# Étude théorique des matériaux adaptatifs

Maxime Ducamp

PSL
UNIVERSITÉ PARIS

**THÈSE DE DOCTORAT**
**DE L'UNIVERSITÉ PSL**

Préparée à Chimie ParisTech

# Étude théorique des matériaux adaptatifs


Soutenue par
**Maxime DUCAMP**
Le 7 Juillet 2022


École doctorale n°388
**Chimie Physique et Chimie Analytique de Paris-Centre**

Spécialité
**Chimie Physique**

Composition du jury :

| | |
|---|---|
| Caroline MELLOT-DRAZNIEKS<br>Directrice de recherche CNRS<br>Collège de France | *Présidente du jury* |
| Céline MERLET<br>Chargée de recherche CNRS<br>Université Paul Sabatier | *Rapporteur* |
| Benoit COASNE<br>Directeur de recherche CNRS<br>Université Grenoble Alpes | *Rapporteur* |
| Michele CERIOTTI<br>Associate professor<br>EPFL | *Examinateur* |
| François-Xavier COUDERT<br>Directeur de recherche CNRS<br>Chimie ParisTech | *Directeur de thèse* |

Chimie Paris ParisTech | PSL




UNIVERSITÉ PSL

DOCTORAL THESIS

# Computational study of responsive materials

*Author:*
Maxime DUCAMP

*Supervisor:*
Dr. François-Xavier COUDERT

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

Institut de Recherche de Chimie Paris
Chimie Organométallique et Catalyse de Polymérisation

July 20, 2022

*"We are not lost. We are locationally challenged."*

Jhon M. Ford

# Contents

# List of Figures

# List of Tables

# Introduction

Nanoporous materials with high specific surface area have been of great interest for a long time. They are used in a wide variety of applications such as gas storage, catalysis, sensing or drug delivery. The last 15 years have seen the birth of new classes of crystalline nanoporous materials based on weaker bonds. These new classes include metal-organic frameworks (MOFs), covalent organic frameworks and porous molecular organic solids. Among them, a new class of material started to emerge, named "stimuli-responsive materials" or "soft porous crystals", which exhibit large or anomalous responses to external physical or chemical stimulation. The modification of the framework structure induced by the response also involves, in turn, a modification of other physical and chemical properties, making such materials multifunctional. Stimuli-reactive materials exhibit then a large variety of responsive behaviors: negative adsorption, negative linear compressibility, photoresponsive frameworks or negative thermal expansion to name only a few. Each of these properties can be used in different applications such as sensors or actuators, or engineering composite materials with targeted mechanical and thermal properties for example.

Despite their interesting behaviors, the development of applications for them is slowed down due to their complex synthesis. Indeed there is no systematic method to synthesize such materials. While the number of stimuli-responsive materials is growing over time, most of them are found by serendipity and the majority have never been characterized or tested under simulation. The full characterization of a soft porous material, whether experimentally or theoretically, is a consequent task requiring enormous resources. Usually, once a material is discovered, it is tweaked to obtain derivative materials with satisfying properties. This current way of generating materials is not efficient and needs to be improved. For this, the systematic study of the properties of soft porous materials is necessary to understand the occurrence of such properties. This can also highlight feasibility criteria guiding the chemical synthesis. The current thesis was built around this idea.

During my PhD, I conducted studies on two families of materials that have both potentials in industrial applications: zeolites, that are already used in the industry and are present in a lot of applications of everyday life, and metal-organic frameworks (MOFs), which have a great potential for adsorption-based applications and for which industrial processes are currently developed.

Zeolite is a family of natural aluminosilicates discovered in the $18^{th}$ century that have been widely studied until now. The first experimental synthesis of a zeolite goes back to the 1940s, and since then, more and more frameworks have been obtained experimentally. Due to their porous nature, zeolites have been extensively studied for their properties of adsorption and their potential use as catalysts and molecular sieves. They are present in numerous fields of our life such as agriculture or pollution control. Although their adsorption was thoroughly studied this past 50 years, their other features such as mechanical or thermal properties have drawn

only a little interest. However, zeolites have recently caught the attention due to their specific characteristics, especially in the computational field. Indeed, owing to their particular structural arrangements, zeolites exhibit fascinating mechanical properties that are considered "anomalous", such as a negative linear compressibility or a negative Poisson's ratio (a behavior also called auxeticity). These mechanical aspects of zeolites have been studied in details theoretically through various methods including DFT and machine learning. Thermal properties, on the opposite, have not received the same attention. They nevertheless have "anomalous" behavior in zeolites too as most of the frameworks exhibit a negative thermal expansion, to the point where it is considered a common behavior in this family of material. Very few studies were directed towards the thermal properties of zeolites and the need for large scale study becomes more and more important. Using databases of known frameworks of zeolites, I investigated the thermal behavior of a large number of pure silica structures of zeolites through DFT-based methods. The aim was to give insights into the "anomalous" phenomena occurring in these systems as well as to build a database for further calculations. The database of DFT-calculated zeolites was then used and study as a potential training set for machine learning purposes. I investigated the feasibility of using such techniques along with the database I created and the applicability for the prediction of properties of zeolites.

MOFs are a recent family of materials first discovered around the 90s which have since then drawn a large interest due to their industrial potential. MOFs are materials composed of inorganic units linked together by organic linkers. Their particular arrangement leads usually to a high porosity, most often higher than zeolites. Their structural diversity along with the variety of properties they can offer allows for a lot of possibility of applications. Due to their high porosity nature they are particularly studied for adsorption-based applications. Unlike zeolites which are rigid materials constituted of Si–O bonds, MOFs are built with a combination of strong and weak bonds, making them much more flexible and leading sometimes to stimuli-responsive behaviors of large amplitude. However, their flexibility also constitutes an obstacle to their industrial applications. Indeed, MOFs are known to have a relatively low thermal and chemical stability. The lack of knowledge on their stability and degradation over time prevent the development of numerous applications. Countless studies are however realised each year to fill this need and industrial uses of MOFs are starting to appear. In my thesis, I had the opportunity to work on several MOFs in different projects in collaboration with experimentalists. The first one dealt with the flexibility and opening process of a Co-based MOFs upon adsorption. The second focused on the different compression behaviors of 3 MOFs with same structures and different compositions. Finally, in the last one, I gave atomistic insights into the melting process of a MOF.

This thesis is organised in 4 chapters. The first chapter will introduce the materials I chose to focus on, namely zeolites and MOFs. It will then present the principles of DFT simulations and molecular dynamics. The second chapter will deal with the systematic study of thermal properties of zeolites. The first part of this chapter is dedicated to the introduction of thermal and mechanical properties and will present the methods used to calculate them. The second part contains the results of the study. The third chapter concerns the machine learning study of thermal properties of zeolites, which uses the results obtained from the second chapter. The first part is focused on the machine learning technique and its principles and the second part will present the results. The last chapter contains 3 different projects realised in

collaboration with experimentalists from foreign laboratories. The chapter is organised in 3 parts, each corresponding to a project. Each part will introduce the project by showing some results obtained experimentally and then will present the results obtained with simulations.

# Chapter 1

# Materials and Computational Methods

In this part I will describe the theoretical background needed to apprehend the works done in this thesis. The first part will deal with the materials studied and the second part will focus on the methodologies and fundamentals of computational chemistry.

## 1.1 Materials

During my PhD, I had the opportunity to investigate different materials for different projects, all related to the abnormal behaviors that can occur in the solid state. Despite having studied these materials for different purposes, they all share the same similarities: all of them are crystalline porous materials. Crystalline materials are composed of periodic arrangements of atoms in three dimensions. The key feature of such structure is the periodicity, which allows to describe large systems with only a fraction of it. This small fraction is called the unit cell and is the pattern used to describe the material. This representation is convenient especially for simulations as only a small portion of the structure has to be computed. The term porous designates the presence of void in the structure. Indeed, depending on the atomic arrangement (periodic or not), cavities called "pores" can appear inside the structure. The materials exhibiting this kind of feature are usually interesting in adsorption related fields or for applications as molecular sieves for example. The size of pores can vary a lot and is strongly depending on the structure of the material.

Within this category of crystalline porous materials, I studied two families, zeolites and metal organic frameworks, for which I will describe the main characteristics in this part

### 1.1.1 Zeolites

The history of zeolites started in the $18^{th}$ century when a Swedish chemist named Axel Fredrik Crönstedt discovered the first material of this kind, the mineral called stilbite (represented on fig. 1.1). While studying the thermal stability of this mineral, he observed that, when heated, moisture and bubbles would appear on its surface, giving off the impression that the stone was boiling[2]. He thus called this family of materials zeolites, associating the Greek terms *zeo* (to boil) and *lithos* (stone). This discovery set the path for all the following works on this type of material. In the following decades, a dozen of natural zeolites were discovered and in 1862 the first synthetic zeolite called levinite was obtained, opening then fully the field of research

on zeolite materials. As of now, there are 244 different known frameworks for zeolites.

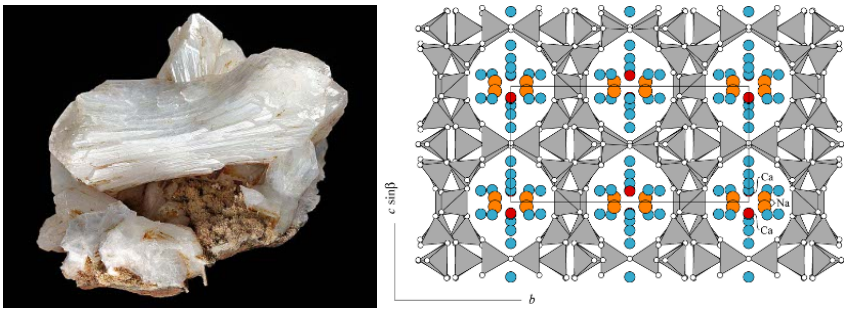


FIGURE 1.1: Image of the mineral stilbite (left panel) and its crystal structure (right panel). Taken from [3].

### Generalities

Zeolites are a type of three-dimensional aluminosilicates (also called tectosilicates) whose structure is built with corner-sharing tetrahedra. This particular 3D arrangement can lead to the presence of voids in the structure which is one of the most important features of zeolites. Pore size and accessibility are known to be the key characteristics to look for when investigating properties such as adsorption, or processes like ion-exchange. Due to their specific arrangement and to its variety throughout the different frameworks, the zeolite family presents a large diversity of pore sizes ranging from a few angströms to a few nanometers. They also exhibit a very large diversity of pore shape which can be, depending on the structure, cylindrical or spherical for example, and can also be inter-connected or not. This high versatility of porosity in the zeolite family has been the object of a great number of studies focusing on adsorption, ion-exchange and molecular sieves properties for industrial applications[4, 5, 6, 7].

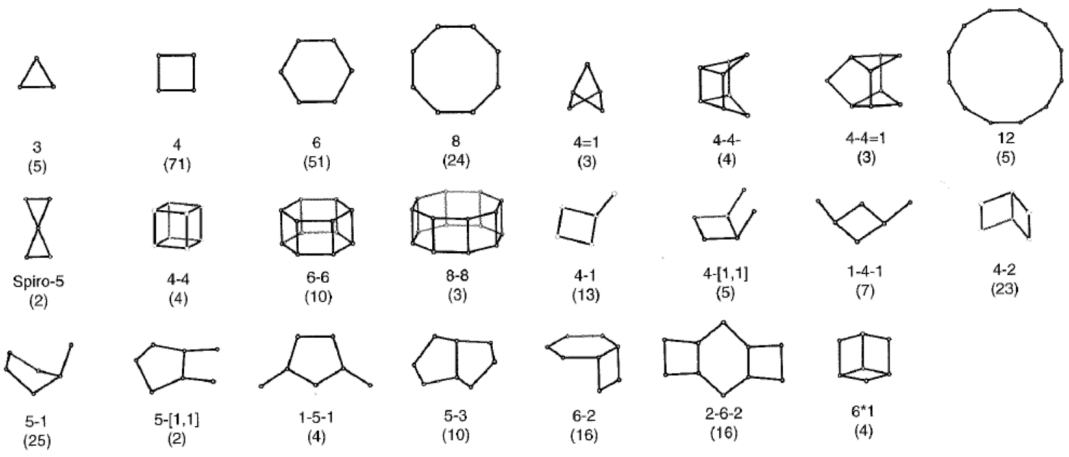


FIGURE 1.2: Secondary Building Units of zeolites. Vertices designate Si/Al atoms and edges correspond to oxygen atoms. Taken from [8]

Elements present in these materials are mostly aluminium Al, silicon Si and oxygen O with tetrahedral structures of $[AlO_4]^{5-}$ and $[SiO_4]^{4-}$. Each tetrahedron is bonded such that all oxygen atoms of a tetrahedron are shared with the corresponding adjacent tetrahedra. A pure silica structure $SiO_2$ leads to an overall neutral framework

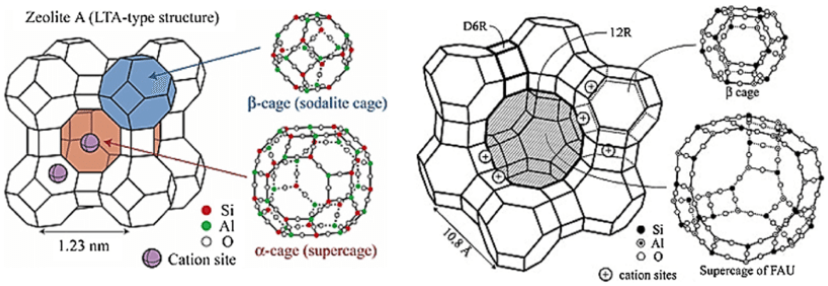


FIGURE 1.3: Representation of the LTA and FAU framework. Vertices represents Si/Al atoms and edges correspond to the shared oxygen atoms. Taken from [9].

(like the $\alpha$-quartz for example). However, swapping silicon with aluminium in the structure induces an excess of negative charges that is balanced by the presence of alkali or alkaline earth cations, such as sodium $Na^+$, potassium $K^+$ or calcium $Ca^{2+}$. These cations are located within the empty space of the structure, that is the pores or the channels.

To describe the framework of zeolite materials, a simple representation has been adopted: Si/Al tetrahedras of the structure are called the primary building units (PBU). Their connections give birth to simple geometrical shapes such as rings or cages that are called secondary building units (SBU). Some SBUs existing in the zeolite family are depicted in fig. 1.2 where we can observe rings of different sizes based on the number of tetrahedra composing them (the most common have four, five, six or eight). These rings as well as the pore sizes and shapes can also differ depending on the Si/Al ratio which causes an even more important diversity of frameworks. In theory, there is an infinite number of different topologies for zeolites. In order to classify them, a "three-letter name code" has been adopted to designate all the different existing frameworks. The codes are approved by the International Zeolite Association (IZA) which established as of now 244 different ordered frameworks. As examples, the structures of linde type A (LTA) and faujasite (FAU) are represented on fig. 1.3.

The general formula of an aluminosilicate is as follow: $M_{a/n}(AlO_2)_a(SiO_2)_b \cdot wH_2O$ where M is any alkali or alkaline earth atom and n is the charge of the atom M. This formula allows to clearly see the ratio $\frac{b}{a}$ indicating the fractions of Si and Al in the zeolite. This ratio is of great importance in this family of materials and is at the origin of many differences observed in their properties. Usual values of $\frac{b}{a}$ ratio are from 1 to 5[10, 11] but much larger values up to 100, although much rarer, can exist[12]. Some topologies of zeolites can also be synthesized in pure-silica form, with a $\frac{b}{a}$ ratio being infinite.

### Natural and synthetic zeolites

Natural zeolites can be found as crystals in cavities of basaltic rocks or as volcanic tuffs. They are formed in natural condition of pressure and temperature in various environments such as lake sediments, alkaline deserts or ash ponds. The general formula for natural zeolites is as follow:

$$(Li, Na, K)_p(Mg, Ca, Sr, Ba)_q[Al_{p+2q}Si_{n-(p+2q)}O_{2n}] \cdot wH_2O$$

The most favored applications for natural zeolites are environmental remediation, waste treatment and agriculture. As examples, clinoptilolite was used around 1960 for the immobilization of radioactive isotopes of cesium and is still used nowadays [13]. Cadmium, lead and nickel ions were successfully trapped using the natural zeolites faujasite and Phillipsite[14]. Hence natural zeolites still have applications often, however they are not suited for industrial applications. These materials are contaminated by other minerals such as quartz or other zeolites, and thus cannot be used in industry where the purity and uniformity is essential.

Compared to the natural zeolites, properties of synthetic ones are more uniform and controlled as they are designed by the synthesis. Raw materials used for synthesis are minerals or chemicals rich in silicon and/or aluminium. The Si/Al molar ratio defines the ratio of the final zeolite which can be anywhere between 0.5 to the infinity (pure-silica zeolite). Modulating the ratio can change the properties of the resulting material, which is a strong advantage for industrial applications. Several methods are possible for synthesizing zeolites and usually require high pressure and temperature. Examples include seed-assisted methods[15, 16] or template-based methods[17, 18].

Among all the topologies of zeolites approved by the IZA, there are both natural and synthetic ones. For example, Linde Type A (LTA) topology is not occuring naturally and can be obtained only synthetically. In contrary, Heulandite (HEU) is an abundant topology occurring naturally and being very difficult to synthesize. But there are also topologies that can occur naturally and can be synthesized such as the Chabazite (CHA) topology.

#### Properties of zeolites

The main interests of zeolites are for adsorption, ion exchange and molecular sieves applications, which means that their porosity is their main characteristic. Thus, properties such as the density are of great importance for this class of materials. Extensive studies are performed also on the morphology aiming to measure and control their pore sizes and shapes, their surface area and accessible volume[19, 20]. These properties of zeolites are strongly linked to the target applications. As mentioned earlier, the versatility of zeolite frameworks allow for a large variation of porosity throughout the landscape of possible zeolites, meaning that zeolites represent a candidate family for achieving materials with highly fine-tuned adsorption or ion exchange properties. On par with their porosity, the surface features of zeolites are important when it comes to applications as molecular sieves. Indeed the pores at the external surface are the "entry doors" for molecules to penetrate in the internal void space of the structure. The pore size and shape in the external surface is the key factor for restricting the access to the molecules of choice. Such "entries" to the internal void space can be tuned by organic functionalization.

Zeolites are in general quite robust and stable materials, due to their inorganic nature and strong Si–O and Al–O bonds. Indeed, although the increase in temperature causes a loss of adsorbed water, the structure usually remains stable at high temperature[21, 22]. This stability is influenced by several factors among which the crystallinity and the Si/Al ratio take a big place. It is known that a higher Si/Al ratio usually improves the thermal stability with high-silica zeolites being stable up

to almost 1000°C[23]. It is also interesting to note that pure-silica zeolites are chemically stable in all mineral acids except hydrofluoric acid, which makes them suitable candidates for corrosive-resistant coatings, replacing then the commonly used chromium-based compounds known to be highly toxic and carcinogenic[24].

