'adsorption sont très polyvalents grâce au large choix de matériaux disponibles, et de la possibilité d'adapter l'adsorbant à un système de gaz spécifique. Le choix d'un adsorbant et la taille d'une unité de production de séparation d'un mélange gazeux, nécessite une bonne connaissance des propriétés de co-adsorption de ces gaz.

La caractérisation expérimentale de la co-adsorption est en général longue et couteuse, à cause du grand nombre de paramètres à faire varier. Ce coût important est à l'origine de plusieurs modèles théoriques permettant de prédire la co-adsorption à partir de données d'adsorption de corps purs. La méthode la plus couramment utilisée sur la théorie des solutions adsorbée idéale (*Ideal Adsorbed Solution Theory*, IAST)[62], qui est robuste et relativement simple à mettre en œuvre. En pratique, l'utilisation de la méthode IAST revient à travailler dans l'ensemble thermodynamique grand-canonique, et à considérer la matrice d'adsorption comme étant rigide. Pour pouvoir décrire la flexibilité des matériaux lors de l'adsorption, il faut utiliser l'ensemble osmotique à la place. Cet ensemble est à la base de la méthode OFAST (*Osmotic Framework Adsorbed Solution Theory*) développée dans le groupe[80, 81]. Pour une description plus en détail de ces méthodes, je renvoie le lecteur au chapitre 2 de ce manuscrit ou à l'article correspondant[82].

Durant ma thèse, j'ai utilisé à la fois IAST et OFAST pour calculer la sélectivité d'adsorption dans les MOFs Cu(dhbc)$_2$(4,4′-bpy) et RPM3-Zn, et démontrer l'inadéquation d'IAST lorsque les structures adsorbantes sont flexibles. Je vais rappeler rapidement ici les résultats obtenus dans le cas de Cu(dhbc)$_2$(4,4′-bpy). L'adsorption dans ce matériau, présenté en figure R3, est représentatif du phénomène d'ouverture de portes : en dessus d'une pression "d'ouverture", aucune molécule de gaz n'entre dans la structure, et au-delà

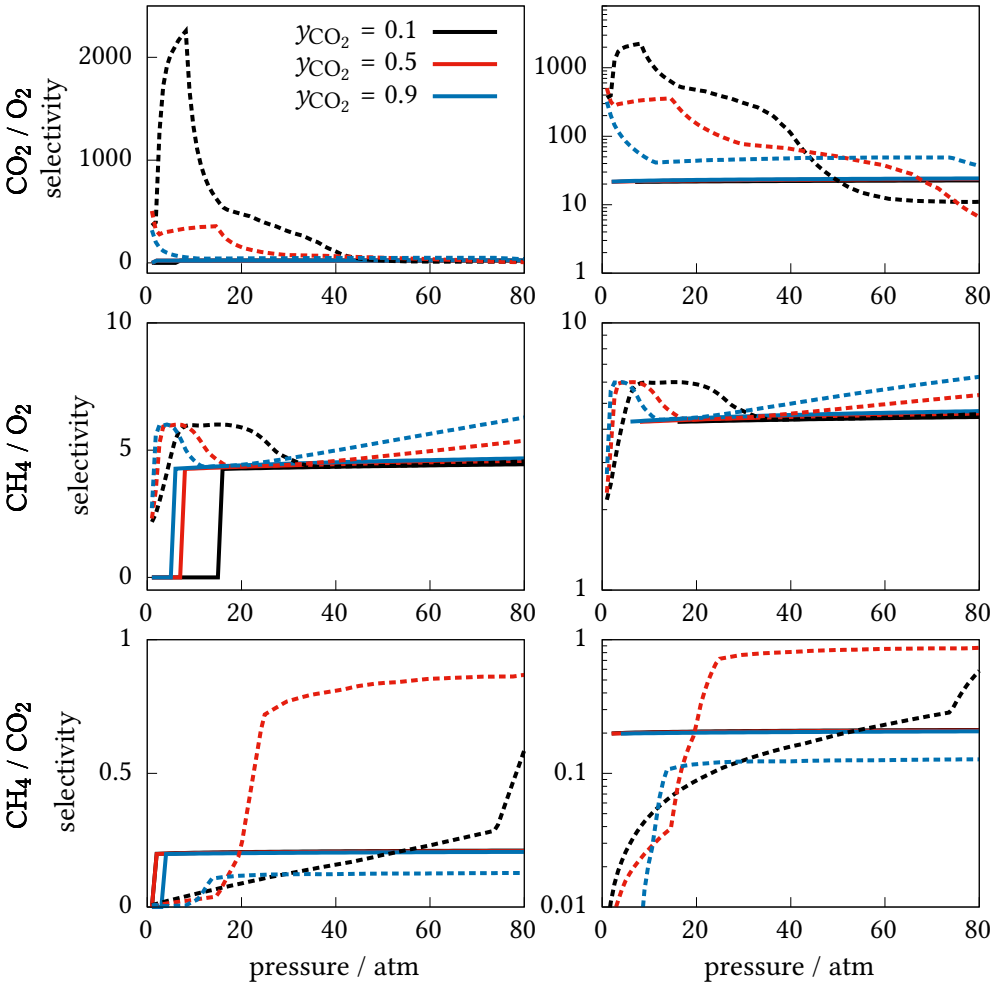


FIGURE R4 – Comparaison de la sélectivité prédite par la méthodes IAST (lignes pointillées) ou OFAST (lignes pleines) pour les couples CO_2/O_2 (haut); CH_4/O_2 (milieu) et CH_4/CO_2 (bas) dans Cu(dhbc)$_2$(4,4′-bpy); en échelle linéaire (gauche) et logarithmique (droite).

de cette pression l'adsorption se produit normalement. La sélectivité prédite à partir de ces isothermes d'adsorption par IAST et OFAST est présentée en figure R4.

La sélectivité d'adsorption calculée avec OFAST se comporte comme attendu : à basse pression, les pores sont fermés et aucun gaz n'entre dans la structure. Ensuite, à une pression dépendant de la composition de la phase gazeuse, l'ouverture de porte se produit. Pour les pressions plus importantes, le matériau est sous sa forme ouverte, et la valeur de sélectivité dépend de la différence en capacité d'absorption entre les deux gaz, environ 20 pour les mélanges CO_2/O_2 et 4 pour les mélanges CH_4/O_2. Au contraire, les sélectivités calculées par la méthode IAST sont clairement non physiques. Toutes les courbes de sélectivité présentent un maximum dans la plage de pression où se produit l'ouverture de la structure, avec des sélectivités qui peuvent être plusieurs ordres de grandeur trop élevées, par exemple 2 000 au lieu de 20 pour CO_2/O_2. Même loin de la pression de transition, les sélectivités prédites par IAST ne reproduisent pas celles calculées avec OFAST, le comportement incorrect à basse pression affectant directement les calculs au-delà. J'ai donc pu confirmer par une étude quantitative que IAST n'est pas adapté à l'adsorption dans les matériaux flexibles nanoporeux.

5 Adsorption de N_2 dans la ZIF-8

Les ZIFs (*zeolitic Imidazolate Frameworks*) sont une famille de MOFs construits autour de centres métalliques tétravalents (Fe, Co, Zn, Cd, Cu) assemblés par des ligands imidazolate ou imidazolate fonctionnalisés. Les premières ZIF (ZIF-1 à ZIF-12) ont été synthétisés en 2006[19], et se sont révélés être plus résistantes à l'eau et la chaleur que les MOFs typiques, ce qui les rend particulièrement attractif pour des applications commerciales. J'ai spécifiquement étudié la ZIF-8 pendant ma thèse. Cette structure est construite avec des ligands 2-méthyl-imidazolate (mim) et des ions Zn^{2+} comme centre métallique, assemblées dans une structure sodalite représentée en figure 4.1. Dans cette topologie, les pores principaux sont quasi-sphériques et correspondent aux cages sodalite, reliés par des fenêtres formées par 6 et 4 atomes de zinc. Deux nouveaux matériaux analogues à ZIF-8 ont été synthétisés récemment par Li et al.[125] utilisant des ligands imidazolate chloro- et bromo-substitués à la place du méthyl-imidazolate de la ZIF-8. Ces nouveaux matériaux, que nous appellerons ZIF-8(Cl) et ZIF-8(Br), sont représentés dans la figure 4.2.