Thermal and mechanical properties of zeolite will not be further discussed here as a dedicated part for them can be found later in chapter 2

### 1.1.2 Metal Organic Frameworks

Zeolites have been extensively studied since their discoveries as they were the most accessible and primary source of cristalline porous materials. Yet the accessible pore size ranges accessible remained limited to a certain extent for crystalline materials and pore diameters of 1 or 2 nm were not observed before the late 1980s and early 1990s[25, 26]. In 1995 however, Yaghi and coworkers discovered a brand new material, depicted in fig. 1.4 showing high porosity and surface area[27]. Such type of material was already observed by Hoskins and Robson in the 1990s[28]. They used at the time the term "infinite polymeric frameworks" to designate their discovery. Yaghi and his group continued to work on this type of materials and published a new MOF called MOF-5[29] which sparked the interest of the porous materials' community. As the structure was composed of inorganic units linked by organic linkers, they proposed for the first time the term "Metal Organic Frameworks" (MOFs) to refer to these type of materials. The keen interest in this new family showing a compromise between a good stability along with a way more important porosity than zeolites caused an exponential growth of the field with extensive study on their synthesis, stability and properties. As of today, thousands of works are published every year on this class of materials.

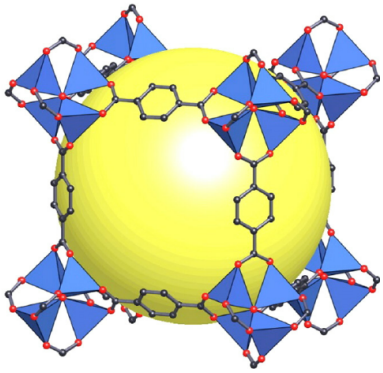

FIGURE 1.4: Representation of the MOF-5 framework. The yellow sphere represents the porous volume of the structure, blue tetrahedra designate the zinc tetrahedra, red spheres represents the oxygen atoms and black spheres corresponds to the carbon atoms.

#### Generalities

MOFs are crystalline materials combining inorganic building units linked together by organic linkers. Both inorganic units and organic linkers are commonly called Secondary Building Units (SBUs). Due to their nature, they often show a high porosity and surface area which makes them suitable candidates for applications in adsorption or catalysis, similar to the applications of zeolites. However, compared to

the latter, MOFs are less stable due to their bonds between organic linkers and inorganic parts being weaker. This result in a poorer thermal stability which can still be up to 500°C[30, 31] whereas zeolites can withstand temperatures up to 1000°C. The mechanical stability also suffer from the weaker bonds and from the very high porosity, which can in some cases pose a great difficulty for industrial applications, as MOFs are much softer than zeolites. However, MOFs are just starting to be employed in the industrial field and countless studies are done on this family of material, despite the stability issue. This is due to their extremely high flexibility and versatility which allows to build a vast variety of MOFs with different properties even within the same structure. It is also important to note that the MOFs field is pretty recent (around 30 years) so the room for innovation is still large; solutions to the challenges proposed by this class of materials are yet to be found.

In the following parts I will briefly introduce the synthesis of MOFs as well as their properties and applications.

#### Synthesis and functionalization

The most conventional way of synthesizing MOFs is the solvothermal route via a conventional electric heating. Inorganic sources and organic linkers are mixed together and heated to provoke the self-assembly of the MOF. The resulting material is then treated to empty the pores from all the residual solvent or impurities. This process is usually done by applying a vacuum or thermal treatment, and is crucial to "activate" the structure for its use in adsorption-related applications. The solvothermal synthesis is the preferred way in laboratory as it is easy to realize and is quite powerful to accelerate the discovery of new MOFs and to optimize synthesis protocols. However, since the series of MOF introduced by Yaghi and coworkers during the 90s, new ways of synthesizing MOFs appeared and constant development is made in this field. One example is the microwave-assisted synthesis. It consists in putting the mixture of raw materials in a sealed Teflon vessel and to heating it with a microwave unit. This method results in a fast heating of the liquid phase ensuring then a fast crystallization, and is adapted for industrial use as commercial microwave equipment are adapted to this method. This technique allowed to synthesize the well known HKUST-1 in crystals of 10 to 20$\mu$m within 1 hour and with a high yield of around 90%[32]. The process to synthesize MOF-5 was also optimized with this method resulting in a shorter synthesis time although the crystal quality decreased[33].

Another synthesis route is mechanosynthesis, which consists in grinding the raw materials at room-temperature in a solvent-free environment. The first mechanosynthesis was realised in 2006[34] and was quickly followed later by the synthesis of HKUST-1 and ZIFs[35, 36]. Advantages of this technique are the short reaction time (less than an hour) and the absence of solvent, which simplifies the activation process. Nevertheless, a small quantity of solvent can be added during the grinding so as to conduct a liquid-assisted grinding which further accelerates the reaction, the solvent allowing for more mobility of the reactants. Many other possibilities exist such as the electrochemical synthesis or the dry-gel conversion synthesis, which has been extensively applied to zeolites and zeolite membranes.

The key parameter in designing and synthesizing new MOFs is obviously the choice of organic linkers and SBUs. MOFs can be constructed from a very large variety of organic linkers and SBUs (as demonstrated in fig. 1.5) allowing then an extreme

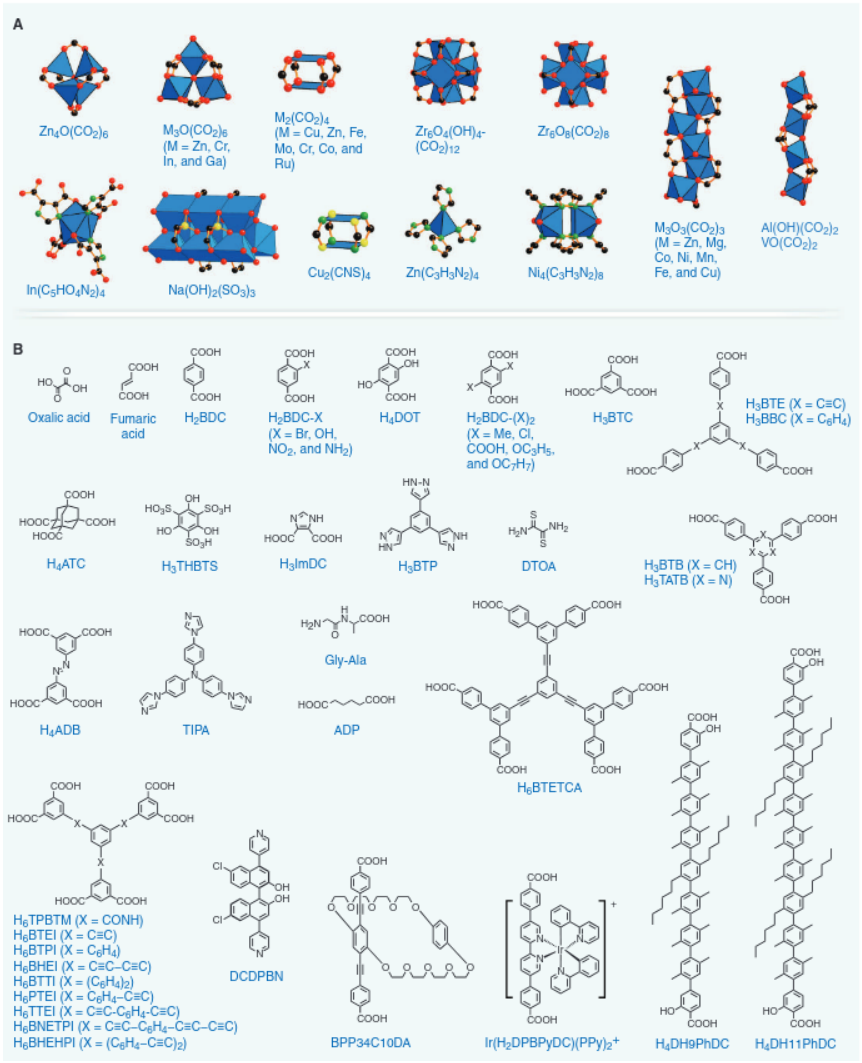


FIGURE 1.5: Representation of several possible (A) inorganic units and (B) organic linkers. Taken from [37].

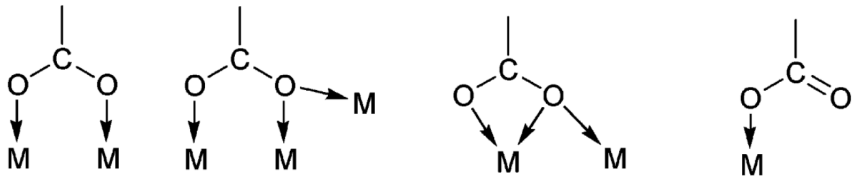


FIGURE 1.6: Representation of the different possible coordinations of a carboxylate linker with the metal node: (left) bidentate, (central schemes) bridging and (right) monodentate. Taken from [39].

diversity of possible frameworks. This is considered to be one of the most important advantage when using MOFs: their structures are easily customizable and their properties tunable. Usual organic linkers are polycarboxylates or polynitrogens which can both have different coordination configuration around the inorganic clusters, constituting the first possibility to create diversity. Those different coordination configurations are depicted in fig. 1.6. The second option is to use linkers of different sizes which creates so-called "iso-reticular" MOFs that share the same net but have different size of porosity. This was achieved for example with the well-known IR-MOFs family[38], which uses the same bidentate bridging carboxylate linkers. This family of MOFs is represented in fig. 1.7. The length of the linker was varied from the smallest one being IRMOF-1 (or MOF-5) with a pore size of around 4 Å to the longest one being IRMOF-16 which shows a pore size of around 29 Å. However, the length of linkers that can be used has a limit, as several problems arise when using too long chains. One such problem is interpenetration - which can be solved to a certain extent with the proper experimental conditions[38]. Another one is the stability of the resulting structure.

Thanks to the high versatility of MOFs, we can direct the synthesis of MOFs towards a particular application by designing a specific MOFs with suitable properties. This customization is further improved by the possibility to functionalize the organic linkers, creating then specific properties for the pores. The aim of that tool is to further specialize MOFs on a specific targeted property or feature and to optimize them for industrial applications. For example, catalysis can be improved by introducing specific groups to create catalytic sites[40, 41]. This is also true for adsorption for which the introduction of specific functional groups can improve the affinity of a MOFs towards certain molecules[42, 43]. Functionalization can happen before the synthesis by constructing MOFs with already functionalized linkers, or after the synthesis by chemical reaction on the linker. Functionalizing a linker can have several goals, one of them being the stabilization of the structure. The insertion of hydrophobic groups on linkers can protect the metal-linker bond from water and thus giving water stability to the MOF. Commonly used hydrophobic groups are fluoro, alkyl or aromatic groups which can all be incorporated before or after the synthesis. Examples of successful improvement of water stability after insertion of hydrohobic moieties include the MOF-5 framework which can be stabilized by adding trifluoromethoxy and methyl groups to the linkers[44, 45] and the structures based on paddlewheel units[46, 47, 48]. Moreover, addition of hydrophobic groups can still be useful in the case of an already water-compatible MOF, as it can further improve its hydrophobicity leading then to a better selectivity of adsorption in humid environments where water often competes with the targeted molecules.

Other methods aiming for a high optimization of the properties of MOFs have arisen

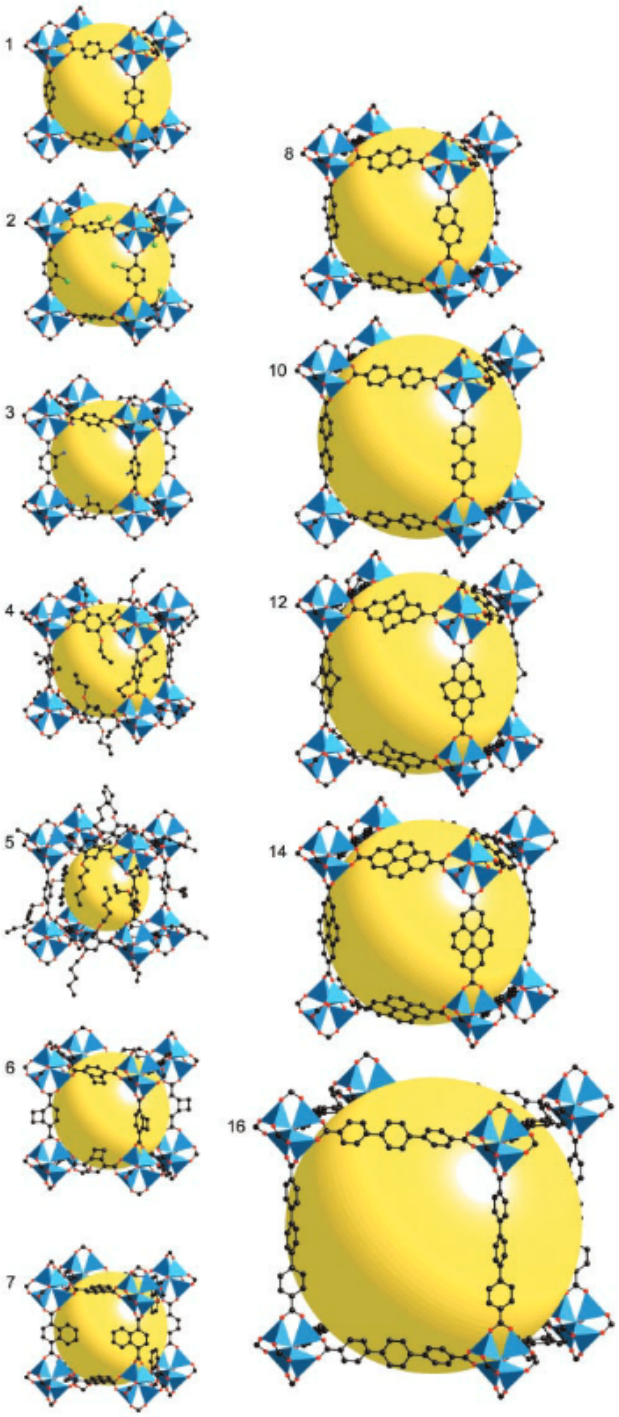


FIGURE 1.7: Image of the IRMOF family with the number n on the side of each image corresponding to IRMOF-n. Taken from [38].

over the course of the years, among which multivariate MOFs (MTV-MOFs) offer interesting perspectives for industrial applications. The principle of MTV-MOFs is to add several different functional groups to a MOFs giving then different properties to a same structure. It is even possible to imagine a material where the different steps of a chemical reaction could occur in the different cavities of the structure each containing different catalytic site. Some structures of this kind have been successfully synthesized as MOF-5 containing several functionalities[49]. They reported ordered structures of MOF-5 with a disordered repartition of functionalities throughout the structure. They also found that the properties of the final materials are not a simple combination of the properties of the raw materials due to the complex arrangement of the moieties. In their case, they observed a MOF-5 with an unusually high selectivity of carbon dioxide up to 4 times better than its original counterpart.

**Properties and applications**

MOFs have created a great interest due to their high potential of flexibility and high porosity. Going from a few Angströms to several nanometers, they filled the void between zeolites and mesoporous materials while also covering the mesoporous range. This large porosity range is of great interest for industrial purposes, especially for catalysis and adsorption-related applications. Their potential in this field is far superior to zeolites, which are limited in pore sizes. Extensive works were realized in order to develop stable MOFs with higher and higher porosity. Famous examples of this regroup the IRMOF family[38] and the Cu paddlewheel-based MOFs like the HKUST-1[50, 51, 52, 53]. MOFs are built from metal nodes linked by organic ligands which means that they are composed mainly of strong bonds (C–C, C–H, M–O, etc.) and weak coordination bonds between the metal nodes and the ligands. This results in an overall good thermal stability[30, 31, 32, 54]. However their chemical stability still represent a great challenge as many of these materials are sensitive to treatment with polar solvents. This issue is commonly addressed by functionalization methods to raise the resistance of the structure to specific environment such as water-containing or acidic/basic environments[31, 54, 55]. The mechanical stability of MOFs constitutes also an important field of research. Indeed they tend to show mechanical properties of the order in between molecules and inorganic solids as they are composed of both. This results usually in soft materials with low to medium bulk moduli[56]. It is however interesting to note that MOFs can exhibit singular mechanical responses due to their particular nature combining both inorganic and organic units[57].

The number of potential applications of MOFs is relatively large, but are mostly focused on adsorption-related fields as their flexibility and potential are enormous in this field. The first application one can imagine is then the adsorption, in particular MOFs have been investigated for the storage of hydrogen and methane as the storage of fuel gases is an important barrier to their development. The first study showing interest in the storage of hydrogen by MOFs confirmed the high potential of this family by investigating the adsorption properties of MOF-5[58]. This was followed later by a study reporting the hydrogen storage potential for hundreds of MOFs[59]. As only increasing the accessible surface area is not enough to increase the uptake of hydrogen, various methods were considered to enhance the adsorption of hydrogen including the use of open metal sites[31, 59, 60] or metal impregnation[61, 62, 63]. Adsorption of methane was also studied as it is another potential source of fuel. In

particular, MOF-177, MOF-200 and MOF-210 were found to show impressive uptakes of methane of around 350 to 470 mg/g, much higher than any other MOF[64]. A tank filled with one of this MOF can store more methane than an empty tank. This technology has since then been developed and commercialized by BASF. This is not the first time a MOF material was developed for commercial use as there are other reports of MOFs in commercial products (for example, the ION-X designed to deliver arsine, phosphine and boron trifluoride[65]).

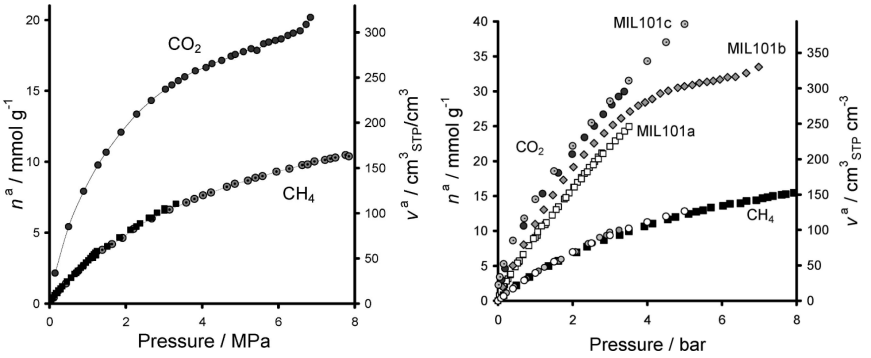


FIGURE 1.8: Adsorption isotherms of carbon dioxide and methane on (Left) MIL-100 and (Right) various samples of MIL-101 at 303 K. Taken from [66].

Other than for adsorption-based storage, MOFs have been investigated for their potential in separation and purification of gases. Difficulties in this field arise from the MOFs having to be stable around the components of natural gas such as water or sulfur-based compounds, while being able to adsorb a specific gas with a high selectivity. Overcoming this difficulty requires a precise engineering of the MOF structure and fine-tuned properties of adsorption. The applications are however numerous, with the capture of carbon dioxide being one of the most studied due to its potential use in reducing the industrial waste and pollution. The current process to separate the $CO_2$ uses an amine-based solvent which has a high selectivity for $CO_2$. However this method is fastidious as the C–N bond formed with the reaction between the solvent and $CO_2$ is very strong and thus, recovering the solvent after the capture has a high cost because of the necessity to heat. This is why a method based on MOFs, which can adsorb and desorb $CO_2$ easily would make this process much cheaper. Researchers are extensively studying this area in order to find the best suited MOF for this particular application. Examples of materials having a high potential for the adsorption of $CO_2$ are the MIL-100 and MIL-101[66]. Examples of adsorption isotherms for these two compounds are shown on fig. 1.8. These two materials present several advantages including an ultrahigh porosity as well as a good thermal and chemical stability as they are stable around water and until 300°C[67, 68]. Moreover their adsorption properties can still be increased by functionalization especially by adding amine group such as ethylenediamine or N,N'-dimethylethylenediamine[69, 70, 71, 72].

Due to their high porosity and high density of tunable active sites, MOFs have also drawn a great interest for catalysis applications. They generally have a continuous and steady porosity easily accessible which allows for a great diffusion of molecules inside the pores and towards the active sites. Thus they can be used either for encapsulation of catalysts, as heterogeneous catalysts or to support homogeneous catalysis[73, 74]. Several considerations have to be investigated to determine the

potential of a MOF for catalysis: the performance compared to reference catalysts, possible leaching issues (dissolution of MOFs in the reaction medium) and the recyclability as a catalyst has to be reusable. Some MOFs have been studied for catalytic applications among which open metal sites structures are probably the most explored materials. The removal of weakly coordinated ligands such as water in these structures creates coordination vacancies on the metal which acts then as a Lewis acid accepting electrons from donor molecules. Examples of this types of catalysts are the well known HKUST-1 and MIL-101. Such MOFs have been already used in specific applications like the oxidation of alcohols[75], the cyanosilylation of carbonyl[76, 77] compounds or the Fridel-Craft benzylation[78], etc.

## 1.2 Computational methods and theories

All of the work done in this PhD thesis was realised using simulation techniques commonly used in the theoretical field such as the density functional theory or the molecular dynamic. Depending on the systems and objectives, I used different methods with different levels of approximation. The knowledge of such techniques is thus required by the reader to fully understand the conclusions made in this thesis. This section will therefore introduce the main principles behind the methods of interest. I will first start with the density functional theory by describing its main principles before defining the theory itself, and then describe the molecular dynamic technique.

### 1.2.1 Basic principles

#### Schrödinger's equation

The starting point to explain the Density Functional Theory (DFT) is to first introduce Schrödinger's equation[79, 80, 81, 82]. Schrödinger was born in 1887 in Vienna and became a professor in this same city in 1921 where he conducted researches until 1933 when he resigned from this position. During that time, he developed a famous equation widely used worldwide, that is depicted in equation eq. (1.1)

$$\hat{H}\Psi = E\Psi \tag{1.1}$$

This equation is a key tool in understanding quantum physics and is the origin of many computational methods that are used nowadays. It consists in an operator $\hat{H}$ (called the Hamiltonian) applied to a function called the wave function $\Psi$ which depends on the coordinates of all the particles of the system. $\Psi$ is then a function of 3n coordinates (being the $x$, $y$ and $z$ for a cartesian frame) where n is the total number of particles of the system. The result of this operation is the same function $\Psi$ multiplied by a term $E$ called the eigenvalue, corresponding to the energy of the system. Schrödinger's equation often admits several eigenfunctions each having a different associated solution $E$ resulting then in an ensemble of discrete levels of energy, corresponding to the stable bound states of the system.

#### The Hamiltonian operator

The Hamiltonian operator, shown in equation eq. (1.2), is composed of two parts: $\hat{T}$ the operator associated to the kinetic energy and $\hat{V}$ the operator associated with the potential energy of the system.

$$\begin{aligned}\hat{H} &= \hat{T} + \hat{V} \\ &= -\frac{\hbar^2}{2m}\nabla^2 + V(r,t)\end{aligned} \tag{1.2}$$

with $\hbar$ the Planck constant and $\nabla^2$ the Laplacian operator. These two parts developed for a system containing several atoms will result in 5 terms contributing to the total energy: first the kinetic energies of the nuclei and the electrons, then the energy of attraction between nuclei and electrons and finally the two repulsive energies of nuclei and electrons between themselves. Using these new terms, the Hamiltonian can be written:

$$\hat{H} = -\sum_i \frac{\hbar}{2m_e}\nabla_i^2 - \sum_k \frac{\hbar}{2m_k}\nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{k<l} \frac{e^2 E_k Z_l}{r_{kl}} + \sum_{i<j} \frac{e^2}{r_{ij}} \tag{1.3}$$

where i and j represent the electrons, k and l the nuclei, $m_e$ the mass of an electron, $m_k$ the mass of the nucleus, Z the atomic number, e the charge of the electron. This form of the Hamiltonian shows that it contains attraction and repulsion terms, meaning that particles do not move independently: all movements of particles are "correlated".

This dependency makes the problem more complicated but we can simplify this a bit. Indeed the masses of the nuclei are far superior to the mass of electrons and thus their kinetic energy are much lower. To the eye of the electrons, nuclei are almost completely fixed and this is the approximation we can make. By considering that motion of electrons is instantaneous compared to the nuclei and considering fixed positions for nuclei, the kinetic energy of nuclei becomes independent of the electrons, the correlation is removed from the nuclear-electron attraction term and the nuclear-nuclear repulsion becomes an analytical calculation. It is then common to rewrite the Hamiltonian to regroup the simply evaluated nuclear terms on one side and the more complicated electronic terms on the other. In practice the nuclear-nuclear repulsion potential is not taken into account when solving the equation and is only added at the end as it is a constant, leaving then only the electronic parts of the Hamiltonian to solve. This approximation is called the Born-Oppenheimer approximation and is valid for the dynamics of most of the systems. It is however interesting to note that, due to symmetry-related avoided crossing, it can sometimes fail and give poor results especially in regions where two or more solutions to the Schrödinger equation are close in energies[83].

#### The Hartree-Fock theory

The Schrödinger equation constitutes a differential equation that can be solved by the standard method of finding a right set of coordinates which can separate the equation in simpler equations. This can be easily achieved for hydrogenoid systems, however for systems containing more than one electron the equation can't be solved analytically due to the electron-electron repulsion term preventing the separation of variables. In such cases the equation needs to be handled by computational (numerical) methods. Several methods exist for this, but in this part I will only briefly introduce the Hartree-Fock theory as it is one of the first methods to deal with this

issue. Later in this chapter I will focus much more on the Density Functional Theory (DFT) as most of the calculations in this Ph.D. thesis were realised under this method.

One of the earliest methods was created by Hartree who suggested to transform the problem of a single many-electron wave function into a product of several one-electron wave function[84] (also called orbitals or molecular orbitals):

$$\Psi(x_1, x_2, ..., x_n) = \frac{1}{\sqrt{n}}\psi_1(x_1)\psi_2(x_2)...\psi_n(x_n) \tag{1.4}$$

where $x$ represents the coordinates and spin for all $n$ electrons of the system. This approximation is one of the simplest one can make, however it comes with some drawbacks. The main problem arise from the approximation not taking into account the exchange interaction. This interaction comes from the Pauli exclusion principle and should be integrated by making the wave function antisymmetric such that:

$$\Psi(x_1, x_2, ..., x_i, ...x_j, ..., x_n) = -\Psi(x_1, x_2, ..., x_j, ...x_i, ..., x_n) \tag{1.5}$$

which is not respected with the Hartree approximation. This problem was later addressed and corrected[85] to include the exchange interactions by writing the wave function as an antisymmetrised product of orbitals giving birth to the well-known Hartree-Fock approximation of the wave function:

$$\Psi_{HF} = \frac{1}{\sqrt{!N}}[\psi_1(x_1)\psi_2(x_2)...\psi_n(x_n) - \psi_1(x_2)\psi_2(x_1)...\psi_n(x_n) + ...] \tag{1.6}$$

Later, Slater discovered that the Hartree-Fock wave functions could be written in a much more pleasant way taking the form of a $(n \text{ x } n)$ matrix now called the Slater determinant[86]:

$$\Psi_{HF} = \frac{1}{\sqrt{!n}} \begin{bmatrix} \psi_1(x_1) & \psi_1(x_2) & \dots & \psi_1(x_n) \\ \psi_2(x_1) & \psi_2(x_2) & \dots & \psi_2(x_n) \\ \vdots & \vdots & & \vdots \\ \psi_n(x_1) & \psi_n(x_2) & \dots & \psi_n(x_n) \end{bmatrix} \tag{1.7}$$

with the columns of this matrix representing the single-electron wave functions (or orbitals) and the rows corresponding to the electron coordinates.

As the Hartree-Fock approximation includes all possible permutations of the particles, one advantage of this method is that the exchange energy obtained is exact. However the main problem lies in the correlation part of the energy. Indeed, as the many-electron wave function cannot be expressed with a single Slater determinant, the approximation does not fully include the correlation interaction and, as a result, energies with the HF method tend to be too high. Over the course of the years, several possibilities were created in order to integrate the missing correlation and to overcome this issue. All these possibilities are grouped in what we call now the "post Hartree-Fock methods". In this group we find the Configuration Interaction, the Möller-Plesset, the Coupled-Cluster or the Multi-Configuration Self-Consistent Field techniques. I will not develop any further on these methods but it is interesting

to note that for some of them it is in theory possible to solve the Schrödinger equation exactly and obtain the exact true energy of the system. For example, this is the case for the Configuration Interaction if one includes a complete basis of single-electron wave functions which can allow to express the many-electron wave function as a linear combination of all the possible determinants made of these wave functions. However, such calculations require obviously a tremendous amount of computing ressources and cannot be applied to big systems. Nevertheless, post-treatment methods of the HF theory can still be efficient and obtain the energy with a good accuracy for big systems and this is why this approximation is still used nowadays.

**The basis set**

We saw in the previous part one of the first method to deal with the problem of the electron-electron repulsion term. But let us put aside the issue of many-electrons systems for the moment and concentrate for the time being on the wave function $\Psi$. Indeed while we saw how to approximate the many-electrons wave function $\Psi$ into a linear combination of single-electron wave functions, we have yet to define how to express those functions. For small systems such as atoms and diatomic molecules, we can solve directly the Hartree-Fock equations by mapping the orbitals on a set of grid points. However larger systems typically need a set of basis functions, used to express the unknown Molecular Orbitals (MO). The choice of this basis set is a key parameter of the electronic structure calculations as the wave function should represent the coordinates of all particles of the system and is associated through the Hamiltonian operator to the resulting energy of the system. The variational principle states that:

$$\frac{\int \phi \hat{H} \phi dr}{\int \phi^2 dr} \geqslant E_0 \tag{1.8}$$

or in Dirac notation

$$\langle \phi | \hat{H} | \phi \rangle \geqslant E_0 \tag{1.9}$$

with $\phi$ being any possible wave function. The main outcome of this principle is that we can evaluate the quality of the wave function simply by obtaining its energy. We can then choose whatever functions we want to describe the system and we can use all the mathematical tools we can to find the function with the lowest energy possible. Several possibilities can be then considered for the creation of the basis set: polynomial, exponantial, gaussians, plane waves, etc. However there are some rules for choosing this set of basis functions. First, these functions should be relevant with respect to the physics of the system. This ensures that adding more basis functions will make the convergence faster. For atoms and molecules this basically means having a function which follows the probability density of finding an electron in given point of space. The second rule is to use functions that are computer-friendly and can be computed efficiently. Following the first rule, one may think that exponential functions centered on the nuclei would be a good choice of base functions, however multi-centered integrals of exponentials are quite complicated to deal with computationally. Gaussian functions are less accurate but are easier to compute and the trade between loss of accuracy and increased speed of calculation is worth enough to choose Gaussians functions over exponentials.

The usual form of a Gaussian base function, also called Gaussian Type Orbital (GTO), is:

$$\chi_{\alpha,i,j,k}(x,y,z) = Nx^i y^j z^k \mathrm{e}^{-\alpha r^2} \tag{1.10}$$

where $x$, $y$ and $z$ are cartesian coordinates and the sum of $i$, $j$ and $k$ determines the type of orbital (for example $i+j+k=0$ would mean that we are dealing with a spherical $s$-orbital). There is another type of function often used as basis functions for a basis set, called the Slater Type Orbitals (STO). In theory these are more efficient in describing the physics of the atoms than Gaussians and thus one would require fewer functions to make a complete basis set. However the ease of computing Gaussians being really attractive makes so that GTOs are more than competitive for a basis set construction. Another possibility for periodic materials is to use a plane-wave basis set which have the advantage of ensuring a smooth and monotonic convergence for the wave function. In this case there is no need to choose the number of base function to use, this is decided by fixing a limit on the energy of the plane-waves which will be introduced. All plane-waves with a lower energy than the cut-off are used in the basis set. One then needs to define this cut-off energy by running convergence test in order to find the best value which gives accurate results while keeping the cost of the calculation affordable. However, only GTOs were used in this Ph.D thesis which is why I will focus only on this type of basis set.

A single Gaussian function as described in equation eq. (1.10) is not enough to describe the physics of the atom. Therefore basis sets are constructed as a sum (or contraction) of several Gaussian functions which are called "primitive". These contracted GTO basis sets have particular names indicating how they are constructed. This notation includes all the informations on how many functions are used on which type of electrons. The first number designates the number of primitives used to describe the core electrons of the atoms. Following an hyphen we then find the number of primitives used for each orbitals of the valence electrons. Since only valence electrons contribute to the bonding, it is common to use more basis functions to describe them. An example of a famous basis set using this notation is the 6-31G which uses one base function composed of 6 primitives for the core electrons, and 2 basis functions using respectively 2 and 1 primitives to describe the valence electrons. A first improvement can be made to this basis set by doubling all the basis functions creating a Double Zeta (DZ) basis set. This has the advantage of allowing more different bondings in different directions. For example it would help in describing better the HCN molecule where a $\sigma$-bond and a $\pi$-bond are present as the basis set would be able to describe different distribution of electrons in the different directions. Even larger basis set can be created, such as triple Zeta (TZ), Quadruple Zeta (QZ) and so on. In practice the multiplication of basis functions is only applied to valence orbitals as it does not improve the description of core electrons.

This basis set can be further improved with the integration of what is called a polarization function and is noted as an asterisk in the name of the basis set. The point of adding such function is to add more mathematical flexibility for the description of the wave function. They are generally incorporated in the form of a basis function corresponding to one quantum number higher angular momentum than the valence orbitals. For example carbon atoms would require the polarization function to be a $d$-GTO.

**The Self-Consistent Field algorithm (SCF)**

Now that we have defined the different elements of the Schrödinger equation, especially how to approximate the wave function under the HF theory and how to construct it with a basis set, we can finally focus on how to solve this equation and obtain the best result for the energy in practice. For that we use an algorithm called the Self-Consistent Field (SCF) algorithm. This algorithm is used in order to converge progressively to the best energy possible starting with a relatively good guess on the wave function.

First the user has to choose the basis set to use in the calculation and the geometry of the system (input coordinates, atomic species, etc.). From this point, the necessary integrals are calculated and an initial guess for the MO coefficients is formed. These MO coefficients are used to create an initial guess for the density matrix which is an alternate way of expressing the quantum state of a system. Using this density matrix as well as the integrals, the Fock matrix is constructed and then diagonalized. The diagonalization of the Fock matrix results in new MO coefficients that can be used to construct a new density matrix. This new density matrix is then used to construct a new Fock matrix and which in turn is diagonalized to get a new density matrix and so on until the obtained density matrix is similar enough to the previous one which stops the algorithm. Once this condition is satisfied we obtain the energy of the system.

Doing only one SCF loop is called a single-point calculation and is used to obtain an accurate value for the energy for a specific atomic configuration (position of the nuclei). However this requires to optimize the geometry beforehand which adds a second loop to the SCF algorithm. After doing an SCF loop, the condition of geometry optimization are checked (either residual forces on atoms, stress on unit cell, etc.) and if they are not satisfied then the geometry is updated and a new SCF starts again. Once optimization conditions are satisfied then the algorithm stops and we obtain the equilibrium structure. This is what is called optimizing a structure: finding its configuration of minimum energy (locally)

There are some numerical issues hidden behind this process, one being the fact that we do not know if the energy will actually converge. Small basis sets and structures close to equilibrium will converge quite fast, but it can be harder for more complicated structures or basis sets. It is even worst in the case of metals in which several states with similar energies are possible. Some tricks can be used to help convergence in these cases such as "damping" technique which consists of mixing a certain percentage of the $n-1$ calculated density matrix to the $n$ calculated one. This helps preventing the algorithm from oscillating and stabilizes it. Another common tool is the "level shifter" which adds a shift in the energies of the occupied orbitals reducing their coupling with the unoccupied ones.

**Periodic systems and periodic boundary conditions**

Currently we know how to express the different elements of the Schrödinger equation and how to find the best possible wave function through the use of SCF method. However this requires to fully describe the system in order to generate a correct initial guess for the density matrix. Even though some systems can be quite large (especially in biology where molecules such as proteins can contains thousands of atoms), this can be achieved for most molecular systems as they are finite systems. However in the case of solid materials, and crystals in particular, the size of system

to represent, which can be considered infinite, makes it impossible to describe all of the particles. In that case, we need a method to describe solid state materials entirely while limiting the amount of particles to compute. This is achieved by the use of periodic boundary conditions. The idea is to create a box containing some representative pattern of particles of the system and to repeat the same pattern around the box created. This box is referred to as the simulation box or input cell and is taken to be most of the time the crystallographic or the primitive unit cell of the structure for solid state materials. It can be noted however that this method can be applied also to glasses, liquids and gases. A representation of the mechanism of the periodic boundary condition can be found on fig. 1.9.

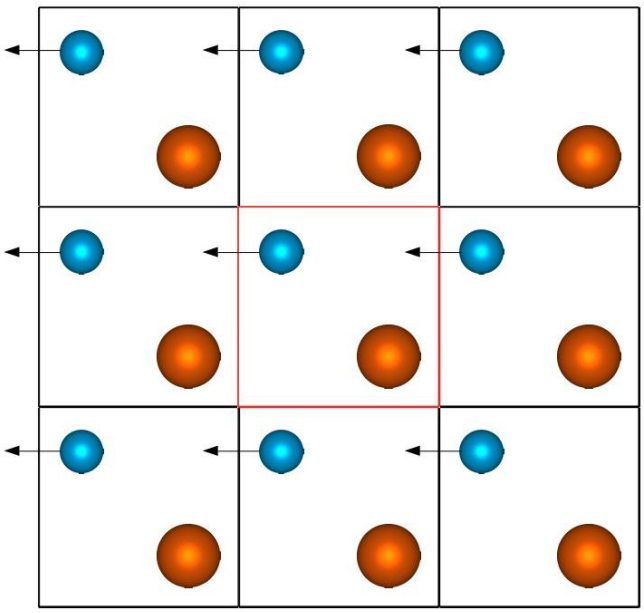

FIGURE 1.9: Illustration of the periodic boundary condition with an atom moving out of the box. The red square represents the input cell and blue and orange spheres represent atoms inside this cell.