Les isothermes d'adsorption de l'azote à 77 K dans ZIF-8, ZIF-8(Cl) et ZIF-8(Br), sont présentés dans la figure 4.3. Elles ont été mesurées dans le cadre d'une collaboration avec l'équipe expérimentale du laboratoire IS2M de l'Université de Haute Alsace. Sur ces isothermes, nous remarquons tout d'abord les deux sauts pour le cas de ZIF-8(CH_3) : un premier correspondant au remplissage initial des pores jusqu'aux environs de 300 cm^3/cm^3, et un second jusqu'à 400 cm^3/cm^3. Ce comportement est celui d'une isotherme de type IV en la classification de l'IUPAC[43] (voir la figure 1.12). Concernant les deux nouveaux matériaux, l'adsorption dans la ZIF-8(Cl) se fait de la même manière, avec une isotherme de type IV ; alors que dans ZIF-8(Br), l'adsorption a lieu avec une isotherme de type I, sans second saut dans la quantité adsorbée. La présence du second saut dans les isothermes pour ZIF-8(CH_3) et ZIF-8(Cl) est surprenant ici, ce type d'isotherme étant en général associée à la présence de plusieurs niveaux de porosités ou à des transitions de phase de l'hôte. Il semblerait ici que la ZIF-8 subisse une transition de phase associée à une rotation des ligands[128, 129].

Afin de mieux comprendre la relation entre la structure des ZIFs utilisées l'adsorption de N_2 dans ces ZIFs, j'ai utilisé des simulations de dynamique moléculaire. Afin de pouvoir décrire pleinement la flexibilité de la ZIF-8, j'ai favorisé la dynamique moléculaire *ab initio* à la dynamique moléculaire classique, basée sur un champ de force. Cette étude est publiée dans *The Journal of Physical Chemistry C* (2018)[124].

Un premier indicateur de déformation de la structure est l'angle dièdre Zn – Zn – Zn – X — présenté dans la figure 4.1 — où X est CH_3, Cl or Br selon la ZIF utilisée. Il représente la rotation des ligands autour du plan de la fenêtre à 6 éléments, 0° étant le point où les ligands sont alignés avec ce plan. L'évolution de la distribution de cet angle quand le remplissage augmente est représentée sur la figure R5, où nous observons pour ZIF-8(CH_3) une augmentation progressive de la valeur moyenne de l'angle à mesure que la quantité adsorbée augmente alors que la distribution reste gaussienne. Les deux autres structures se comportent différemment. Pour ZIF-8(Cl), presque aucun changement n'est présent dans la distribution lors de l'adsorption jusqu'au remplissage le plus élevé (*i.e.*, N = 50). À ce moment, la distribution se déplace et le profil n'est plus de type gaussien, mais ressemble à la somme de deux distributions gaussiennes, l'une centrée autour de 25° et

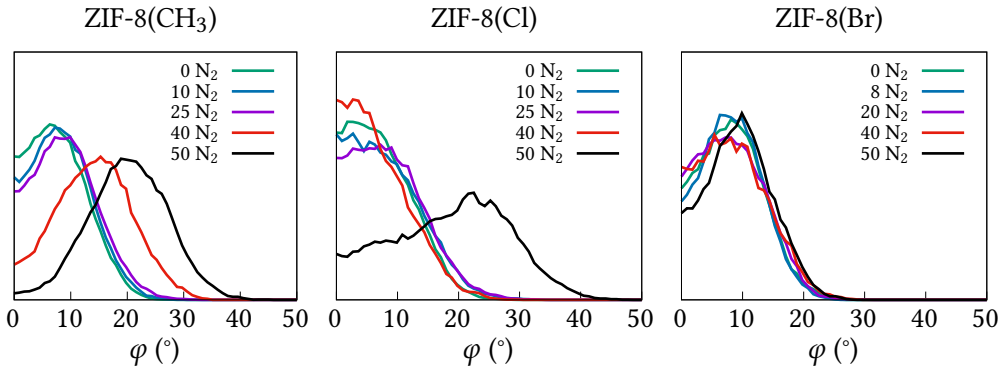


FIGURE R5 – Distribution des angles dihèdre Zn – Zn – Zn – X — où X est CH_3, Cl or Br selon la ZIF utilisée — à différent remplissage d'azote.

l'autre autour de 10°. Ceci indique vraisemblablement que certains des ligands ne tournent pas même à fort remplissage. Enfin, pour ZIF-8(Br), aucun changement significatif ne se produit pour la distribution des angles dièdre même aux plus grandes valeurs de remplissage, ce qui indique que les ligands ne tournent pas pendant l'adsorption.

Bien que ce comportement semble être corrélé à la présence ou à l'absence d'un second saut dans les isothermes d'adsorption, il n'est cependant pas suffisant l'expliquer. Pour tenter de faire ce lien, j'ai calculé la distribution des tailles de pores (voir la figure 4.5) et le volume poreux accessible (voir la figure 4.6). Pour obtenir ces valeurs, j'ai commencé par vider la structure de toutes les molécules d'azote, puis j'ai calculé le volume poreux et la distribution de tailles de pores dans les structures vides restant à l'aide du logiciel zeo++[135]. Sur ces figures, on peut observer que la distribution de taille des pores reste constante — même lorsque les ligands tournent. De même, le volume poreux accessible reste constant. Il y a donc un autre phénomène en œuvre qui est à l'origine du second saut dans les isothermes pour ZIF-8(CH_3) et ZIF-8(Cl).

Une autre hypothèse que nous pouvons formuler pour expliquer la présence du second saut dans les isothermes est que les molécules d'azote subissent une ré-organisation dans les pores, entraînant ainsi une augmentation du nombre de molécules adsorbées dans un volume poreux constant ainsi que la rotation des ligands. Afin de visualiser l'organisation des molécules, j'ai projeté les positions de tous les atomes d'azote adsorbés dans le plan xy et créé une carte de densité de la phase adsorbée. Cette carte de densité est présentée figure R6, pour différents remplissages et pour les trois structures.

Pour ZIF-8(CH_3), on rencontre deux organisations moléculaires différentes en fonction du remplissage. Avec 10 ou 25 molécules par maille, les cartes de densité montrent des positions clairement délimitées sur un arrangement cubique, tandis qu'avec 40 et 50 molécules, elles montrent un arrangement tétragonal de ces mêmes molécules. Pour ZIF-8(Cl), le comportement est à peu près similaire : les molécules s'organisent d'abord de façon cubique lorsqu'il y a 10, 25 et 40 molécules par mailles, avant de se réordonner vers un arrangement tétragonal à 50 molécules par mailles.

Ceci est cohérent avec les distributions des angles dièdres en figure 4.4, et la preuve que le réarrangement des molécules d'azote se produit conjointement avec la rotation des ligands. Il est intéressant de noter qu'un certain désordre dans l'arrangement tétragonal

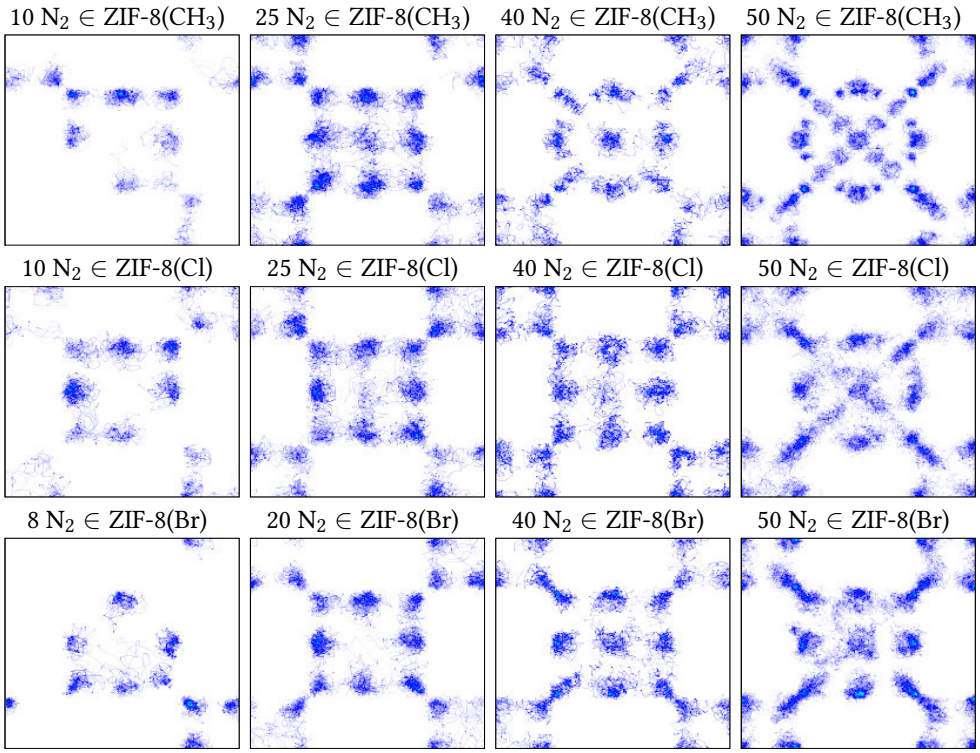


FIGURE R6 – Cartes de densité 2D de l'azote adsorbé dans le plan xy à divers remplissages dans ZIF-8(CH_3) (haut), ZIF-8(Cl) (milieu), et ZIF-8(Br) (bas).

reste présent — même à un remplissage de 50 molécules par maille — les positions des molécules n'étant pas aussi bien définies que dans la ZIF-8(CH_3). Encore une fois, ceci est cohérent avec la distribution des angles dièdres pour un remplissage de 50 molécules par mailles pour ZIF-8(Cl) qui n'est pas de type gaussien, indiquant que ce désordre se retrouve aussi dans la structure. Pour ZIF-8(Br), le comportement est différent. On trouve le même agencement cubique aux remplissages de 8 et 20 molécules par mailles, alors que l'agencement pour 40 et 50 molécules par mailles n'est pas le même que celui observé dans la ZIF-8(CH_3) et ZIF-8(Cl). Les molécules semblent être principalement réparties sur un cube avec des molécules supplémentaires sur les diagonales du cube.

Le mécanisme que je propose pour l'adsorption de l'azote dans les trois structures ZIF-8 est illustré schématiquement en figure 4.9. Lorsque les pores sont vides, les ligands sont dans leur position d'équilibre, autour de 0°. Au fur et à mesure que le remplissage augmente, les pores se remplissent selon un agencement cubique. Puis, comme le remplissage continue à augmenter, les molécules se ré-arrangent dans ZIF-8(CH_3) et ZIF-8(Cl) pour passer à un arrangement tétragonal. La rotation des ligands accompagne ce ré-arrangement des molécules. Aucun de ces deux phénomènes n'a lieu dans ZIF-8(Br), sans doutes à cause des différences de taille et de forme des pores, les pores de ZIF-8(Br) étant plus petits que ceux de ZIF-8(Cl) et de ZIF-8(CH_3).

6 Champs de force à partir de données *ab initio*

Pendant ma thèse, j'ai également collaboré avec Johannes Dürholt et Rochus Schmid de la Ruhr Universität à Bochum en Allemagne. Nous avons paramétré un champ de force classique pour ZIF-8(CH_3), ZIF-8(Cl) et ZIF-8(Br) en utilisant les simulations *ab initio* présentées précédemment. L'article correspondant est publié dans *Journal of Chemical Theory and Computation* (2019)[136].

Idéalement, nous aimerions que les champs de force que nous utilisons soient aussi précis que possible lors du calcul de l'énergie d'un système, afin d'être aussi sûrs que possible que le modèle décrit par le champ de force décrive la réalité chimique. Cela nous impose de créer un champ de force spécifique pour chaque molécule et pour chaque combinaison de molécules. Cependant, le fait d'avoir des champs de force distincts pour chaque système peut nous empêcher de comparer les propriétés prédites par les simulations. Nous ne pouvons pas attribuer avec certitude les différences de prédiction à la réalité chimique sous-jacente, et non au modèle utilisé. C'est la raison pour laquelle des champs de force génériques ont également été développés. Ils sacrifient un peu de précision pour plus transférables : ces champs de forces sont utilisables pour de multiples systèmes moléculaires avec à peu près la même précision partout. Malheureusement, les champs de force génériques existants ne sont pas toujours optimaux pour simuler les MOF, principalement en raison des liens de coordination existant entre les centres métalliques et les lieurs organiques.

Une façon de surmonter ces problèmes (temps de paramétrage long, compromis entre champs de force transférables et champs de force précis) est d'utiliser un algorithme de paramétrisation systématique. L'utilisation du même algorithme pour tous les systèmes d'intérêt permet d'améliorer la précision, car chaque système a des paramètres qui lui sont propres, tout en permettant la comparaison entre différents systèmes, car ils partagent la même forme fonctionnelle et les mêmes données de référence. Plusieurs approches ont été développées pour générer de nouveaux champs de forces pour les MOF, telles que Quick-FF[141] et MOF-FF[142]. Les deux utilisent des données de référence *ab initio*, telles que la géométrie optimisée, et la matrice Hessienne du système. Pour optimiser les paramètres du champ de force, ils utilisent des algorithmes d'apprentissage machine, tels que des algorithmes génétiques.

Johannes Dürholt a utilisé MOF-FF pour générer de nouveaux champs de force pour ZIF-8(CH_3), ZIF-8(Cl), et ZIF-8(Br) à partir des données de référence *ab initio*. J'ai contribué à ce travail en fournissant les simulations de ZIF vides décrites précédemment et en aidant à l'analyse des trajectoires afin de valider le champ de force. Les versions précédentes de MOF-FF utilisaient des clusters finis représentatifs dans le vide pour paramétrer le champs de force. Au cours de ces travaux, la stratégie d'ajustement des paramètres a été améliorée pour permettre l'utilisation de conditions aux limites périodiques. Il n'est donc plus nécessaire de trouver des clusters finis représentatifs et non chargés, ce qui n'est pas toujours possible en fonction de la topologie du MOF considéré. Les champs de force ainsi généré sont capable de reproduire non seulement les propriétés géométriques statiques des trois ZIFs, mais aussi certaines propriétés dynamiques comme les modes de vibration ou globales comme les coefficients d'élasticité (voir la figure 4.10 et les tables 4.1 et 4.2).