As depicted on this figure, the input cell is duplicated in all directions and all the movements of the atoms are equivalently replicated in the cells surrounding the input cell. During an optimization process, when an atom (like the blue one) is mowing out of the box, there will be an atom from an outer box moving inside the box on the opposite side. Thus, following this scheme, every atom moving outside of the input cell will reappear on the opposite side. This method allows to mimic an infinite material without having to compute it entirely. All solid state material computation uses this technique and it has been used as well in all the calculations performed in this work.

### 1.2.2 Density functional theory

The core idea behind the Density Functional Theory (DFT) is to abandon the idea of working with and approximating a wave function admitting $3N$ parameters and to work instead with a physical property that can be observed and measured and use it in the determination of the energy. Since the Hamiltonian is built using the positions and the atomic numbers of nuclei as well as the number of electrons, we must find a property or an observable which contains these informations. In that sense, the electron density $\rho$ seems to check all the boxes. Indeed, the total number

of electrons can be obtained by integrating the electron density over all space, the positions of nuclei are deduced from the maxima in the electron density and even atomic numbers can be retrieved from it. This suggests that the electron density could be used to simplify the problem and this is the main idea in DFT.

### Thomas-Fermi model

The very origin of the DFT goes back to the year 1927 when Thomas and Fermi [87, 88] were both working independently on the electronic structure. The major problem at the time (which still applies today) was the calculation of the electrostatic potential which was impossible for atoms with 2 or more electrons. In order to calculate the energy of more complex systems, these two researchers had the idea to completely bypass the multi-electron wave function and to work with the electron density instead. They decided to express the different terms of the hamiltonian using the electron density. Considering the system to be classical, the potential terms can be easily rewritten:

$$V = \sum_i \int \frac{Z_k e^2}{r_{ik}} \rho(r_i)\, dr_i + \frac{1}{2} \int \int \frac{\rho(r_i)\rho(r_j)}{r_{ij}}\, dr_i dr_j \tag{1.11}$$

with the first term being the attraction between the nucleus and the electron and the second corresponding to the electron-electron repulsion. $i, j$ correspond to the electrons and $k$ represents the nucleus. While these two terms are straightforward, the kinetic energy is not as easy to determine. For that they considered a theoretical environment called the uniform electron gas or "jellium" [89]. This approximation considers the electron density to be uniformly distributed over all the space, disregarding then the effect of the lattice in the definition of the electron density. This theoretical substance can be used in inorganic chemistry for metals as the resulting smooth uniform density is suitable for representing qualitatively the behavior of metals. In this case, the use of jellium along with Fermi-Dirac statistics allowed Thomas and Fermi to define the kinetic energy as a function of the electron density:

$$T = \frac{3\hbar^2}{40 m_e} \left(\frac{3}{\pi}\right)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(r_i)\, dr_i \tag{1.12}$$

This constitutes the first attempt in making a DFT method, however the results they obtained from their model were not satisfying[90]. Results were not accurate enough and even if a few corrections were added to try to improve the model, the idea of using the electron density has been dismissed until 1964 when Hohenberg and Kohn proved the key theorems for defining the DFT.

### Hohenberg and Kohn theorems

The Thomas-Fermi model has made the first step towards a DFT method, yet it failed to provide the wanted results. Some necessary considerations were missing and these were added by Hohenberg and Kohn who defined the two important theorems building the base of the DFT [91].

The first theorem is the existence theorem which states that the external potential $V_{ext}$ (designating the interaction between nuclei and electrons) uniquely determines the electron density $\rho_r$ and thus $V_{ext}$ is a functional of the electron density. There can't be

two external potentials yielding the same ground state electron density. The proof of this theorem is surprisingly simple which makes us wonder how it took so long from the Thomas-Fermi model to find this theorem. Let us assume two different external potentials $V_{ext}$ and $V_{ext}^{'}$ which lead to the same electron density $\rho(r)$ associated to the non-degenerated ground state. We can define two different hamiltonians $\hat{H}$ and $\hat{H}^{'}$ associated to two different wave functions $\Psi$ and $\Psi^{'}$ which differ only in their external potential $\hat{H} = T + V_{ee} + V_{ext}$ and $\hat{H}^{'} = T + V_{ee} + V_{ext}^{'}$ with $V_{ee}$ being the electron-electron repulsion term. Following the variational principle, we can write:

$$\begin{aligned} E_0 &< \langle \Psi^{'} | \hat{H} | \Psi^{'} \rangle \\ &< \langle \Psi^{'} | \hat{H}^{'} | \Psi^{'} \rangle + \langle \Psi^{'} | \hat{H} - \hat{H}^{'} | \Psi^{'} \rangle \\ &< E_0^{'} + \langle \Psi^{'} | V_{ext} - V_{ext}^{'} | \Psi^{'} \rangle \end{aligned} \tag{1.13}$$

A similar result is obtained when switching the primed and unprimed characters as follow:

$$E_0^{'} < E_0 + \langle \Psi | V_{ext}^{'} - V_{ext} | \Psi \rangle \tag{1.14}$$

Adding these two previous equations we obtain the following result:

$$E_0 + E_0^{'} < E_0^{'} + E_0 \tag{1.15}$$

which is not possible. This proves then that there can't be two different $V_{ext}$ defining the same ground state electron density, so the external potential uniquely defines the electron density.

The second essential theorem of Hohenberg and Kohn states that, just like MOs, the density obeys the variational principle being:

$$\langle \Psi | \hat{H} | \Psi \rangle = E \geqslant E_0 \tag{1.16}$$

This leads to the same conclusion as for the Hartree-Fock theory, meaning that we can optimize the density by choosing different densities and evaluating their quality by calculating their energy. However this does not solve our problem. Indeed this is equivalent to the Hartree-Fock method where the Schrödinger's equation is solved by approximating the wave function. The point of a DFT method is to avoid the resolution of the Schrödinger's equation. In order to avoid this problem, we should avoid defining the wave function or the hamiltonian with respect to the density and instead work on determining the energy directly from the electron density without the help of the wave function. This issue was solved by Kohn and Sham in 1965[92] when they expressed the hamiltonian as a sum of non-interacting one-electron operators. They wrote then the energy as:

$$E[\rho(r)] = T[\rho(r)] + \int V_{ext}\rho(r)\, dr + \frac{e^2}{2} \int \int \frac{\rho(r_i)\rho(r_j)}{r_{ij}}\, dr_i dr_j + E_{xc}[\rho(r)] \tag{1.17}$$

with the first term being the kinetic energy of the non-interacting electrons, the second term is the interaction between the nuclei and the electrons and the third is the electron-electron repulsion. Now in this expression, no terms for the correlation or the exchange part of the energy are appearing. In fact they are grouped and "dumped" in a term called the exchange and correlation energy $E_{xc}[\rho(r)]$. The particularity of this equation is that the first three terms can be calculated exactly for all systems. The only unknown part of this expression is the exchange and correlation term for which we don't know the form and have to approximate. The key part in DFT in then to express the last term of the equation with a suitable function and this still remains a very active domain of research.

In practice the Kohn-Sham energy is calculated self-consistently like in HF theory. The algorithm is in fact quite similar: the user must first choose the basis set and the geometry of the system. Then the initial guess for the density matrix is formed and the necessary integrals are calculated. However instead of constructing and diagonilizing the Fock matrix, the DFT aims to construct and solve the Kohn-Sham equations:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{eff}\right]\psi_i(r) = \epsilon_i\psi_i(r) \tag{1.18}$$

where

$$V_{eff} = V_{ext} + \int \frac{\rho(r_i)}{r_{ij}} dr_{ij} + \frac{\partial E_{xc}[\rho(r)]}{\partial \rho(r)} \tag{1.19}$$

$\psi_i(r)$ are the single electron wavefunctions constructed from the linear combination of atomic orbitals. From the results, the density matrix is evaluated and compared to the previous one. If it's close enough from the $n-1$ density then the SCF cycle stops. If not, then the Kohn-Sham equations are built from the latest density matrix and solved again to obtain a new density as long as the conditions of convergence are not satisfied.

**Exchange and correlation treatment**

It exists several ways of treating the exchange and correlation term of the Kohn-Sham equation. Each techniques have their own accuracy and downsides. I will briefly present the most famous methods as well as the ones used in this Ph.D thesis.

*The Local Density Approximation (LDA):* This method considers that the variations of density are slow and treats then the density as a uniform electron gas. The $E_{xc}[\rho(r)]$ term is then obtained as an approximate value of the exchange and correlation of an electron in an uniform electron gas system of same density $\rho(r)$. This approximation gives quite good results for systems with slowly varying charge densities and its accuracy is good in most cases. Due to its tendency to over-bind molecules and solids, it usually results in shorter bond lengths than experiments as well as an overestimation of the total energy. LDA does not admit several applications (or functionals), this approximation only has one flavor.

*The Generalized Gradient Approximation (GGA):* This method was introduced as an improvement of the LDA. It uses the same considerations but takes additionally the gradient of the electron density, allowing to account for the rapid changes in density that the LDA fails to represent. In contrary with the LDA, GGA exists in several

flavors each having their own advantages and drawbacks. Some are obtained semi-empirically, some are derived from experimental data and others are obtained from first-principles. Examples of famous and widely used functionals are PW91[93] or PBE[94].

*The hybrid functionals:* The principle of this method is to make use of the exact exchange obtained in the HF theory. A part of exact exchange is introduced in the calculation which generally improves a lot the quality of the result. This improvement is due to the exchange energy introduced being exact and to the reduction of self-interaction error (SIE) induced by this introduction. Self-interaction error arises in DFT due to the residual interactions of electrons with themselves. Such error does not happen in the HF theory, hence the reduction of SIE when replacing DFT exchange energy with HF exact exchange. Famous examples of hybrid functionals are B3LYP and PBE0[95] which is derived from the PBE GGA functional.

#### Dispersion correction

Despite the great results of DFT, there are still some behaviors that the current functionals fail to represent and for which they give generally poor result. Among those cases we find the lack of description of weak interactions. These interactions are due to dispersion forces arising from long-ranged electron correlation effect in the form of $\frac{C_6}{R^6}$ (London forces). Because of their attractive nature, they are often crucial in calculations as they might be at the origin of several phenomena such as the crystal packing of molecular solids or the adsorption of molecules on surfaces or porous materials. These interactions has since been corrected only in an empirical way by adding a correction term to the expression of the DFT energy. This term is obtained with an efficient and cost-effective approach using a two-body correction including the long-range dispersion through an expression of the form:

$$E^{Disp} = -\sum_{AB}\sum_{n=6,8,\ldots}\left(\frac{C_n^{AB}}{R_{AB}^n}\right) \tag{1.20}$$

Different attempts were made for the implementation of this technique but the most used are the D2 or D3 methods proposed by Grimme[96, 97]. D3 is however usually preferred over D2 as it is generally more accurate. Indeed, in D3 the $C_n$ coefficients depend on the coordination number of atoms while in D2 they depend only on the pair of elements involved.

### 1.2.3 Molecular dynamic

#### Basic principles of classical molecular dynamics

The limitations of DFT calculations arise due to their static nature. Indeed DFT calculations are done at zero Kelvin, making it impossible to obtain properties related to thermal motions or entropic effects. Some approximations and methods exist to get around this problem (and will be discussed later in chapter 2) however the method of choice to investigate motion is the molecular dynamic (MD). Molecular dynamics consist in predicting the movement of particles over time by calculating their interaction over a fixed period of time. This is achieved by integrating the Newton's equation of motion for all the particles of the system simultaneously:

$$F_i = m_i a_i = m_i \frac{d^2 x_i(t)}{dt} \tag{1.21}$$

with $i = 1...N$ particles. If one has access to $F_i$, then one can integrate the trajectory $x(t)$ of the whole system as a function of time. The forces can be calculated as the gradient of the potential energy of the system:

$$F_i = -\frac{\partial U(r^n)}{\partial r_i} \tag{1.22}$$

which leaves us with the problem of the determination of the potential. Several contributions have to be taken into account. First, the intramolecular interactions which correspond to the degrees of freedom of the molecule. The different types of vibrations have to be accounted for, i. e. the stretching (vibration of the length of the bond), bending (vibration of the angle between 3 atoms) and torsion energies (vibration of the dihedral angle of a bond).

As for interatomic interactions, they can be divided in three parts. The first part is linked to the two-body Coulomb interaction between different particles with point charges. The expression of the Coulomb energy is as follow:

$$U_{coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{1.23}$$

This interaction weakens as the distance between particles increases at the rate of $\frac{1}{r_{ij}}$. The second and third part correspond to the Van der Waals interactions and the electronic repulsion at short distances. These two parts are generally grouped together in the same expression, usually a Lennard-Jones or a Buckingham potential:

$$U_{LJ} = 4\epsilon \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1.24}$$

$$U_{Buck} = \alpha e^{-\beta r} - \gamma r^{-6} \tag{1.25}$$

In any cases of classical MD, all of these interactions are then described empirically which, as a result, causes the properties to be computed numerically but allows the classical MD to be quite efficient for very large systems or very long time scales. This technique is thus an ideal choice for biological simulations where molecules can have tremendous sizes.

MD calculations work with an algorithm that is used to update the positions by integrating the equation of motions. There are several possibilities, each one having downsides and advantages. The most simple method is called the Euler method and consist in replacing the differential in the differential equations by finite differences. This method is easy to understand but leads to large errors for time step that are not small enough. In general, this algorithm is not recommended as it does not respect the necessary conditions for an integration algorithm such as the time-reversability or the ergodicity. Other algorithms are usually preferred such as the Verlet algorithm for example which is almost as simple as the Euler method. Considering a small time

step $\Delta t$ between the current positions $r_t$ and their previous and next ones $r_{n-1}, r_{n+1}$, the Verlet algorithm is derived from two Taylor expansion:

$$r_{t+\Delta t} = r_t + \frac{dr_t}{dt}\Delta t + \frac{1}{2}\frac{d^2r_t}{dt}\Delta t + \frac{1}{3!}\frac{d^3r_t}{dt}\Delta t + O(\Delta t^4) \tag{1.26}$$

$$r_{t-\Delta t} = r_t - \frac{dr_t}{dt}\Delta t + \frac{1}{2}\frac{d^2r_t}{dt}\Delta t - \frac{1}{3!}\frac{d^3r_t}{dt}\Delta t + O(\Delta t^4) \tag{1.27}$$

Adding together these equations and simplifying it we obtain:

$$r_{t+\Delta t} = 2r_t - r_{t-\Delta t} + \frac{d^2r_t}{dt}\Delta t + O(\Delta t^4) \tag{1.28}$$

This scheme has the advantage of being almost as simple as the Euler method while satisfying the conditions of an integration algorithm. It only requires the current positions $r_t$ and the previous ones $r_{t-\Delta t}$ in order to determine the next positions and its error $O(\Delta t^4)$ is also reasonable. Improvement of this scheme were made to give birth to the more efficient and reliable leap frog and velocity Verlet algorithm. As I did not use classical MD, I am only mentioning these techniques for reference and not developing them. However, other algorithms, especially thermostats, will be discussed later in this part.

#### First principles molecular dynamics

During my PhD, I did not use classical MD as it fails to properly representsystems where bond cleavage or bond formation occur. Instead, I used the so-called Ab-Initio Molecular Dynamics (AIMD) which, along with a much higher computational cost than classical MD, allows to treat the thermal motion of the system at the quantum chemical level. Different flavours of AIMD are available such as Born-Oppenheimer MD (BO-MD), Car-Parinello method[98] or Ehrenfest, method but I only used BO-MD in my PhD and thus will only describe this method.

Comparatively to the Born-Oppenheimer approximations mentioned earlier in this chapter, the BO-MD consists in decoupling the motions of nuclei and electrons by making use of their tremendous mass difference. Nuclei and electrons are then treated separately and in this case, are treated at different levels. Nuclei are still treated classically using the Newton equations of motions while electrons are treated at the DFT level. BO-AIMD involves thus DFT calculations of the energy to determine the forces on atoms and update the positions. This results in a much higher computational cost than classical MD or quantum calculations but allows for more accurate trajectories and allow for specific behavior such as bond cleavage or bond formation to be computed.

#### Ensembles

In molecular dynamics, we need to define which properties will remain constant and which ones will variates during the simulation. Do we want to let the system variates freely without any interaction? Do we want to keep the temperature of the system constant throughout the whole simulation by forcing it to interact with a thermostat? These "rules" under which the system will operate are defined by what we call the ensembles. Ensembles are commonly used in thermodynamics to simplify

the system and isolate some parts from its surroundings. It is possible to isolate all the system, resulting in a so-called microcanonical ensemble, or to completely open it, resulting then in a grand canonical ensemble. Various thermodynamic ensembles exist, each isolating different properties but in the case of molecular dynamics only three ensembles are used:

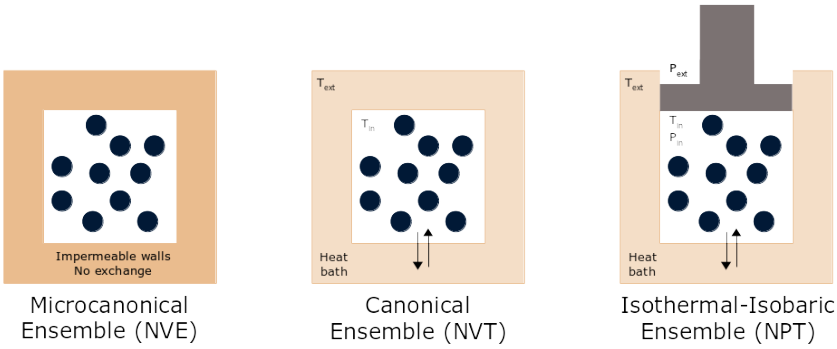


FIGURE 1.10: Schematic of the (left side) microcanonical ensemble where no exchange are allowed with the surrounding of the system, (center) canonical ensemble where the system is allowed to exchange heat with a thermostat resulting in $T_{in} = T_{ext} = constant$, (right) isothermal-isobaric ensemble where the system can exchange heat and work with its surrounding resulting in $T_{in} = T_{ext} = constant$ and $P_{in} = P_{ext} = constant$.

- *NVE Ensemble:* This ensemble is the most simple. As depicted in fig. 1.10, it considers a system completely isolated from any interaction. As a result, this ensemble considers the quantity of particles N, the volume V and the total energy E of the system to remain constant. Such ensemble is thus usually called the "NVE ensemble" but is more formally known as the "microcanonical ensemble". Its use is however limited in molecular dynamics due to its nature. Indeed the total energy being constant, a small variation of potential energy will cause an opposite variation of the kinetic energy leading then to a variation of temperature. It is then complicated and near impossible to keep a constant temperature with this ensemble as it would require the system to be in a perfect state at the beginning of the simulation. This is even more true for large systems as they will most likely not start in a perfect state and thus lower their potential energy to a preferred state, leading to an increase of kinetic energy and obviously in temperature as well due to the total energy being constant. It is then common to observe an increase of temperature when using the NVE ensemble. While limited, this ensemble can still be of use in certain cases, for example when studying small systems for which we already know how it will behave, or for the exploration of constant energy surface of conformational space or simply for classical MD.

- *NVT ensemble:* This ensemble is probably one of the most used in MD. In contrary with the microcanonical ensemble, here the system is allowed to exchange heat with a thermostat set up by the user. This results in a constant temperature throughout the whole simulation except in the beginning where the temperature is gradually converging to the thermostat value. Because the amount of substance N and the volume is still kept fixed, this ensemble is usually called the "NVT ensemble" or the "canonical ensemble". The method to achieve a constant temperature is quite easy: as velocities are related to the temperature, it is simply needed to rescale the velocities at each step depending on the difference between the system's temperature and

the thermostat. If the temperature is too low then velocities are scaled up a bit, and inversely if the temperature is too high velocities are scaled down. The NVT ensemble is very useful when one wants to realize simulations at constant temperature, however it presents drawbacks like the NVE ensemble, the most important one being the regulation of pressure. This problem can arise usually when using solvents for examples as it does not coincide perfectly with the surface of the system. This leads to a vacuum between solvent molecules and the system which is filled when the simulation starts by rearranging the disposition of the solvent and the system. Due to the volume being constant, this rearrangement is accompanied by a decrease of the pressure leading then to an overall lower pressure than the one desired for the simulation. Nevertheless, the canonical ensemble is widely used for MD as it is a good ensemble for sampling systems at constant temperature and also a necessary ensemble for computing the helmholtz free energy.

- *NPT ensemble:* The last ensemble in molecular dynamics is as useful as the previous canonical ensemble. While keeping the temperature constant through the same process as the NVT ensemble, the NPT ensemble also keeps the pressure constant which was the main lacking point of the previous ensemble. To achieve such result, the volume is not anymore a fixed quantity and can vary to accommodate the changes in pressure. To take up the example of solvents, if solvent molecules are drawn to the system due to the presence of void, leading then to a decrease of pressure, the volume of the system will be able to reduce in order to keep the pressure constant. This ensemble keeping at the same time pressure and temperature constant is conveniently called the "isothermal-isobaric ensemble". Such ensemble is useful for chemistry as most of the chemical reactions are carried out at a constant pressure. It is also necessary to use this ensemble when computing the Gibbs free energy.

**Thermostats**

During my PhD, I only used the NVT ensemble. In order to sample this ensemble, the use of a thermostat is mandatory. A thermostat is an algorithm that maintains the temperature constant by realistically adding or removing energy from the boundaries of an MD system. There are many options to use as a thermostat but in my calculations I only used the Canonical Sampling through Velocity Rescaling (CSVR) algorithm[99].

The simplest way of sampling a canonical ensemble one could imagine would be to scale the velocities at each time step. Indeed the velocities of the particles are directly linked to the temperature of the system. We could then imagine a factor $\lambda = \sqrt{\frac{T_0}{T(t)}}$ with $T_0$ being the targeted temperature and $T(t)$ being the current temperature. At each time step, velocities are scaled by this factor $\lambda$ until the desired temperature $T_0$ is reached. This is the most simple thermostat but it is usually not used as it does not take into account the fluctuations in temperature. Another simple and closely formulated option is the Berendsen thermostat which considers the system to be connected to a heat bath. Like in the velocity scaling thermostat, the velocities are scaled at each step with the difference that the rate of change of temperature is proportional to the difference in temperature:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_0 - T(t)) \tag{1.29}$$

With $\tau$ being the coupling parameter between the heat bath and the system and describing how tightly they are coupled. The choice of value for $\tau$ is empirical and crucial for the simulation. A too high value leads to the thermostat being inactive and a too small value leads to unrealistic changes of temperature. Typical values for this parameter are around 0.1 ps. The temperature is changed at each step following:

$$\Delta T = \frac{\partial t}{\tau}(T_0 - T(t)) \tag{1.30}$$

The scaling factor $\lambda$ is then expressed as:

$$(\lambda^2 - 1)T(t) = \frac{\partial t}{\tau}(T_0 - T(t)) \tag{1.31}$$

This thermostat is smoother than the simple velocity scaling and is usually good for large systems. However, like the latter, it does not sample the desired canonical ensemble. The CSVR algorithm proposed by Bussi and coworkers[99] is an extension of the Berendsen thermostat and samples correctly the canonical ensemble. I redirect the reader to their article for a full description of this algorithm.

# Chapter 2

# Systematic Study of the Thermal Properties of Zeolites

In this chapter, I will describe the systematic study of thermal properties I realised during my PhD. In contrary with the mechanical properties which have been extensively studied on zeolites in the past literature, thermal properties are much more difficult to obtain both experimentally and computationally. As a result, fewer studies have been realised on these properties despite the need of knowledge for industrial applications. Zeolites are known for showing abnormal thermal behavior as they often exhibit negative thermal expansion (NTE). Up until now, no systematic study on the thermal properties of these materials has been realised and the need for such large scale studies is increasing. Such works are useful for identifying common behaviors of a group of materials or for creating databases to support further high throughput works. This was the intent behind the study described in this chapter. The first section of this chapter will deal with the theoretical background on thermal properties and how to calculate them and introduce the scarce literature on the subject. The second section will present the parameters used and the results.

The work described in this chapter has been published in the Journal of Physical Chemistry C[100].

## 2.1 Thermal and mechanical properties of zeolites

### 2.1.1 The negative thermal expansion

#### The concept of thermal expansion

The thermal expansion designates the variation of lattice parameters upon heating or cooling a crystal. The vast majority of solids exhibit a positive thermal expansion (PTE) meaning that they expand upon heating. The mechanism and origin of PTE are well known and can be visualized with the help of fig. 2.1 representing the potential energy with respect to the interatomic distance. Upon heating, the vibrational energy increases and, due to the slope of the potential energy being lower in the direction of the lengthening of distances, the bond length increases. This anisotropy of the potential energy is due to the short-distance repulsion being typically very strong. No examples of bond shortening have been observed so far which is understandable when looking at fig. 2.1. The stronger the bonds are in a material, the steeper the potential energy curves will be and the weaker the thermal expansion will be. The thermal expansion is usually measured with the so-called thermal expansion coefficients $\alpha$ which can be obtained by the following expressions:

$$\alpha_V = \frac{\Delta V}{V_0 \Delta T} \tag{2.1}$$

$$\alpha_l = \frac{\Delta l}{l_0 \Delta T} \tag{2.2}$$

A material showing PTE will have a positive value of $\alpha$ and, inversely, a material contracting upon heating will have a negative $\alpha$. As shown in eq. (2.1) and eq. (2.2), $\alpha$ can be expressed in 2 forms: volumetric ($\alpha_V$) or linear ($\alpha_l$). The volumetric coefficient describes the global expansion of the structure, that is to say the volume, and the linear parameter describes the expansion in a specific direction, i.e. each of the lattice basis vectors. These two coefficients are obviously correlated and in the case of an isotropic solid the relation $\alpha_V = 3\alpha_l$ can be easily understood (with $\alpha_l = \alpha_a = \alpha_b = \alpha_c$). However in the case of an anisotropic solid the relation is not straightforward as the three axes could show different amplitude or even different signs, and could not be contributing equally to the global volumetric coefficient.

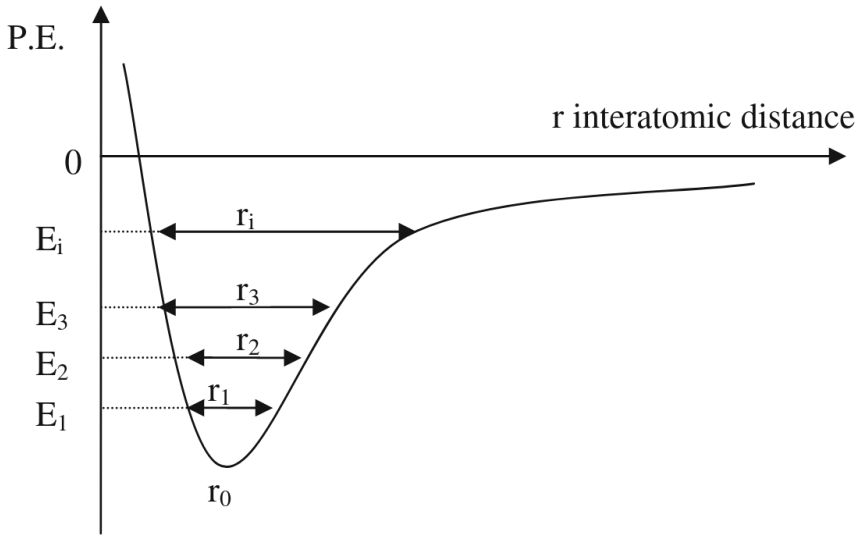


FIGURE 2.1: Typical graph of the potential energy (P.E.) with respect to the interatomic distance for a bond. $E_i$ corresponds to the energy level and $r_i$ corresponds to the range of distances accessible in each microstate. Taken from [101]

While the most observed and the most logical behavior is the positive thermal expansion, there are some materials that show a negative thermal expansion (NTE) meaning that they retract upon heating. What is even more counter-intuitive is that this phenomenon occurs in common materials such as metal-oxides like $ZrW_2O_8$ or $PbTiO_3$[102, 103], zeolites[104, 105], AlPOs[106] and MOFs[107]. This means that there are specific mechanisms guiding the NTE, either related to a specific structural arrangement or to higher order mechanisms, i.e. supramolecular mechanisms.

#### Negative thermal expansion and its mechanisms

The negative thermal expansion cannot come from the shortening of bonds as previously seen on the diagram of potential energy. Instead, this phenomenon was found to come from a variety of supramolecular mechanisms which cause the shrinkage of the overall volume of a structure. Multiple origins are then possible: low frequency

phonons, phase transition, libration or Rigid Unit Modes (RUM). Each one of these mechanisms can be responsible for NTE behavior depending on the material studied. However it has been found that a lot of solids exhibiting NTE contains specific M–O–M linkages such as the Si–O–Si present in zeolites. These linkages are present in most metal oxides and zeolites and are an important feature to investigate when studying NTE. In the following of this section I will shortly describe each mechanism before mentioning some interesting materials showing NTE and describing the applications of this behavior.

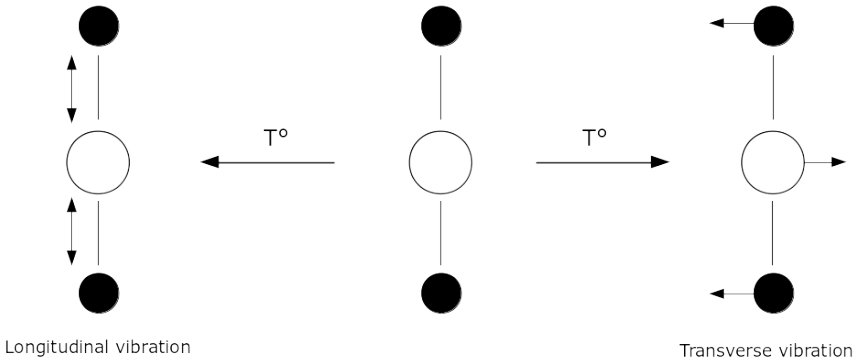


FIGURE 2.2: Schematic of the two different types of vibrations inside a crystalline material, which can be activated by an increase in temperature.

*Low frequency phonon modes:* A phonon mode describes the uniform vibration of a network with a particular frequency and amplitude. They are also often called "normal modes" and are quantified modes of vibration. To each mode of a structure corresponds a frequency and thus an energy required for the mode to be activated, for example through heating. There are two types of vibration modes competing in a structure: longitudinal and transverse modes. Both modes are represented in fig. 2.2. In longitudinal vibrations, atoms move in the same direction as the bond, causing the bonds to lengthen when these type of modes are activated. In metal oxides, longitudinal modes would increase the length of M–O bonds increasing in turn the M···M distance resulting in a global expansion of the structure. The transverse modes on the other side can sometimes reduce the M···M distance.

A useful formalism has been established by Grüneisen[108, 109]:

$$\gamma_i = -\frac{V_0}{\omega_i(V_0)}\frac{\partial \omega_i(V)}{\partial V} \tag{2.3}$$

With $\gamma_i$ being the mode Grüneisen parameter of mode $i$, $V_0$ the equilibrium volume and $\omega_i$ the frequency of mode $i$. The mode Grüneisen parameter links the vibration's frequency with the volume, allowing to study the direct effect of the mode in the thermal expansion. It is also possible to formulate a global Grüneisen parameter by adding the weighted contribution to the heat capacity of all the mode Grüneisen parameters as follow:

$$\gamma = \frac{\sum_i \gamma_i C_{V,i}}{\sum_i C_{V,i}} \tag{2.4}$$

Where $C_{V,i}$(T) are the partial vibrational contributions to the heat capacity of each mode $i$. This global Grüneisen parameter is linked to the thermal expansion $\alpha$, the bulk modulus $K_T$, and the constant volume heat capacity $C_V$ through the relation:

$$\alpha_V = \frac{\gamma C_V K_T}{V} \tag{2.5}$$

Looking at a specific mode $i$, a decrease of frequency when the volume decreases results in a negative value of $\gamma_i$. Considering eq. (2.4) and eq. (2.3), this means that the overall contribution of such a mode to the thermal expansion is negative. This is often depicted as a "guitar string" effect: If one plucks a guitar string (or excites a transverse mode that leads to NTE) and then stretches it, the note will move to higher frequency. These types of modes are thus prone to cause NTE. In order to understand the thermal expansion of complex systems, one has to look at the population of such modes at a given temperature to determine which type of modes are dominating. Transverse modes are often dominating at low temperatures as they generally have a much lower excitation energy than longitudinal modes, which means that they need less energy to be populated. Thus transverse modes usually dominate the Grüneisen parameter at low temperature and this leads to NTE. The best everyday example to this would probably be ice (see fig. 2.3). Ice is known to have a lower density than water, leading ice to float on water. But even ice has been shown to exhibit NTE at very low temperature. This is probably due to the transverse modes observed by Tanaka[110] being the only one populated at those temperatures.

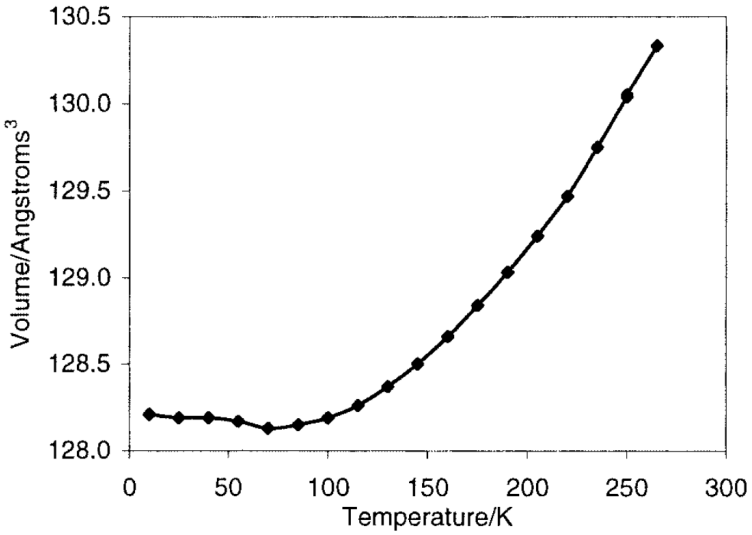


FIGURE 2.3: Thermal expansion of ice. Taken from [103]

*Rigid Unit Modes:* The idea of RUM comes originally from the work of Megaw[111] who described the transition between the $\alpha$ and $\beta$ phases of quartz with a model considering coupled rotation of relatively rigid corner sharing $SiO_4$ tetrahedra. This idea was then further extended[112, 113, 114, 115] and a general model to describe the negative thermal expansion emerged. A typical vibration corresponding to a RUM is represented in fig. 2.4. The primary building units of a system (i.e. the octahedra or tetrahedra) are generally stiff and require a high energy to distort due to their strong M–O bonds. However in a structure of corner sharing polyhedra like the one depicted in fig. 2.4, the polyhedra are linked together by M–O–M hinges

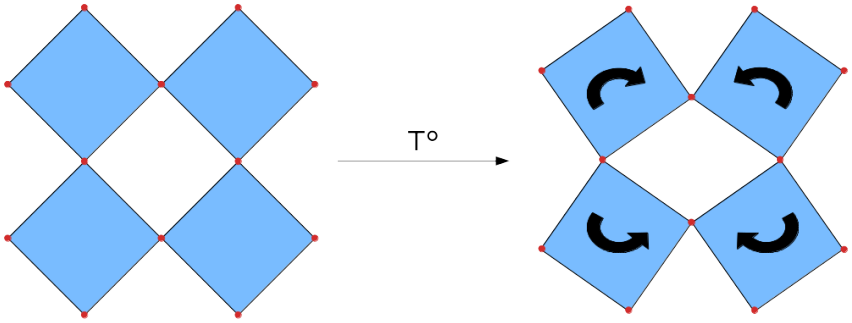


FIGURE 2.4: Schematic of a Rigid Unit Mode in a 2-dimensional representation.

which are much weaker and can easily rotate. These low energy rotations lead to the coupled rotation of polyhedra without any distortion inside them and are often (but not always) the lowest energy vibrations of the structure and have large negative Grüneisen parameters. Such vibration modes have been termed "Rigid Unit Modes" or "floppy modes". Due to the coupled rotation of polyhedra they induce a reduction of second-nearest-neighbor distance, leading then to a linear or global NTE.