7 Intrusion d'électrolytes dans la ZIF-8

L'intrusion de liquides non mouillants, et en particulier du mercure, est utilisée depuis longtemps pour caractériser les matériaux poreux ayant des largeurs de pores comprises entre 50 nmnm et 500 µm[44]. L'intrusion peut être vue comme une adsorption de fluides au-dessus de la pression de vapeur saturante, c'est-à-dire quand le fluide est dans son état liquide. Depuis les années 2000, le mouillage forcé des solutions électrolytiques a été étudié, révélant des effets très intéressants de pression osmotique géante[50, 51], et des applications potentielles pour le stockage et la dissipation de l'énergie mécanique[47]. Les lecteurs intéressés peuvent consulter la revue de littérature que j'ai aidé à écrire sur le sujet, publiée dans *Chemical Society Reviews* (2017)[45]. L'utilisation d'ions dans le liquide d'intrusion permet non seulement de modifier la pression d'intrusion par des effets osmotiques[53, 54], mais aussi de changer la forme de l'isotherme d'intrusion, passant d'un comportement de stockage de l'énergie à la dissipation de cette énergie[52].

Au cours de ma thèse, j'ai étudié l'intrusion haute pression d'électrolyte dans la ZIF-8. Ce matériau ZIF-8 est hydrophobe[155] et présente un comportement intéressant pendant d'intrusion. Alors que les effets de pression osmotique ne dépendent pas de la nature chimique des ions mais seulement de leur concentration, la ZIF-8 passe d'un comportement énergétique à l'autre lorsque la nature ionique change à concentration constante[55]. Le mécanisme exact et le comportement au niveau moléculaire de l'intrusion d'électrolytes dans la ZIF-8 sont encore inconnus. J'ai utilisé des simulations de dynamique moléculaires classiques pour étudier la structure, la dynamique et les implications énergétiques du confinement de l'eau et des solutions aqueuses de LiCl dans la ZIF-8. Pour ce faire, j'ai réalisé des simulations d'eau libre et confinée à différentes pressions (de 0 à 1 Gpa) et avec différentes concentrations d'ions (de 0 à 20 mol/L).

Du point de vue du liquide, le principal effet de l'intrusion est le confinement du fluide dans un espace poreux de dimensions nanométriques. Afin de caractériser la structure du liquide et la solvatation des ions, j'ai calculé le nombre de voisins dans la première couche de solvatation (figure 5.2) et des cartes de densité 2D pour la répartition des molécules dans le pore (figure 5.3). Sur ces figures, on peut observer que le confinement modifie la structure de l'eau et la solvatation des ions, créant une structuration à longue distance du réseau de liaisons hydrogène. Les molécules d'eau occupent des sites très bien définis, notamment à l'intérieur des fenêtres entre deux cages voisines. Au fur et à mesure que la concentration augmente, cette organisation est légèrement perturbée par les ions insérés dans le réseau des molécules d'eau. La présence de la ZIF-8 empêche aussi l'eau de se ré-organiser, et diminue donc la solvatation des ions à haute concentration, quand il n'y a plus assez de molécules d'eau pour chaque ion.

J'ai caractérisé la dynamique de l'eau via celle des liaisons hydrogène. J'ai considéré qu'une liaison hydrogène était présente entre deux molécules d'eau si les deux atomes d'oxygène sont séparés par moins de 3.5 Å, avec un angle oxygène-oxygène-hydrogène ($\widehat{\mathrm{OOH}}$) inférieur à 30°. J'ai ensuite calculé l'auto-corrélation temporelle des fonctions d'existence de la liaison hydrogène $\mathrm{H}(t)$ — qui vaut 1 si la liaison existe au temps t, 0 si elle n'existe pas :

$$\mathrm{C}_{\mathrm{hbonds}}(t) = \left\langle \mathrm{H}(t_0) \cdot \mathrm{H}(t_0 + t) \right\rangle_{t_0} . \tag{A.1}$$

La décroissance de cette fonction d'auto-corrélation est caractéristique de la dynamique du réseau de liaisons hydrogène et de la constante de temps des liaisons hydrogène individuelles. Cette décroissance n'est pas décrite adéquatement par un modèle exponentiel pur, c'est pourquoi j'ai utilisé un modèle bi-exponentiel pour extraire les constantes de temps :

$$f(t) = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}. \tag{A.2}$$

où τ_1 et τ_2 sont les deux constantes de temps de la décroissance, et A_1 et A_2 sont leur poids relatif. Les paramètres d'ajustement résultants sont présentés dans table 5.1 et figure R7.

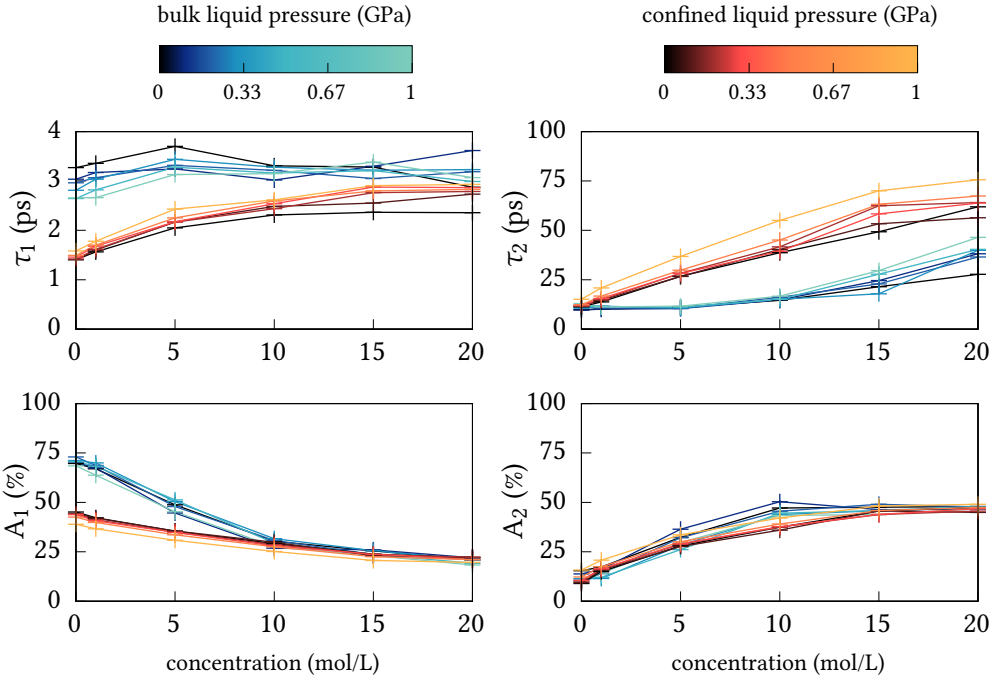


FIGURE R7 – Variations de la constante de temps et de la proportion des deux termes exponentiels dans l'auto-corrélation de liaisons hydrogène en fonction de la pression et de la concentration en ions dans les liquides purs (teintes bleues) et confinés (teintes rouges).

Tout d'abord, nous remarquons que la pression a une influence relativement faible sur la dynamique des liaisons hydrogène. Dans le liquide pur, la constante de temps la plus courte est relativement constante autour de 3 ps lorsque la concentration augmente, mais le poids de ce processus rapide (A_1) diminue. Cela suggère que cette constante de temps rapide est associée à des liaisons hydrogène entre molécules d'eau entourées seulement par d'autres molécules d'eau. La deuxième constante de temps, associée au processus le plus lent, augmente avec la concentration, ainsi que le poids correspondant (A_2). Cela indique des liaisons hydrogène entre les molécules d'eau liées aux ions. Dans le liquide confiné, les poids évoluent de la même façon par rapport à la concentration, ce qui indique qu'ils sont associés au même type de liaisons hydrogène que dans le liquide pur. La deuxième constante de temps augmente dans le liquide confiné par rapport au liquide pur. Ce ralentissement de la dynamique de l'eau en confinement est bien connue[172], et a été observé dans de nombreuses classes de matériaux nanoporeux[160, 174-176].

J'ai également étudié la thermodynamique du processus par lequel des molécules d'eau et des ions peuvent entrer dans les pores de la ZIF-8 en passant par les fenêtres à 6 membres. J'ai donc modélisé une interface eau/ZIF-8 explicite, représentée dans la figure 5.10 : le système contient à la fois de l'eau pure et de l'eau confinée dans la ZIF-8. J'ai utilisé des simulations *umbrella sampling* et la méthode d'analyse WHAM pour reconstruire le profil d'énergie libre des molécules (Li^+, Cl^-, ou H_2O) entrant ZIF-8 selon l'axe cristallographique (111), reproduit en figure R8.

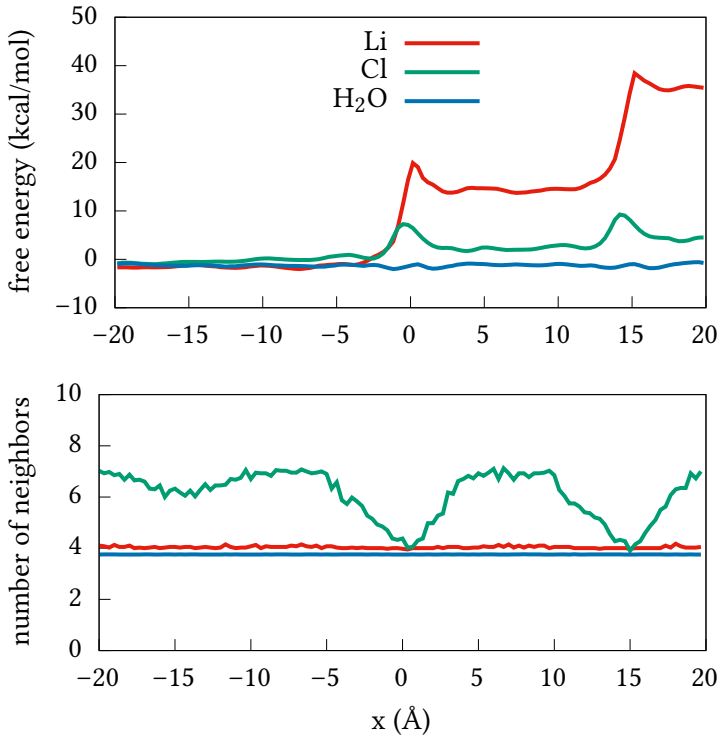


Figure R8 – (haut) Profil d'énergie libre d'une molécule entrant dans la ZIF-8 et nombre correspondant de voisins (bas) dans la première couche de solvatation en fonction de la position de la molécule sur l'axe cristallographique (111). La première fenêtre de la ZIF-8 est à $x = 0$; la zone $x < 0$ correspond à l'eau pure, et la zone $x > 0$ à l'eau confinée dans la ZIF-8.

Tout d'abord, on peut remarquer que les molécules d'eau entre dans la ZIF-8 sans la moindre barrière. Pour les anions chlorure, nous observons deux barrières sur le profil d'énergie libre, qui correspondent aux fenêtres de la ZIF-8 à 0 et 15 Å. Ces barrières sont corrélées à un nombre inférieur de voisins pour l'ion : l'anion doit se désolvater partiellement pour passer à travers la fenêtre, ce qui explique la barrière d'énergie libre. En dehors de ces barrières, le profil est plat et au même niveau que dans le liquide pur : les ions Cl^- ont une barrière cinétique à l'entrée dans le ZIF-8, mais pas de barrière thermodynamique. Les résultats pour Li sont plus surprenants. Nous voyons à la fois une barrière élevée à la première ($x = 0$ Å) et à la deuxième ($x = 15$ Å) fenêtre ; et une différence énergétique entre l'extérieur et l'intérieur des pores valant approximativement 15 kcal/mol. Cette différence d'énergie n'est pas seulement due à la transition entre le liquide pur et le liquide confiné, car elle est également présente au niveau de la deuxième fenêtre. En même temps, ces barrières ne sont pas liées à une différence de solvatation comme dans le cas du chlore. Ceci indique une différence de nature entre les ions Li^+ et Cl^-, qui devront être approfondis.

8 Adsorption d'eau dans les imogolites

J'ai aussi étudié l'adsorption d'eau dans un matériau hydrophile : l'imogolite. Il s'agit de nanotubes d'aluminosilicate dont la structure est représentée en figure 5.12. J'ai pour cela collaboré avec Laura Scalfi durant son stage de master. Nous avons utilisé des simulations Monte-Carlo grand canonique et la dynamique moléculaire classique, et étudié la structure et la dynamique de l'eau adsorbée. Je vais me concentrer ici sur la structure de cette phase liquide.

Pour caractériser cette structure, nous avons calculé les profils de densité de tous les types d'atomes, présenté en figure 5.15 dans le plan xy, comme si elles étaient vues du haut du nanotube (infini). En regardant distribution de l'oxygène de l'eau O_w, il est clair qu'il existe deux populations différentes de molécules d'eau : de l'eau fortement structurée adsorbée près de la surface interne et de l'eau plus désordonnée au centre du nanotube. On observe également que les nanotubes sont légèrement déformés par rapport à tube circulaire : les atomes de silicium sont placés dans une disposition hexagonale. Cela est due à la symétrie de l'empilement des nanotubes en faisceaux[197]. Cette déformation est très faible, avec des déplacements de Si de 0,1 à 0.2 Å tout au plus, les nanotubes d'imogolite étant très rigides.

Une autre manière de visualiser l'adsorption est présentée en figure R9. Ici, la densité des atomes est représentée par rapport à z et une coordonnée circulaire, comme si une tranche du nanotube avait été coupée et déroulée. On peut imaginer la surface interne des nanotubes comme une séquence périodique de triangles, où chaque sommet est un

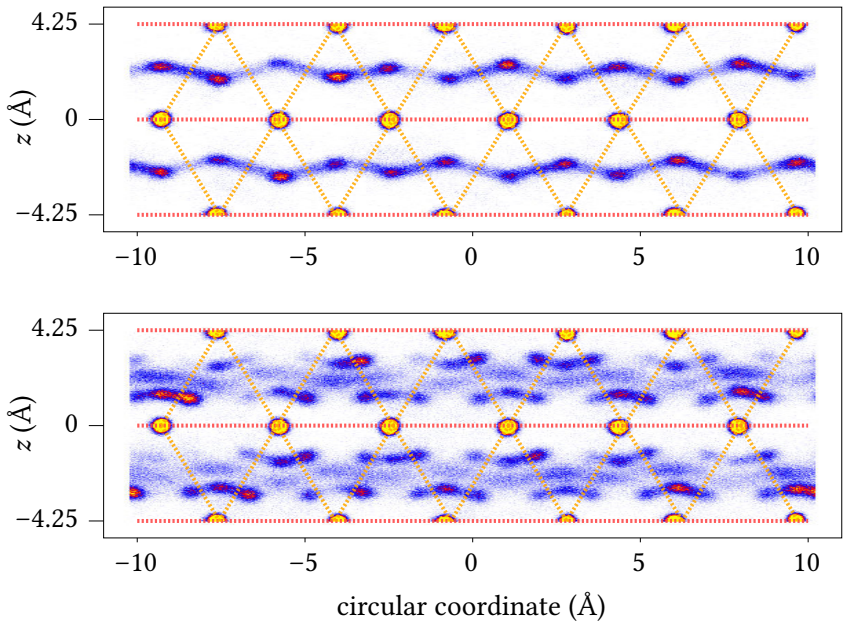


FIGURE R9 – Profils de densité sur les nanotubes "déroulés" pour l'oxygène de l'eau (haut), et l'hydrogène de l'eau (bas). La coordonnée circulaire correspond à une abscisse curviligne qui trace un cercle de rayon R = 6.5 Å *centré sur l'axe z. Sur tous les graphiques, les atomes d'oxygène internes apparaissent sous forme de points jaunes.*

groupe silanol SiOH. Sur la figure R9, les lignes pointillées rouges indiquent les anneaux SiOH, tandis que les lignes pointillées orange indiquent ce maillage triangulaire. Ces triangles sont isocèles avec deux angles de 66,5° et un angle de 47°. Au-dessus du centre de chaque triangle se trouve un site potentiel d'adsorption d'eau. Cependant, l'analyse des sites montre qu'aucun site voisin dans le même plan *xy* ne peut être occupé en même temps, étant donné qu'ils sont trop proches les uns des autres. Au plus la moitié des sites d'adsorption seront donc occupés dans le nanotube rempli d'eau. Nous pouvons aussi observer sur cette figure qu'il y a une forte anisotropie du système. Là où les molécules d'eau peuvent se déplacer dans un mouvement circulaire entre les sites dans le même plan *z*, il n'y a aucune molécule passant d'un plan à l'autre. Il y a diffusion en surface des molécules d'eau, mais seulement dans le plan *xy*. Ceci est corroboré par les coefficients de diffusions, présentés en figure 5.21.