The global thermal expansion depending on the relative contribution from all phonons of the structure, not all materials having RUMs or low energy phonon modes with negative Grüneisen parameters will show NTE. However it is a necessary feature for NTE behavior as well as a phonon band gap that separates the low energy modes with negative $\gamma_i$ from the high energy modes[116]. Among the low energy phonon modes, along the RUMs, we can also find quasi RUMs (QRUMs) which are very close to the RUMs in terms of vibrations. The only difference is that QRUMs have a low energy instead of a near-zero energy. This causes the QRUMs to exhibit small distortions of the polyhedra along with their rotation. Pryde et al.[117] demonstrated this in $ZrV_2O_7$ where the coupled rotations of tetrahedra induced a small distortion of the $ZrO_6$ octahedra. QRUMs generally arise when the rigidity of the polyhedra is decreased. This rigidity is linked to the strength of the anion–anion repulsion in oxides and thus, reducing their repulsion leads to the softening of polyhedra and likely to the emergence of QRUMs. This was done on $A_2M_3O_{12}$-type structures by increasing the size of the "A" cation increasing the oxygen-oxygen distances and thus reducing their repulsion[118, 119]. This specific family of materials cannot exhibit a rotation of the polyhedra without a distortion at the same time. It was found that the reduction of rigidity due to the substitution with bigger cations allows for more pronounced distortions, provoking in turn more rotation of the polyhedra, which leads to an increase of the NTE behavior.

*Phase transition:* There are some structures having rigid polyhedra and yet not showing NTE due to the presence of RUMs. In this case, NTE can arise from other factors such as phase transitions. Many structures undergoing phase transition at a specific temperature show NTE. Examples of this behavior include the ferroelectric-paraelectric phase transition of $PbTiO_3$[120], the zeolite ferrierite[121] or the $A_2(MoO_4)_3$ -type molybdate[122].

As mentioned previously, RUMs are low energy vibrations of the systems which correspond to coupled rotation of rigid polyhedra. There is generally one key vibration responsible of the phase transition and it provokes the phase change when

it becomes unstable due to reaching the stability limit. As the system approaches the temperature of the phase transition, the energy of the mode gets closer to zero and upon reaching the limit, the system becomes unstable and changes to a stable form. Depending on the type of phase transition, this can be accompanied by an increase or decrease of volume. For example, first order displacive phase transitions usually feature a rapid increase of volume followed sometimes by a contraction of the structure.

### Materials exhibiting negative thermal expansion

Over the course of the years NTE became an important subject of research as many materials showing NTE were found, gradually increasing the interest for this topic. This is demonstrated by the fig. 2.5 which shows a steady increase of the number of publications every year and supported by the fact that special issues of journals have been dedicated to this topic[123]. Among the NTE materials, $ZrW_2O_8$ is one of the most famous and studied, as it was found to show a constant isotropic contraction over its entire stability range[124, 125]. It was first synthesized around 1960[126] by mixing and heating $ZrO_2$ and $WO_3$ followed by quenching in air or water. At that time it was already found that this material was metastable at temperatures lower than 1420 K and it was found later that it is thermodynamically stable around 1440 K. However quenched cubic $ZrW_2O_8$ is kinetically stable from to 2 K up to 1050 K. This metastability can probably be explained by its framework structure which is represented in fig. 2.6. It is composed of $ZrO_6$ octahedra linked to $WO_4$ tetrahedra. All of the octahedra share their 6 corners with the tetrahedra but the tetrahedra are only linked to 3 octahedra. This means that each tetrahedron has a single-coordinated oxygen which is very unusual among oxides. This would explain why $ZrW_2O_8$ is thermodynamically stable at only high temperatures.

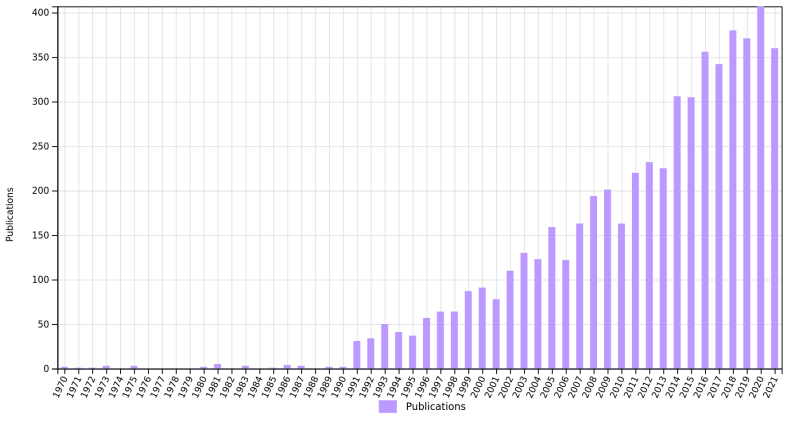


FIGURE 2.5: Number of publications per year based on a Web of Science search for "Negative thermal expansion". It should be noted that some early publications related to the field are missing as they were not using the NTE term as a keyword yet.

$ZrW_2O_8$ exhibits a quite significant NTE with a value of -9.07 x $10^{-6}$ $K^{-1}$[103], a contraction similar to the expansion of conventional ceramics. What is interesting in this material is that its contraction is isotropic, a behavior that is not interrupted by the phase transition it undergoes at around 450 K. During this phase transition the oxygen atoms reorient themselves to connect with their nearest neighbour, resulting in oxygen migration. The fact that NTE is isotropic through its whole stability range

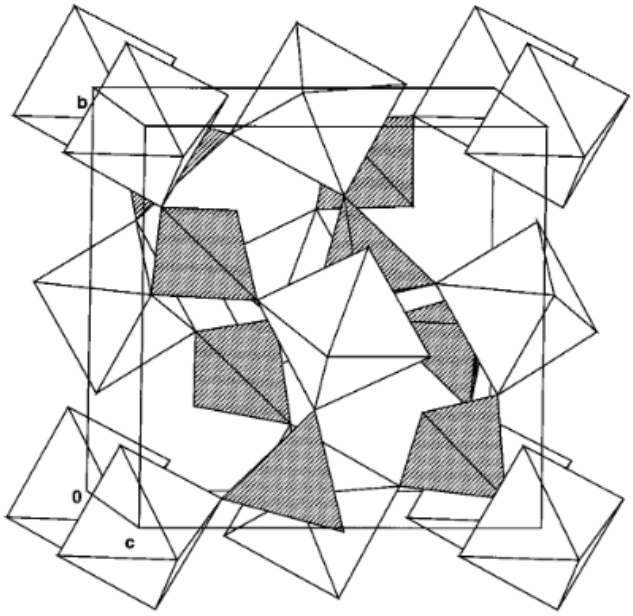


FIGURE 2.6: Representation of the $ZrW_2O_8$ framework. $ZrO_6$ octahedra are represented in white and $WO_4$ tetrahedra are shaded. Taken from [124]

is of particular interest for various applications, especially for the design of composite materials. Its NTE behavior was found to arise from the presence of RUMs and QRUMs at low frequencies[127]. As seen previously, RUMs and QRUMs consist in coupled rotations of octahedra and tetrahedra resulting in a reduction of the cell parameters. NTE in this material is thus caused by a family of low frequency vibrational mode inducing the coordinated rotation of polyhedra.

Other materials closely related to the $ZrW_2O_8$ are the $AM_2O_7$ family as their structure are very similar. Indeed the only difference lies in the $MO_4$ tetrahedra which are arranged in units of $M_2O_7$, getting rid of the M–O termination in the tetrahedra. This new complex structure has been found to allow some QRUMs like for $ZrW_2O_8$ but no RUMs[117]. Their NTE is therefore best described as transverse vibrations of corner-sharing oxygen atoms involving some distortions of the polyhedra. Therefore the $AM_2O_7$ generally presents a weaker NTE. An example of material of this family is the well-known $ZrV_2O_7$[128] which shows NTE at temperatures above 375 K. Under this limit, the structure undergoes a phase transition and the resulting new thermal becomes positive[129].

Another interesting family of material showing NTE is the $A_2W_3O_{12}$ family (A being a trivalent cation such as Al, Sc or Y). Their structure are also close to the two families seen previously as they are composed of corner-sharing $AO_6$ octahedra and $WO_4$ tetrahedra. This family drew some attention due to its wide range of possible compositions allowing for the tuning of expansion behavior[130] as well as a wide range of thermodynamic stability. In contrast with the previous materials, this family does not crystallize in a cubic framework but in an orthorhombic space group. This leads to an anisotropic thermal expansion with a contraction along $a$ and $c$ axes and an expansion along the $b$ axis. NTE in this case happens through the same mechanism of coupled rotation of polyhedra, however it has been observed that this was only possible if it was accompanied by some slight distortions of them[119]. The rigidity

of a tetrahedron being based on O–O repulsions, the size of the cation 'A' is of critical importance when one wants to improves the NTE behavior. Smaller tetrahedra will lead to stronger O–O interaction and thus to more rigid polyhedra and a weaker NTE behavior. Using bigger cations will result then in a more pronounced NTE as the ability of the structures to deform will increase. As a result, the coefficients of thermal expansion in this family are strongly dependent on the choice of cation.

There are many more types of materials exhibiting NTE, and zeolites occupy an important place. Yet, as zeolites are the main subject of this chapter, the thermal expansion and the previous studies on this topic will be discussed later in this part.

**Applications**

Thermal expansion is a phenomenon happening in all solids but its amplitude and sign depend on the material considered. This behavior needs to be taken into account when elaborating an application for any material, as it is a crucial factor when the temperature variates or when high accuracy is needed. Examples of this are bridges which undergoes changes of temperature through the year and especially between summer and winter, or high accuracy measurement equipments that need precise positioning of their components. It can also be problematic in certain scientific fields such as thin films where a difference of thermal expansion between the substrate and deposit can cause a misfit strain and a difference in the properties observed compared to the expected ones[131, 132]. All of these categories of applications are impacted by the thermal expansion and need to address this problem. This is where NTE materials come in handy: The principal interest of NTE materials is to build composites with tailored thermal expansion. The goal with composites is to be able to obtain a material with the properties of interest remaining unchanged and a tuned thermal expansion for the targeted application. In theory, NTE materials can be used to create composites with zero thermal expansion coefficients or to offset the thermal expansion to a desired value. In practice, composites are much more complex and represent a interesting challenge of engineering and design.

There are several aspects to consider when using NTE materials for composites. Firstly, the type of thermal expansion: the use of isotropic expansion materials is desirable in most cases but anisotropic expansion can be preferred when working with a highly oriented material or in cases where a compensation is needed in a specific direction (for example for thin films). This requires then the use of NTE materials with particular expansion along the directions of interest. Secondly, the stability of the composite is obviously necessary during the formation process and during its use but it also has to be compatible with the other components of the device. Some of these composites are also sensitive to moisture and thus must be stored in sealed environment to prevent the formation of hydrates through the contact with ambient air. Water molecules can be removed by heating but repeated hydration and dehydration cycles will lead to a faster deterioration of the composite. The next important aspects concern the production: except for zeolites, all materials showing NTE contain transition metals which raise the global price of production. Indeed, the materials discussed in the previous section all contains V, Zr or W atoms which are much more expensive than aluminosilicates such as zeolites. This additional cost does not usually weigh much in the balance for a high end application which already has a high cost of manufacturing. However, this cost is not negligible in the case of production of cheaper devices. Moreover, one has to consider the commercial availability of NTE materials being limited in choice. Zeolites, as well as $ZrW_2O_8$,

are readily available but it is difficult to find other commercial NTE materials. Nevertheless, it is possible to synthesize NTE materials for industrial use. For example $ZrV_2O_7$ can be easily obtained through traditional methods for ceramics[128, 133]. In contrary, $ZrW_2O_8$ family is metastable at room temperature and requires rapid quenching[134]. Solution based routes have also been reported for all families of NTE materials, providing the advantage of particle size control[135].

Besides the stability and overall cost of NTE materials which are important factors for industrial use, their properties are of course of central interest. Their amplitude of NTE has to match the requirements for the desired composite and their range of NTE behavior should include the room temperature and the range of temperatures at which the composite will be used. The stability with respect to external pressure also plays a role in choosing NTE materials as many of them exhibit mechanical instability at low pressure. These instabilities can take the form of phase transitions or amorphization. This is the case for $ZrW_2O_8$ which shows an irreversible phase transition at 0.2 GPa from cubic to orthorhombic[136]. This orthorhombic phase has a NTE at low temperatures but a positive one at room temperature which is not desirable for composites. Further compression of $ZrW_2O_8$ leads to pressure-induced amorphization starting at around 1.5 GPa[137].

Composites are the main field of study for negative thermal expansion and NTE materials are mostly considered as fillers improving the thermal behavior of already existing matrices. As $ZrW_2O_8$ is commercially available and is showing an isotropic NTE it is often the material used for making composites. It has been incorporated in a large variety of systems such as metal composites (mostly with copper)[138, 139, 140], ceramics[141, 142, 143] and polymers[144, 145, 146, 147]. However, we can imagine other applications than composites for NTE materials. One example for zeolite could be the temperature-activated desorption of adsorbed molecules. Indeed, contraction of the structure with temperature would reduce its porosity and thus triggering desorption of particular molecules. It could also be used to restrict the access of the pores by shrinking them, allowing then only specific small molecules or even prohibiting adsorption.

#### Thermal expansion of zeolites

Among the materials showing NTE, zeolites were one of the firsts to be discovered with this behavior. Indeed the first observation of NTE behavior on zeolites was made on quartz and vitreous silica in 1907[148, 149]. Yet the subject was not studied further until the 90s when several works focusing on NTE reported this behavior in several zeolite systems. In 1995, Tschaufeser and Parker[150] reported the simulated thermal expansion coefficients for several siliceous and nonsiliceous zeolites as well as some aluminophosphates (AlPO). They used at that time a code called PARAPOCS which uses the quasi-harmonic approximation to model the thermal properties. This approximation will be thoroughly described in the next section. The computer resources being however less efficient than today, they had to use an empirical potential for their calculations. Nevertheless, as depicted in fig. 2.7, they found that some of the zeolites studied such as the faujasite show NTE behavior over the whole range of temperature studied (0–500 K) and some others like the sodalite have the complete opposite PTE behavior over the entire range. Their work at the time was quite significant as the detailed measurement of thermal expansion in situ was (and still is nowadays) complicated. In comparison, the simulation of

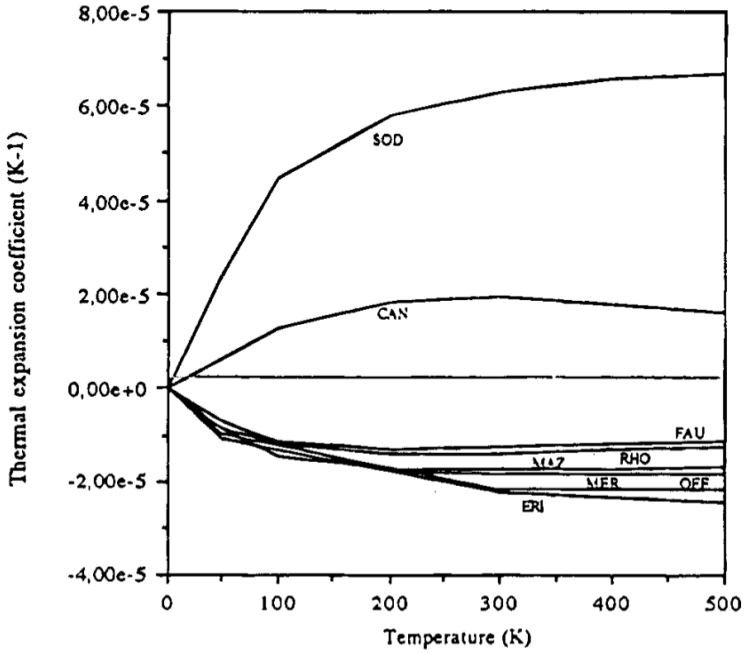


FIGURE 2.7: Simulated thermal expansion coefficients of several zeolites as a function of temperature. Taken from [150].

these properties is far more easier which is why the exploration of thermal behavior using computer techniques was pursued. This study was quickly followed by an experimental study on several siliceous zeolites in 1997[151]. In this work, Park and co-workers measured the thermal expansion coefficients of the pure silica zeolites MFI, DOH, DDR and MTN using X-ray Guinier technique. They found that all the as-synthesized structures have a global PTE, even if some directions exhibit a negative coefficient. They however calcined all of their compounds to reiterate their measurements on the calcined form and found different results: all of the frameworks showed PTE at low temperatures (until around 400 K to 600 K depending on the topology) but a transition occurs to all of them at high temperature leading to their high symmetry form and giving rise to NTE for the rest of the range (until 1000 K/1100 K depending on the framework). They attributed the NTE behavior to the rotations of tetrahedra without mentioning however the RUM model. They linked their work to the previous simulation study which predicted a correlation between NTE and the empty cavities of the structure. It shows that sample preparation, water removal and polymorphism is of great importance for the experimental measurement of thermal expansion. These two studies opened the field of research for zeolites.

In the next decade, the field was more focused on the experimental measurement of the thermal expansion rather than its computation. Thermal expansion coefficients of pure silica zeolites chabazite and ITQ-4 were determined by neutron powder diffraction[104]. It was found at the time that siliceous chabazite was the material with the strongest NTE behavior with a linear coefficient up to -16.7x$10^{-6}$ $K^{-1}$. The authors attributed the NTE to the rotations of Si–O–Si linkages but also mentioned the RUM model as a possible explanation. Another framework which caught a lot of attention is the MFI framework. MFI zeolites (also known as silicalite) are widely used as industrial catalysts as they show a high catalytic activity and an excellent shape selectivity. Its pure silica version (called silicalite-1) is also used as a

molecular sieve. An early molecular dynamic study[152] was able to reproduce the vibrational spectra as well as the already known temperature-dependent phase transition of the MFI framework from orthorhombic to monoclinic. They also found that its thermal expansion becomes negative at high temperature, suggesting the existence of a higher temperature orthorhombic phase at around 450 K. This transition was observed later by Park and co-workers[151]. The thermal expansion of MFI framework has been measured for different chemical compositions through high temperature X-ray diffraction[153, 154, 155]. The previous simulated and measured properties were confirmed as MFI frameworks show PTE until around 420 K and NTE above. Like previously (and as commonly found for zeolites), this NTE was attributed to transverse vibrations and rigid polyhedra rotations. Substituting different atoms to the Si atoms had various consequences: for example substituting Si with Zr or Fe has the same effect on the amplitude of the thermal expansion, both increasing the NTE behavior. However, while compounds with Zr exhibits the same PTE behavior between 298 and 420 K as siliceous MFI, Fe-substituted framework no longer shows PTE at low temperature.

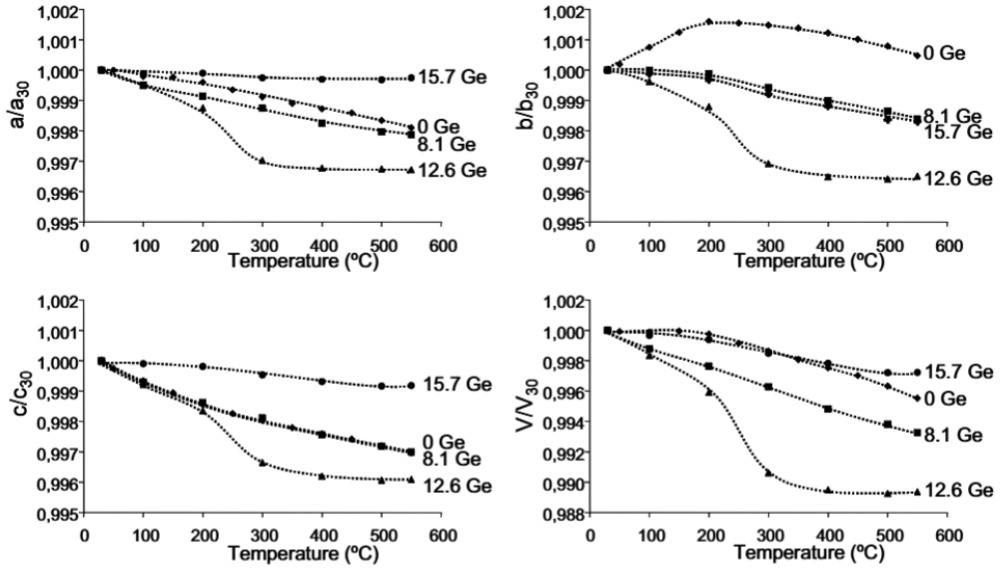


FIGURE 2.8: Variation of cell parameters and volume measured experimentally for four different compositions of zeolite ITQ-24. Taken from [156].

The effect of composition was also studied experimentally by Jordà and co-workers [156] who introduced Ge into the siliceous orthorhombic ITQ-24. The results on thermal expansion was not linear, as illustrated on fig. 2.8 taken from their work. The substitution with Ge increases the global thermal expansion until a certain point. They found that a composition of 15.7 Ge per unit cell results in a zero linear thermal expansion of the *a* parameter. Overall they observed a similar behavior between *a* and *c* parameters and a clearly anisotropic one along *b*. This study along with the studies on MFI showed that a change of composition can alter the thermal expansion of a material and that it can happen in an anisotropic manner. But it is not the only factor that affects this property in zeolite materials. Indeed zeolites are porous materials that can adsorb different species, filling the voids in their structure and thus impacting the thermal expansion. It is known that the rotation of tetrahedra responsible of NTE in zeolitic structures are possible thanks to the void space and could not

necessarily be possible in a dense framework. Filling this void with various species can lead to a reduction of NTE as demonstrated in the litterature[157, 158, 159].

### Calculation of thermal properties

Nowadays the calculation of ground-state energy in static calculations (in the 0K limit) are well mastered and quite common, even if progress is still made in this field. However, the computation of excited states as well as the simulations at T≠0 still represent a challenge as they require heavy computational processes. This usually prevents systematic studies of optical and thermal properties over a large number of compounds and generally limits the studies to some chosen materials. It also restricts the size of systems studied, as the size can tremendously increase the computational resources and time needed. Fortunately, over the course of time, approximations were made to render these types of calculations more affordable which, combined with the fast improvement of computer power, allowed for more and more materials to be characterized theoretically. Yet, while being possible today, the calculation of excited and thermal states is still a long process requiring a lot of computer resources. In this part I will outline the principal approximations and methods used to calculate thermal properties in this thesis.

In the previous Chapter we saw that, in the Born-Oppenheimer approximation, equations determining the electronic states and equations related to motions of nuclei are completely decoupled. Thus, the motions of atoms are studied with the approximation of the electronic state remaining constant in the ground state. The total energy of the system can then be written in the form of:

$$E = E_0 + V_{ion-ion} \tag{2.6}$$

with

$$V_{ion-ion} = \frac{1}{2}\sum_{n,m}\frac{Z_n Z_m e^2}{|R_n - R_m|} \tag{2.7}$$

and where $E_0$ is the ground state energy and $V_{ion-ion}$ is the electrostatic ion-ion interaction. Labeling the different nuclei with $\kappa$ and $\kappa'$ and the cartesian coordinates ($x$,$y$,$z$) with $\alpha$ and $\alpha'$, we can write the expression of the energy around its minimal value:

$$E = E_0 + \sum_{\kappa,\alpha}\left(\frac{\partial E}{\partial u_{\kappa,\alpha}}\right)u_{\kappa,\alpha} + \frac{1}{2}\sum_{\kappa,\kappa',\alpha,\alpha'} H^{\kappa,\kappa'}_{\alpha,\alpha'} u_{\kappa,\alpha} u_{\kappa',\alpha'} + \dots \tag{2.8}$$

with $u_{\kappa,\alpha}$ corresponding to a small displacement of the atom $\kappa$ in the direction $\alpha$. At the equilibrium, all residual forces are zero and thus the first-order derivative is null. The harmonic approximation neglecting the high-order terms of this equation, if we do not take into account the third and higher order terms, we are left with the equilibrium lattice energy $E_0$, and $H^{\kappa,\kappa'}_{\alpha,\alpha'}$ the second-order force constants (or Hessian) corresponding to the second derivative of the energy with respect to the displacement:

$$H^{\kappa,\kappa'}_{\alpha,\alpha'} = \left(\frac{\partial^2 E}{\partial u_{\kappa,\alpha}\partial u_{\kappa',\alpha'}}\right) \tag{2.9}$$

For a periodic structure, we can write the displacements $u_{\kappa,\alpha}$ in terms of a plane wave with respect to the cell coordinates:

$$u_{\kappa,\alpha} = \varepsilon_{m\kappa,\alpha q} e^{(iq.R_{\kappa,\alpha} - \omega_m t)} \tag{2.10}$$

with $q$ a phonon wave vector and $\varepsilon_{m\kappa,\alpha q}$ a polarization vector. This leads to an eigenvalue equation involving the dynamical matrix:

$$D^{\kappa,\kappa'}_{\alpha,\alpha'}(q)\varepsilon_{m\kappa,\alpha q} = \omega^2_{m,q}\varepsilon_{m\kappa,\alpha q} \tag{2.11}$$

with
$$D^{\kappa,\kappa'}_{\alpha,\alpha'} = \frac{1}{\sqrt{M_\kappa M_{\kappa'}}} \sum_a H^{\kappa,\kappa'}_{\alpha,\alpha'} e^{-iq.r_a} \tag{2.12}$$

This eigenvalue equation can be solved numerically if the dynamical matrix has been calculated, and the eigenvalues are obtained as the square of the vibrational frequencies. Second order force constants, which are needed to solve this equation, can be obtained through DFT calculations either by using the finite-displacement method or the perturbation theory.

This approximation, while being fairly simple, is quite powerful when one aims to study the lattice dynamics: the calculations required are cheap and the resulting accuracy is sufficient for most phenomena. Within such model, several quantities can be obtained accurately such as the frequencies of vibration modes or the heat capacity of the crystal. However the harmonic approximation has some limits: the anharmonic terms (like phonon-phonon interactions) not being included, some quantities cannot be obtained through this approximation. For example, in the harmonic approximation approach, thermal conductivity is infinite due to the infinite lifetime of phonons and thermal expansion is null as the vibrational part of the energy is not considered dependent on the volume. One way of correcting this is to include the higher order terms of Equation eq. (2.8) however doing so increases tremendously the computational cost. Another way of calculating some of these properties, including the thermal expansion, is through the so-called Quasi-harmonic approximation (QHA).

The QHA aims to introduce the dependence of phonons on the unit cell volume, which is missing in the simpler harmonic approximation, while keeping an affordable cost of computation. In practice the Helmholtz free energy is computed and an explicit dependence of phonon frequencies on volume is introduced:

$$F(V,T) = U_0(V) + F_{vib}(V,T) \tag{2.13}$$

where $F(V,T)$ is the Helmholtz free energy, $U_0(V)$ is the zero-temperature internal energy and the vibrational part of the free energy $F_{vib}(V,T)$ can be expressed as:

$$F_{vib}(V,T) = E_0(V) + k_B T \sum_k \ln\left(1 - e^{-\frac{\hbar\omega_k(V)}{k_B T}}\right) \tag{2.14}$$

with $$E_0(V) = \frac{1}{2}\sum_k \hbar\omega_k(V) \tag{2.15}$$

being the zero-point energy of the system. All of these quantities can be obtained through basic geometry optimization or harmonic vibrational calculations. Frequencies $\omega_k(V)$ have to be determined at several volumes in order to determine their dependence on volume, which is why QHA calculations require several computations of vibrational frequencies at different fixed volumes. Once this is known, one can determine the equilibrium volume at each desired temperature by minimizing eq. (2.13) with respect to the volume, keeping the temperature constant. This is usually done with an equation of state. In my case, I used the third-order Birch-Murnaghan equation of state, defined as:

$$E(V) = E_0 + \frac{9V_0K_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^3 K_0' + \left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}} - 1\right]^2 \left[6 - 4\left(\frac{V_0}{V}\right)^{\frac{2}{3}}\right]\right\} \tag{2.16}$$

Once the equilibrium volume is known at all desired temperatures, various properties can be obtained such as the thermal expansion:

$$\alpha_V(T) = \frac{1}{V(T)}\left(\frac{\partial V(T)}{\partial T}\right)_{P=0} \tag{2.17}$$

or the isochoric heat capacity:

$$C_V = -T\frac{\partial^2 F(V,T)}{\partial T^2} \tag{2.18}$$

The Grüneisen parameter is also obtainable once $\omega_k(V)$ values are accessible. It was described earlier in this chapter, and can be calculated through its definition in eqs. (2.3) to (2.5).

### 2.1.2 Mechanical properties

#### Mechanical properties of zeolites

Although they were not as much investigated as the thermal properties, some mechanical properties were obtained from the calculations. The term "mechanical property" refers to the physical properties of a material that define its reaction to applied forces. There are several properties describing the behavior of a material in different cases. The most general one is the bulk modulus $K$. It describes the reaction upon applying uniform hydrostatic pressure on the whole structure. It can be thought of the ratio between the pressure increase and the resulting decrease of volume. Except in very specific cases[160], the value of bulk modulus is positive as a negative value (meaning the material would expand upon compression) is thermodynamically forbidden. Typical example values include diamond with a value of around 450 GPa and sodium chloride with a value of around 25 GPa. There are many other mechanical properties such as the shear modulus, young's modulus or Poisson's ratio. However, only the bulk modulus and its derivative were considered in this study as

they are readily obtained from the QHA calculations without any additional computational cost.

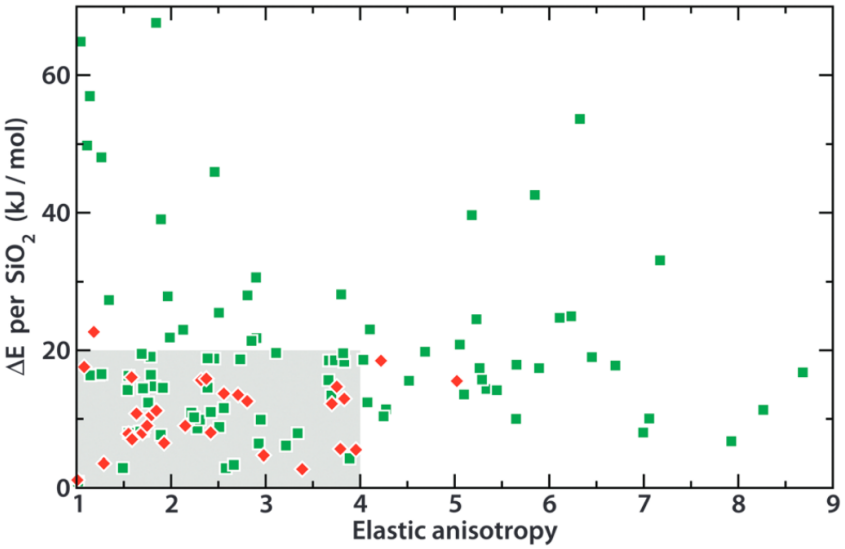


FIGURE 2.9: Elastic anisotropy of pure silica zeolites against lattice energy relative to $\alpha$-quartz; red points correspond to synthesized pure silica zeolites. The grey area corresponds to the feasibility criterion proposed proposed by Coudert. Taken from [161].

Considering that zeolites are inorganic solids, they overall tend to have good mechanical properties and a good mechanical stability. A few experimental determinations of mechanical properties were performed on a limited number of frameworks [162, 163]. It resulted that the properties were strongly dependent on the type of framework. This trend was also observed theoretically on a small number of siliceous zeolites[164], as they calculated the bulk modulus for their structures and found a wide range of 46 to 93 GPa. However, no systematic studies on a large number of zeolites were performed until 2013 when Coudert[161] investigated the mechanical behavior of 121 all-silica frameworks. He used static DFT simulations to determine several mechanical quantities such as the young's modulus and shear modulus. This large-scale study proved the previous observation that the topology of the framework influences strongly the resulting mechanical properties. He also showed that the previously established feasibility criterion of being in the $+30\ kJ.mol^{-1}$ range compared to the $\alpha$-quartz could be extended to include the elastic anisotropy. Indeed, as depicted by the grey area in fig. 2.9, he demonstrated that the topologies that were already synthesized had a low elastic anisotropy as well as a low lattice energy (compared to the $\alpha$-quartz). The idea of the present work is closely linked to the past study of Coudert, but in this case it is mainly focused on thermal properties. Later, Evans et al.[165] performed machine learning simulations using the results obtained by Coudert. He used the 121 zeolite structures as a training set and built a model based on geometric descriptors. By doing so, he highlighted several correlations between the characteristic features of a zeolite and the elastic properties. He also used his model to predict the elastic properties of a large data set of hypothetical zeolites. He found that the less dense a zeolite, the lower the shear resistance. He observed as well that most of the hypothetical structures have a lower stiffness and shear resistance than the IZA structures.

**calculations of mechanical properties**

Besides the thermal properties, the quasi-harmonic approximation allows to compute many other characteristics including the mechanical properties. Indeed, as this method gathers energy/volume data, one can obtain the bulk modulus $K$ using an equation of state. I already used the third-order Birch-Murnaghan equation of state in the quasi-harmonic approximation for the determination of the equilibrium volume. Therefore, the zero-temperature bulk modulus $K_0$ and its derivative $K_0'$ were directly obtained from this method. The temperature-dependent bulk modulus $K(T)$ was also retrieved for the whole range of temperature considered from the minimized Helmholtz free energy $F(V,T)$. It corresponds to the isothermal second derivative of $F(V,T)$ with respect to the volume:

$$K(T) = V(T)\left(\frac{\partial^2 F(V,T)}{\partial V^2}\right)_T \tag{2.19}$$

It can be noted that QHA allows for the combination of pressure and temperature effects on structural and elastic properties of materials. Indeed the $P-V-T$ relation can be obtained easily. The thermal pressure can be readily calculated from the Helmholtz free energy as:

$$P(V,T) = -\frac{\partial F(V,T)}{\partial V} \tag{2.20}$$

From this, the bulk modulus can be obtained at specific pressures and temperatures. By using eq. (2.13) and exploiting eq. (2.20), it can be obtained again as the isothermal second derivative of the Helmholtz free energy with respect to the volume:

$$K(P,T) = V(P,T)\left(\frac{\partial^2 F(V(P,T),T)}{\partial V(P,T)^2}\right)_T \tag{2.21}$$

Finally, in order to characterize the temperature evolution of the bulk modulus $K(T)$, the temperature coefficient $\delta$ was defined as:

$$\delta = \frac{1}{K(T)}\left(\frac{\mathrm{d}\,K(T)}{\mathrm{d}\,T}\right) \tag{2.22}$$

## 2.2 Systematic study

### 2.2.1 Definition of parameters and preliminary calculations

**The automated CRYSTAL17 algorithm**

The application of the calculation of thermal properties described earlier has been realised with the CRYSTAL17 software. In its lastest version, the software incorporates an automated algorithm for conducting QHA calculations[166, 167, 168, 169]. I used this algorithm to perform the work reported in this chapter which is why I will start by describing it.

In practice, the user can input an already optimized structure or a structure to optimize before the QHA routine. Then the user gives the information needed on the

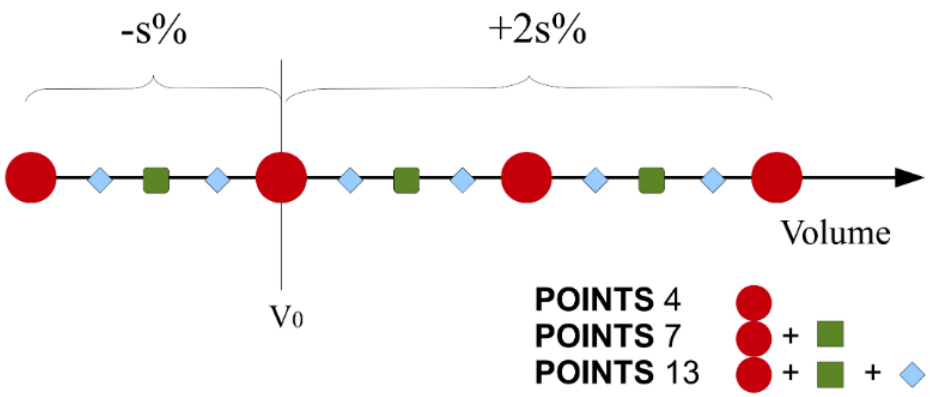


FIGURE 2.10: Definition of volume range and number of points. Taken from [170].

volume by entering the desired number of volume steps for the computation (either 4, 7 or 13) and a scale parameter $s$. This parameter will define a range of volume from a $-s\%$ compression to a $+2s\%$ expansion with a total of 4, 7 or 13 volume values in between (including the extrema) depending on the user's choice. As the computation of harmonic thermodynamic properties requires the knowledge of phonon modes over the complete first brillouin zone (FBZ), it is necessary to increase the sampling of the phonon dispersion in the FBZ. This can be achieved by building a supercell, as the size of the supercell corresponds to the sampling of the FBZ. Supercell information is then also required in the form of a matrix of integers describing the lattice vectors of the supercells to be created. The calculation will then optimize the structure at each different fixed volume and calculate the corresponding frequencies of vibration. This step is the most expensive in terms of computational resources, as it requires as many frequency calculations as the number of different volumes used. The need for a supercell also slows down the calculation due to the increased size of the system. As soon as all the frequencies are obtained, they are individually fitted as a function of volume with polynomial functions of different orders. From these fitted frequencies, the Helmholtz free energy can be obtained at any $T$ from eqs. (2.13) and (2.14). Minimizing this Helmholtz free energy with respect to the volume using an equation of state leads to the equilibrium volume and allows to compute the thermal expansion.

#### Main parameters

All the calculations described in this Chapter were performed under the density functional theory as implemented in the CRYSTAL17 code[171, 170]. All-electron localized basis sets were used for both silicon and oxygen. They can be found on the CRYSTAL17 online library with the acronyms and references as follow: Si_88_31-G*_nada_1996[172] for silicon and O_8_411_muscat_1999[173] for oxygen. After testing several functionals, I decided to choose the PBEsol0[174] hybrid functional for the treatment of exchange and correlation and did not use dispersion correction. This choice is only mentioned here but will be properly explained in the next part.