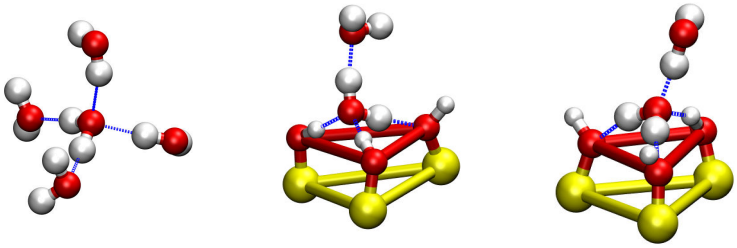


Figure R10 – Représentation des motifs de liaison hydrogène les plus courants pour l'eau dans un nanotube d'imogolite. De gauche à droite, ce sont les motifs 1, 2 et 3. Pour la numérotation des motifs, voir tableau 5.7.

Afin de caractériser le réseau des liaisons hydrogène, nous avons classifié les liaisons hydrogène en fonction du nombre de liaisons donnée et reçu avec d'autre molécules d'eau et les silanols de surface. Les motifs les plus courants sont représentés sur la figure R10. Pour les molécules adsorbées, le motif le plus courant est une molécule d'eau "verticale" (motif 2). Un groupe hydroxyle de l'eau est orienté vers le centre du nanotube, structurant l'eau adsorbée au-delà de la première couche. Le deuxième motif le plus courant est une molécule d'eau située "à plat" sur la surface interne de l'imogolite. Dans ce cas, la direction de la liaison hydrogène est moins contrainte, ce qui explique la partie "diffuse" de la densité d'oxygène dans la deuxième couche d'eau.

Nous avons montré que les interactions entre les molécules d'eau adsorbées sur des sites voisins sont complexes, basées sur (1) l'exclusion d'un voisin le plus proche, et (2) le nombre de liaisons hydrogène données (et acceptées) par les molécules d'eau aux groupes silanol voisins. Maximiser le nombre de liaisons hydrogène de ces règles crée de la frustration dans le système, ce qui conduit à l'émergence d'un état fortement désordonné. Cette situation est très semblable aux règles de Bernal-Fowler qui décrivent l'orientation des molécules d'eau dans la glace[209].

Enfin, nous avons également caractérisé la dynamique des molécules d'eau confinées. Nous avons constaté qu'en plus de l'effet générique du ralentissement de l'eau confinée, les fortes interactions des molécules d'eau avec les groupes silanol donnent lieu à des constantes de temps comparables aux temps de séjour typiques de l'eau dans les sites d'adsorption des protéines.

9 Simulations hybride dans l'ensemble osmotique

Le Monte-Carlo hybride est une amélioration de l'algorithme standard de Monte-Carlo Metropolis qui permet de simuler plus efficacement des systèmes complexes. Cette méthode est basée sur l'utilisation de courte simulations de dynamique moléculaire pour générer une nouvelle conformation dans la chaîne de Markov de la simulation Monte-Carlo. Pour pouvoir simuler l'adsorption dans des matériaux poreux flexibles, il faut pouvoir échantillonner les mouvements collectifs du matériau. Les techniques Monte-Carlo standard ne sont pas adapté à cet effet, car elles doivent déplacer les atomes du système un par un pour être efficaces. La dynamique moléculaire est capable de simuler les mouvements collectifs, mais ne peut pas être utilisé pour des systèmes ouverts. Le Monte-Carlo hybride permet d'améliorer l'efficacité de la simulation en incorporant des informations sur la courbure locale de la surface d'énergie potentielle.

J'ai implémenté la méthode Monte-Carlo hybride dans le logiciel de simulation moléculaire Domino développé au sein du groupe. J'ai utilisé ce code pour une étude préliminaire de l'adsorption du méthane CH_4 dans un modèle simplifié, sans charges, de la MOF-5 à 300 K. Le calcul des interactions électrostatiques n'était pas pris en charge par Domino. Bien que ce modèle ne soit pas une représentation précise de la MOF-5, c'est un test intéressant pour l'utilisation de simulations Monte-Carlo hybrides. L'isotherme prédite par cette simulation est présentée en figure R11

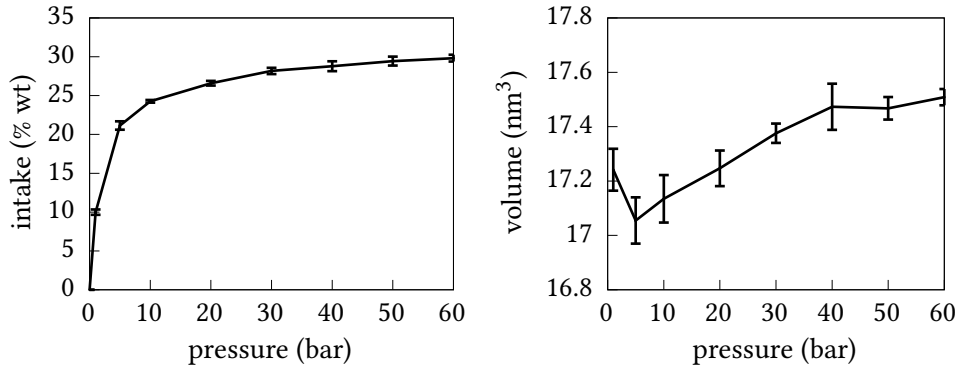


Figure R11 – Résultats de simulations Monte-Carlo hybride dans l'ensemble osmotique pour l'adsorption du méthane dans un modèle de MOF-5 simplifié. (gauche) isotherme d'adsorption à 300 K*, (droite) changement de volume pendant l'adsorption.*

Cette isotherme est de type I, comme prévu pour l'adsorption du méthane dans la MOF-5. Contrairement aux simulations GCMC standard, cette isotherme intègre des effets de flexibilité. Il est plus intéressant d'observer un comportement non monotone dans les variations du volume en fonction de la pression. Nous observons tout d'abord une faible contraction de la maille à basse pression, avant l'expansion attendue à haute pression. Ce comportement de contraction-expansion fait penser aux déformations induite par sorption dans d'autres matériaux poreux[208, 236] : la présence de quelques molécules à l'intérieur des pores induit une contraction de tout le système. Une autre façon de voir ce phénomène est d'imaginer les molécules à l'intérieur du pore *tirant* sur les parois. J'ai donc pu vérifier que la simulation directe des déformations pendant l'adsorption était possible grâce aux méthodes de Monte-Carlo hybride.

Conclusions

Les travaux présentés dans cette thèse portent sur l'étude de l'adsorption et de l'intrusion dans les matériaux flexibles nanoporeux, la déformation de ces matériaux et le couplage entre les deux phénomènes. Le confinement d'un fluide à l'intérieur d'un réseau poreux a des effets significatifs sur ses propriétés thermodynamiques, du fait de la compétition entre les interactions avec les interfaces et les interactions avec le fluide même. Cette compétition génère de nouveaux comportements, tels que de nouvelles phases fluides et des transitions entre ces phases, et est particulièrement présente dans les matériaux nanoporeux, où la largeur typique du pore et la distance typique des interactions sont du même ordre de grandeur. D'autre part, la présence d'un fluide confiné peut également avoir des effets importants sur le solide environnant, créant de nouvelles phases et modifiant l'équilibre entre plusieurs phases méta-stables. C'est particulièrement poignant dans le cas des matériaux nanoporeux flexibles, tels que de nombreux MOFs.

Comme ces matériaux sont relativement récents, leur flexibilité a souvent été négligée et ce n'est que ces dernières années que la communauté scientifique a commencé à en tenir compte. Un exemple d'un tel changement est présenté dans la première section de ce résumé, avec l'incorporation de l'ensemble osmotique dans la théorie de la solution adsorbée idéale (IAST) pour l'étude de la co-adsorption des gaz, menant à la création de la théorie dite *Osmotic Framework Adsorbed Solution Theory* (OFAST). J'ai pu démontrer que la méthode IAST est invalide *par construction* pour le traitement de la co-adsorption lorsque l'hôte adsorbant n'est pas inerte pendant l'adsorption. En particulier, j'ai montré que IAST ne peut pas être utilisé pour la prédiction de la co-adsorption de mélanges de fluides dans des matériaux présentant un comportement d'ouverture de porte, et qu'il prédit une sélectivité non physique, jusqu'à deux ordres de grandeur supérieurs à celle prédite par OFAST. Même lorsque IAST n'est pas explicitement utilisé pour calculer la sélectivité dans matériaux flexibles, il faut rester prudents en lorsque l'on compare des isothermes de corps purs en présence de flexibilité. Les différences de pression d'ouverture des isothermes peuvent conduire à des allégations de forte sélectivité, lorsque l'on applique des concepts qui ne sont valables que pour des matrices hôtes rigides.

Il faut aussi veiller à ne pas aller trop loin dans l'autre sens, et attribuer tous les comportements observés à la flexibilité des matériaux. Dans la quatrième section de ce résumé, j'ai utilisé des simulations de dynamique moléculaire *ab initio* pour expliquer l'origine d'une isotherme étagé pour l'adsorption d'azote dans ZIF-8(CH_3) et ZIF-8(Cl), et son absence dans la structure similaire ZIF-8(Br). J'ai montré que si le réseau se déforme pendant l'adsorption pour ZIF-8(CH_3) et ZIF-8(Cl), les déformations ne changent pas le volume accessible et la distribution des pores de ces matériaux. Au lieu de cela, l'augmentation de l'absorption dans l'isotherme est liée à une réorganisation du fluide confiné dans les pores, réorganisation qui ne se produit pas dans ZIF-8(Br) en raison de la différence de taille des pores. Il est donc fondamental de tenir compte à la fois des effets de flexibilité et de confinement lors de l'étude de l'adsorption dans les matériaux nanoporeux flexibles.

Il en va de même pour l'intrusion, cousin de l'adsorption. Dans les sections 4 et 5, j'ai utilisé des simulations moléculaires classiques pour étudier le confinement sous haute pression de l'eau et de solutions d'électrolytes dans des nanotubes d'imogolite et dans la ZIF-8. J'ai

observé des effets de confinement allant d'une organisation spatiale plus marquée, à des changements dans les propriétés élastiques, et le ralentissement de la dynamique de l'eau. Il est intéressant de noter que la présence d'ions à de forte concentrations peut avoir les mêmes effets sur l'eau non confinée ; structurant le réseau de liaisons hydrogène et ralentissant la dynamique. L'intrusion de solution aqueuses dans un matériau hydrophobe est un moyen prometteur de stocker et de dissiper l'énergie mécanique. Il est possible d'ajuster le comportement et même de transformer un système stockant de l'énergie à un système dissipatif d'énergie à la dissipation en ajoutant des ions dans la solution aqueuse utilisée pour l'intrusion. J'ai examiné l'impact des ions sur le comportement d'intrusion en utilisant des simulations *umbrella sampling* pour extraire le profil d'énergie libre d'entrée dans la ZIF-8, montrant que les différents ions ont des barrières différentes quand ils traversent les fenêtres de la ZIF-8. Cette étude est l'une des premières sur le thème de l'intrusion d'électrolytes dans les MOF, et a permis de mettre en lumière les comportements complexes qui émergent dans ces systèmes.

La nécessité de tenir compte simultanément de l'adsorption et des déformations a été un thème récurrent de toutes ces études. Mais les méthodes actuelles de simulation ne permettent d'aborder qu'une seule dimension du problème : les simulations de dynamique moléculaire peuvent décrire des déformations, mais la modélisation de systèmes ouverts et donc l'adsorption n'est pas possible. Les simulations de Monte-Carlo Metropolis peuvent être utilisées pour des systèmes ouverts, mais elles ont du mal à échantillonner efficacement les déformations collectives. Les simulations Monte-Carlo hybride sont une réponse possible à ce dilemme, combinant l'efficacité de la dynamique moléculaire avec la polyvalence des simulations Monte-Carlo (en particulier la possibilité d'échantillonner des ensembles ouverts). La dernière section présente la méthode de simulation Monte-Carlo hybride et son utilisation pour les simulations directes dans l'ensemble osmotique.

Il y a une autre condition à remplir avant de pouvoir utiliser largement les simulations d'ensembles osmotiques pour l'étude de l'adsorption et de l'intrusion dans les cristaux nanoporeux flexibles : il nous faut pouvoir calculer l'énergie liés à la flexibilité des matériaux et à leurs interactions avec les fluides. Les méthodes *ab initio* (comme la théorie de la fonctionnelle de la densité) permettent de calculer avec précision l'énergie de systèmes atomistiques arbitraires. Ces méthodes nécessitent malheureusement une grande puissance de calcul, ce qui les empêche d'être utilisés en routine sur de grands systèmes. Face à des systèmes d'une telle taille — que ce soit en termes de nombre d'atomes, d'échelle de temps des processus ou de criblage à haut débit — nous avons donc souvent recours aux champs de force classiques.

Les champs de force classiques sont soit *précis*, c'est-à-dire qu'ils reproduisent bien la surface d'énergie potentielle réelle, soit *transférables*, c'est-à-dire utilisables avec plusieurs systèmes différents. Les champs de forces transférables actuels ne sont pas bien adaptés pour décrire la flexibilité résultant des liaisons de coordination, il faut donc créer de nouveaux champs de forces pour ces systèmes. Historiquement, la paramétrisation de nouveaux champs de forces a été un processus assez long et fastidieux. Depuis quelques années, de nouvelles techniques basées sur l'apprentissage statistique ou *Machine Learning* ont été mises au point pour la dérivation constante et rapide de champs de force précis. Je présente une de ces techniques dans la troisième section de ce résumé, que j'ai utilisée pour obtenir des champs de force pour la ZIF-8 et certains de ses dérivés à partir de données

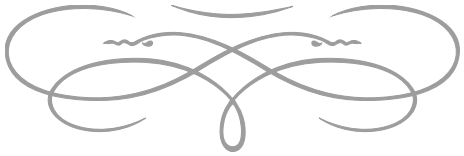
ab initio. Ces techniques automatiques sont particulièrement cruciales pour l'étude des MOFs en raison de la grande diversité de leurs structures. J'espère que la disponibilité de champs de force précis et de logiciels offrant des simulations de Monte-Carlo hybride facilitera l'utilisation de simulations moléculaires pour concevoir de nouveaux matériaux adaptés à des applications spécifiques.

Ces travaux ouvrent des perspectives dans plusieurs directions. En ce qui concerne les méthodes de simulation moléculaire, le Monte-Carlo hybride semble être une technique puissante, utilisable avec une grande variété de systèmes. Tout d'abord, la méthode Monte-Carlo hybride étant basée sur la théorie des chaines de Markov et le critère de Metropolis convergera toujours vers la distribution dans l'espace des phases de l'ensemble statistique voulu. Au contraire, la dynamique moléculaire échantillonne par défaut l'ensemble micro-canonique, et doit s'appuyer sur des thermostats et des barostats pour échantillonner d'autres ensembles. Ces thermostats et surtout les barostats ne sont pas tous égaux, et seuls certains algorithmes sont capables de générer précisément l'ensemble souhaité. En même temps, les mouvements hybrides améliorent grandement l'efficacité des simulations Monte-Carlo en tenant compte de la courbure locale de la surface d'énergie potentielle.

Savoir s'il est possible de simuler des systèmes ouverts avec la dynamique moléculaire reste aujourd'hui encore une question de recherche ouverte. Inversement, de telles simulations sont couramment réalisées dans le cadre du Monte-Carlo Grand Canonique. Le Monte-Carlo hybride pourrait ainsi être utilisé pour la simulation d'ensembles ouverts et de systèmes dilués, tels que les simulations à pH constant, la description de l'environnement ionique des protéines ou la simulation de défauts dans les matériaux cristallins ; améliorant l'efficacité de ces simulations par rapport au Monte-Carlo classique et permettant la simulation d'ensemble ouverts aux utilisateurs de dynamique moléculaire. Grâce au critère d'acceptation de Metropolis, il n'est pas nécessaire que la dynamique moléculaire courte utilisée par les mouvements hybrides échantillonne un ensemble thermodynamique réel ou même un Hamiltonien ayant un sens physique. Cette propriété pourrait être exploitée pour créer des simulations Monte-Carlo encore plus efficaces, par exemple en insérant progressivement de nouvelles molécules dans un système tout en relaxant son environnement.

Une autre perspective concerne les champs de forces classiques et leur capacité à reproduire avec précision les surfaces énergétiques potentielles. L'approche traditionnelle lors de la création de champs de force a été de décomposer l'énergie en une somme de termes dépendant de valeurs scalaires simples avec une signification physique : distances, longueur de liaison, angles et angles dièdres, *etc.* Ces valeurs scalaires sont ensuite combinées dans des expressions mathématiques simples, par exemple des fonctions puissance ou exponentielles. Dans cette approche la surface d'énergie potentielle ne peut pas inclure d'effets à trois corps ou plus, ni reproduire avec précision la forme de la surfaces d'énergie potentielle de référence. Les outils d'apprentissage statistique, en particulier les réseaux neuronaux et les processus gaussiens, peuvent apporter des améliorations à ces deux problèmes. Premièrement, la capacité des réseaux de neurones à reproduire des fonctions arbitraires de $\mathbb{R}^n$ à $\mathbb{R}$ peut réduire l'écart entre les surfaces d'énergie po-

tentielle de référence et du champ de force. Par exemple, au lieu d'imposer un potentiel de Lennard-Jones, les réseaux neuronaux peuvent reproduire les variations exactes de l'énergie. Deuxièmement, les algorithmes d'apprentissage statistique peuvent être couplés à de meilleurs descripteurs de la structure atomique, tenant compte des effets à plusieurs corps. Dans dernières années, de multiples équipes scientifiques indépendantes ont travaillé à la conception, l'entrainement et l'évaluation de ces champs de forces basé sur l'apprentissage statistique et des descripteurs associés. À ma connaissance, ces méthodes n'ont pas encore été utilisées pour l'étude de matériaux nanoporeux flexibles.



RÉSUMÉ

Durant ma thèse, j'ai utilisé la simulation moléculaire pour étudier l'adsorption et l'intrusion de fluides dans les matériaux nanoporeux flexibles. Aujourd'hui, les matériaux à charpente organo-métallique appelés *metal-organic frameworks* (MOF) sont les principaux représentants de cette famille de matériaux. Je me suis en particulier intéressé à la ZIF-8, un MOF constitué de zinc et de ligands imidazolates organisés dans une topologie de type sodalite. Grâce à la dynamique moléculaire quantique, j'ai pu montrer que lors de l'adsorption de diazote dans la ZIF-8 il se produit une réorganisation de la phase adsorbée qui augmente la quantité totale d'azote adsorbée. J'ai aussi montré que changer la nature chimique des ligands permettait de supprimer partiellement ou totalement cette réorganisation.

D'autre part, j'ai utilisé la dynamique moléculaire classique et les simulations Monte-Carlo pour étudier l'adsorption et l'intrusion d'eau dans des matériaux poreux hydrophobes. Ces matériaux ont des applications potentielles dans le domaine du stockage et de la dissipation de l'énergie mécanique. La pression à laquelle se produit l'intrusion, ainsi que la présence et la forme d'une boucle d'hystérèse sont modifiable par l'ajout d'ions dans le liquide d'intrusion. J'ai montré que liquide confiné dans la ZIF-8 ou dans des nanotubes d'alumino-silicates appelés imogolites est fortement structuré, et que la dynamique des molécules d'eau est ralentie par le confinement. La présence d'ions modifie très peu la structuration, mais ralenti encore la dynamique, et rigidifie l'ensemble du système. J'ai aussi étudié l'entrée d'ions dans la ZIF-8, et observé une différence flagrante entre Li^+ et Cl^- d'un point de vue thermodynamique et cinétique.

Enfin, j'ai montré que la prise en compte de la flexibilité était nécessaire pour prédire correctement la co-adsorption de gaz dans un matériau qui se déforme (*respiration, ouverture des fenêtres*, *etc.*) lors de l'adsorption. Cette prise en compte est possible dans le cadre de la méthode *Osmotic Framework Adsorbed Solution Theory* (OFAST) pour décrire des changements de phase du matériau.

MOTS CLÉS

simulation moléculaire, matériaux nanoporeux flexibles, intrusion, adsorption, simulations multi-échelles

ABSTRACT

During my PhD, I used molecular simulation to study the adsorption and intrusion of fluids in flexible nanoporous materials. As of today, metal-organic frameworks (MOF) are the main example of this family of materials. I specifically worked with ZIF-8, a MOF built with zinc metallic centers and imidazolates linkers, organized in a sodalite topology. Using *ab initio* molecular dynamics I showed that nitrogen undergoes a reorganization inside the pores during adsorption; increasing the total adsorbed amount. I also showed that changes to the chemical nature of linkers allows to partially or completely remove this reorganization.

On an other side, I used classical molecular dynamics and Monte Carlo simulations to study adsorption and intrusion of water in hydrophobic porous materials. These materials have possible applications in mechanical energy storage and dissipation. The intrusion pressure, as well as the presence and shape of an hysteresis loop, can be tuned by adding ions in the intrusion liquid. I showed that the liquid confined in ZIF-8 or in alumino-silicate nanotubes called imogolites is strongly structured; and that the water molecules dynamic slows down under confinement. The presence of ions almost does not modify the water structuration, but slows down dynamics even more, and makes the whole system more rigid. I also studied ions entry in ZIF-8 structure, and observed a clear difference between Li^+ and Cl^- both on a thermodynamic and kinetic point of view.

Finally, I showed that it is necessary to take in account flexibility to correctly predict gas co-adsorption in frameworks that undergoes deformation (*breathing, gate-opening, etc.*) under adsorption. This is possible within the scope of Osmotic Framework Adsorbed Solution Theory (OFAST) for materials undergoing phase transition.

KEYWORDS

molecular simulation, flexible porous materials, intrusion, adsorption, multi-scale simulations