Mesh sampling was realized by using the Monkhorst-Pack scheme[175] with a $k$-point mesh of $3 \times 3 \times 3$. For all the structures, supercells were generated in order to contain 150 atoms at most to ensure convergence of the thermal properties within the available CPU time. Geometries were optimized using CRYSTAL17 code with the

default convergence criteria (a maximum of 0.00045 a.u. on atomic displacements during one optimization step and 0.0018 a.u. on forces). Representative input files and optimized structures are available online at https://github.com/fxcoudert/citable-data/tree/master/137-Ducamp_JPhysChemC_2021.

Starting structures were taken from the International Zeolitic Association (IZA) database available at http://www.iza-structure.org/databases/, considering only fully ordered frameworks, of which there are currently 244 known. As ANO and PTO frameworks were added only recently (the $16^{th}$ of July 2021) and were not present in the database when this work was performed, I only considered the 242 topologies that were accepted at that time. Zeolites with unit cell containing more than 150 atoms were not considered, resulting in a set of 190 zeolites for which I optimized the geometry with a pure $SiO_2$ composition. 134 structures were found to achieve convergence within the time constraints imposed. Quasi-harmonic approximation calculations were conducted with the automated QHA algorithm implemented in CRYSTAL17: while keeping the symmetry, several isotropic deformations are performed on the lattice constants and the energies and vibration modes are computed for each resulting structure. After direct fitting of the phonon frequencies to the volume, the properties of interest can be derived as shown in the previous part. I used the algorithm with a volume deformation range of $-1.5\%$ to 3% with 4 different volumes and computed the thermal properties over the range of 10 K to 300 K. Among the 134 structures, 120 were able to converge within the volume and temperature range.

#### Preliminary calculations

When doing a large-scale study, it is important to set correctly the parameters of the simulation in a way that is suitable for all systems studied, to keep a consistent methodology. This step consists usually in a trade-off between accuracy and feasibility. In this case here, I had the possibility to use hybrid functional without creating a too high computational demand. I therefore started by analyzing different exchange-correlation functionals to select the most accurate. I looked at the B3LYP, PBE, PBE0, PBEsol, PBESsol0 functionals. I also investigated the effect of long-range dispersion corrections on the results. They were introduced with the method proposed by Grimme[97]. However, PBEsol and PBEsol0 not being parametrized for this method, I did not introduce these corrections for them. I selected the Linde Type A (LTA) framework to study these functionals as it has a fairly small unit cell and experimental data are available for a pure silica form of this topology[176]. I conducted optimizations of geometry of this framework with all the different functionals and compared the lattice parameters I obtained. The results are shown in table 2.1.

What can be seen first is that all the results are somewhat really close falling in the range of 12 to 12.15 Å. The difference observed with the experimental value is then constant and characteristic of GGA functionals such as PBE and PBEsol which tend to show a higher lattice parameter than experimental measurements. The discrepancy between the calculated and measured lattice parameter is of around 1.3% which could be considered high for a high accuracy study but is more than sufficient for a large scale one. This shift can have several explanations: it can come from the overestimation of the PBE-based functionals. The next factor could be a thermal effect as a larger lattice parameter for static calculation is in line with a negative thermal

expansion behavior. Finally the choice of basis set could also be at play in the difference observed as I chose a relatively small basis set for reasons of computational cost. A more complete basis set could improve the accuracy of the results but it is not affordable in terms of simulation time as a lot of calculations have to be done. Moreover, given the fact that I am aiming for the thermal properties and the evolution of cell parameters with respect to the temperature, I considered that this magnitude of discrepancy was acceptable.

| | Expt. | B3LYP | | PBE | | PBE0 | | PBEsol | PBEsol0 |
|---|---|---|---|---|---|---|---|---|---|
| | | No D3 | with D3 | No D3 | with D3 | No D3 | with D3 | | |
| LTA | 11.853 | 12.095 | 12.057 | 12.142 | 12.119 | 12.050 | 12.027 | 12.090 | 12.014 |

TABLE 2.1: Lattice parameter of the pure silica Linde Type A framework (LTA) calculated for different functionals with or without long-range dispersion corrections (DFT-D3 method). Experimental values of lattice parameters were taken from [176].

Comparing all functionals, I saw that the PBEsol0 functional shows the closest results with the experiment. The introduction of long-range dispersion correction improved a bit the results obtained with the other functionals but it still stayed less accurate than PBEsol0 without adding any correction to it. This shows that PBEsol0 is a really accurate functional for solids (or zeolites at least). It was also already used in a past study of the group focusing on $ZrW_2O_8$[177], a material known for its NTE behavior. I chose then to use this functional as it was the most accurate out of the group I initially defined. Considering the fact that I want to look at properties arising from linear responses to small perturbations (such as thermal expansion and bulk modulus), the impact of dispersion corrections is not crucial in this methodology, and they were not used.

In order to check the validity of my parameters, I decided to calculate the thermal expansion of some topologies of zeolites for which I could find experimental values of pure silica materials in the literature. I found experimental data of thermal expansion for chabazite (CHA), Linde Type A (LTA) and faujasite (FAU), therefore I calculated the thermal expansion of these 3 frameworks. I used the exact same methodology described in the previous part with the PBEsol0 functional and no Grimme-type corrections. The results are shown in table 2.2.

As observed previously, the calculated lattice parameters of FAU and CHA frameworks are slightly overestimated compared to the experimental measurements with the same order of magnitude than for LTA, respectively 1.6% and 1.9% for FAU and CHA. Results are then consistent going from one framework to another. Overall the accuracy on the lattice parameters is satisfying considering the scale of the study that we envision. Following the optimization, I conducted QHA calculations on these 3 topologies and compared the thermal expansion coefficients obtained at 300 $K$ to reported coefficients in the literature. It can be seen that the agreement with experimental values is very good, showing a value of $-7.3 \times 10^{-6}\ K^{-1}$ for the LTA framework which is very close to the value of $-7.4 \times 10^{-6}\ K^{-1}$ reported experimentally by Carey et al.[176] between 100 $K$ and 300 $K$. This good agreement is confirmed with the two others frameworks: I found a value of $-4.9 \times 10^{-6}\ K^{-1}$ for FAU which is very close to the values of $-3.92 \times 10^{-6}\ K^{-1}$ and $-4.2 \times 10^{-6}\ K^{-1}$ measured experimentally[178, 105]. The same is observed for CHA with values of

| CODE | Lattice parameter (Å) | | Thermal expansion coefficient ($10^{-6}$ K$^{-1}$) | |
|---|---|---|---|---|
| | Expt. | Calc. | Expt. | Calc. |
| LTA | 11.853 [176] | 12.014 | −7.4 [176] | −7.3 |
| FAU | 24.256 [178] | 24.650 | −3.92 [178]<br>−4.2 [105] | −4.9 |
| CHA | $a = 13.525$ [104]<br>$c = 14.734$ [104] | $a = 13.671$<br>$c = 15.015$ | −0.5 [104] | −0.6 |

TABLE 2.2: Lattice parameters and linear thermal expansion coefficients of pure silica zeolites with LTA, FAU, and CHA frameworks. Calculated values of thermal expansion coefficients are given at 300 K.

$-0.6 \times 10^{-6}\ K^{-1}$ and $-0.5 \times 10^{-6}\ K^{-1}$, respectively for my calculations and the experiment. These three examples show that the chosen method is valid for the study and it comforted me in my choice of parameters for the DFT and the QHA methodology. These results allowed me to validate the methodology and to start the study on the whole database.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABW | ACO | AEI | AEL | AEN | AET | AFG | AFI | AFN | AFO | AFR | AFS | AFT | AFV | AFX |
| AFY | AHT | ANA | APC | APD | AST | ASV | ATN | ATO | ATS | ATT | ATV | AVE | AVL | AWO |
| AWW | BCT | BEC | BIK | BOF | BOG | BOZ | BPH | BRE | BSV | CAN | CAS | CDO | CFI | CGF |
| CGS | CHA | -CHI | -CLO | CON | CSV | CZP | DAC | DDR | DFO | DFT | DOH | DON | EAB | EDI |
| EEI | EMT | EON | EPI | ERI | ESV | ETL | ETR | ETV | EUO | EWO | EWS | EZT | FAR | FAU |
| FER | FRA | GIS | GIU | GME | GON | GOO | HEU | IFO | IFR | -IFT | -IFU | IFW | IFY | IHW |
| IMF | IRN | IRR | -IRY | ISV | ITE | ITG | ITH | ITR | ITT | -ITV | ITW | IWR | IWS | IWV |
| IWW | JBW | JNT | JOZ | JRY | JSN | JSR | JST | JSW | KFI | LAU | LEV | LIO | -LIT | LOS |
| LOV | LTA | LTF | LTJ | LTL | LTN | MAR | MAZ | MEI | MEL | MEP | MER | MFI | MFS | MON |
| MOR | MOZ | MRT | MSE | MSO | MTF | MTN | MTT | MTW | MVY | MWF | MWW | NAB | NAT | NES |
| NON | NPO | NPT | NSI | OBW | OFF | OKO | OSI | OSO | OWE | -PAR | PAU | PCR | PHI | PON |
| POR | POS | PSI | PTT | PTY | PUN | PWN | PWO | PWW | RHO | -RON | RRO | RSN | RTE | RTH |
| RUT | RWR | RWY | SAF | SAO | SAS | SAT | SAV | SBE | SBN | SBS | SBT | SEW | SFE | SFF |
| SFG | SFH | SFN | SFO | SFS | SFW | SGT | SIV | SOD | SOF | SOR | SOS | SOV | SSF | SSY |
| STF | STI | STT | STW | -SVR | SVV | SWY | -SYT | SZR | TER | THO | TOL | TON | TSC | TUN |
| UEI | UFI | UOS | UOV | UOZ | USI | UTL | UWY | VET | VFI | VNI | VSV | WEI | -WEN | YFI |
| YUG | ZON | | | | | | | | | | | | | |

FIGURE 2.11: List of the 242 fully ordered zeolites from the IZA database and summary of the calculations conducted. Red corresponds to synthesized structures in $SiO_2$ form, blue background corresponds to energy optimization and quasi-harmonic approximation calculations realized in this work and green background corresponds to energy optimization done in this work without QHA analysis due to computational limitations.

Among the 242 fully ordered structures of the IZA present at the time of the study, I eliminated all the structures that contain more than 150 atoms as they are too large to be calculated with the QHA. This resulted in 190 structures for which I optimized the geometry with a pure $SiO_2$ composition. Out of those, 134 were found to achieve convergence within the time constraints. I calculated the thermal properties for all of them with the automated CRYSTAL17 algorithm described earlier, which succeeded for 120 structures. The summary of all the structures considered, optimized or calculated through QHA is represented in fig. 2.11. The study was performed on resources from the Occigen supercalculator owned by the GENCI (Grand Equipement National de Calcul Intensif) and managed and maintained by the CINES (Centre Informatique National de l'Enseignement Supérieur). The production runs took around 2 months to complete.

### 2.2.2 Results

#### Volumetric thermal expansion

Among all the properties I obtained, I first focused on the volumetric thermal expansion coefficients. Within reasonable computational time, I was able to obtain the thermal properties of 120 pure silica zeolites listed in fig. 2.11. Here is shown on fig. 2.12 the thermal expansion of all these structures at 300 K compared to the respective Grüneisen parameters. It can be seen that all the pure silica frameworks studied here show NTE. This appears to be a common behavior in the zeolite family as NTE behavior was already observed in numerous past studies[104, 105, 151, 176, 178, 179, 180], both experimental and theoretical. The mechanism behind this phenomenon has been previously suggested for zeolites as being due to the presence of rigid unit modes[127, 181] arising from the corner-sharing nature of zeolite structures[182]. Here I demonstrate that this mechanism is actually completely generic, as all the 120 frameworks studied show NTE in their pure $SiO_2$ form.

Some works reported in the literature found that a few $SiO_2$ zeolite frameworks can show PTE[150, 151] as does also the nonporous $\alpha$-quartz phase[183]. However, such results are very rare in the literature compared to the number of reported frameworks showing NTE. Combined with my current simulations, this means that PTE behavior is more an exception than a norm for pure silica zeolites. Therefore the causes of PTE in materials such as the $\alpha$-quartz phase should be re-examined. It is possible that the NTE comes not from the geometry of the framework itself, but from the nanoporous nature of zeolites. The void in the structure could allow for the rotation of tetrahedra leading to the presence of RUM thus giving rise to NTE. This would explain why a dense framework such as the $\alpha$-quartz does not show this phenomenon. In this hypothesis, zeolites containing guest molecules would also show PTE as adsorbed molecules would effectively convert these porous frameworks to dense materials. This has been observed on the LTA framework where the adsorption of water changes the thermal expansion from negative to positive[157]. It can be noted that, however, I found no correlation for zeolites in the calculations between the density and the extent of thermal expansion as can be seen on fig. 2.13.

It can be seen also that the magnitude of NTE displayed is on the same order of magnitude as some simple oxides and other porous materials, like AlPOs and metal-organic frameworks[184, 185]. While the histogram in fig. 2.12 show that most zeolites can be found in the range $-2.3\ MK^{-1}$ and $-1.1\ MK^{-1}$ (using the notation $MK^{-1}$ instead of $\times 10^{-6}\ K^{-1}$ for the sake of clarity), the range of thermal expansion values is quite large which shows that the framework topology has a great impact on this property — even with a fixed chemical composition, as is the case here. This was already hinted at by Astala et al.[186] on a small number of zeolites and confirmed here with a much larger set of structures. Previous computational studies, through force-field based molecular dynamics, had drawn the conclusion that zeolites with only 1D channels tended not to show NTE behaviour, where as zeolites with 2D or 3D channels did[101, 150, 187]. The results here, obtained with a higher-accuracy methodology, do not confirm those trends. I found no systematic and simple correlation between the characteristics of the zeolites (channel type, accessible pore volume, etc) and the value of the volumetric thermal expansion coefficient.

In my set of data I have differentiated the theoretical $SiO_2$ structures and the ones that have experimentally been synthesized in a pure $SiO_2$ form. Those are represented respectively in green and red in fig. 2.12 and this color code will be kept

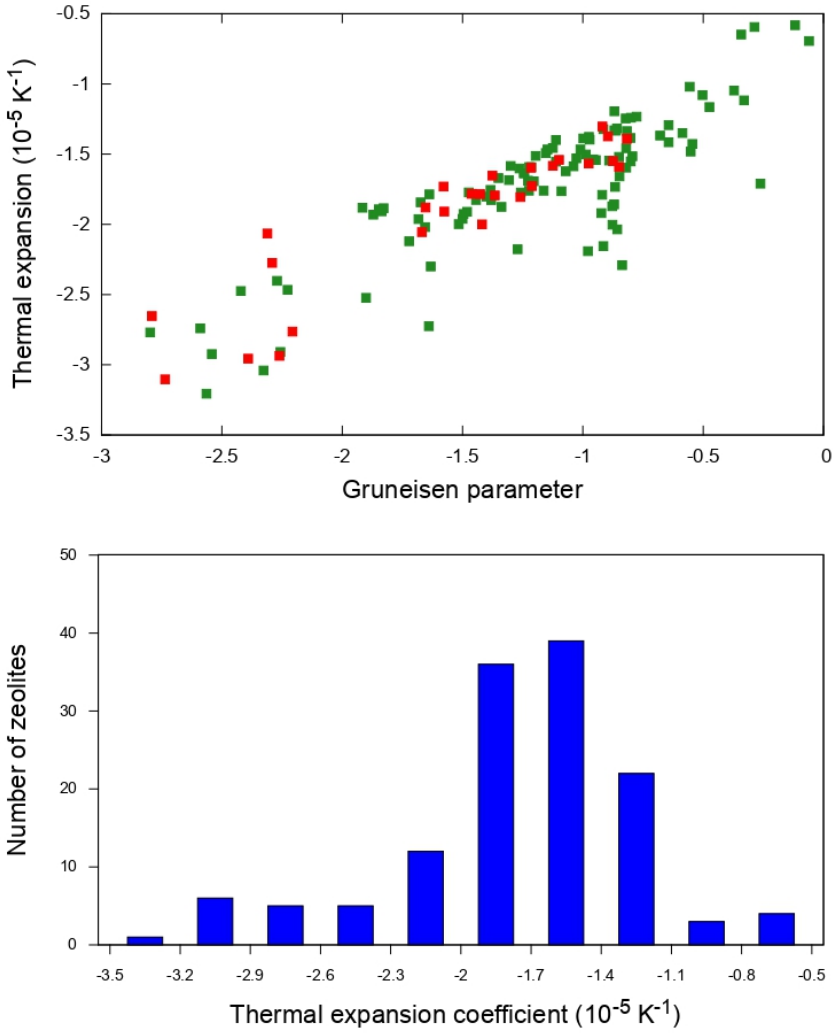


FIGURE 2.12: Top panel: Volumetric thermal expansion coefficients plotted against the mean Grüneisen parameter for each zeolite at 300 K. Green squares correspond to hypothetical zeolitic structures and red ones correspond to experimentally synthesized pure-silica frameworks. Bottom panel: Distribution of the values of thermal expansion coefficients for all the frameworks.

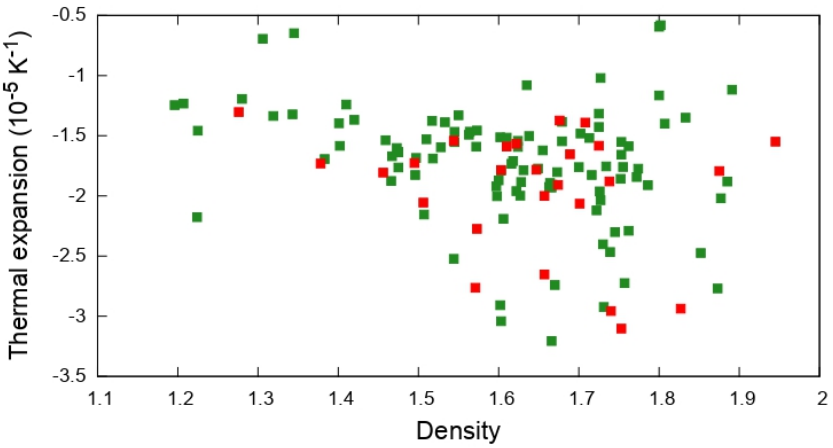


FIGURE 2.13: Thermal expansion coefficient compared to the density at 300K. Green squares correspond to theoretical structures and red squares correspond to experimentally synthesized ones.

through the entire chapter. The aim of separating the data in this fashion is to be able to determine synthesis conditions. However, it can be seen on the figure that no differences can be observed between synthesized and theoretical structures. I could observe that experimental frameworks cover almost the entire range, with perhaps a slight dominance towards the high-NTE region, although that could be due to the relatively small size of sampling of structures. This is in contrast with mechanical properties, where it was shown that the experimental feasibility of the frameworks could be linked to their mechanical stability[161].

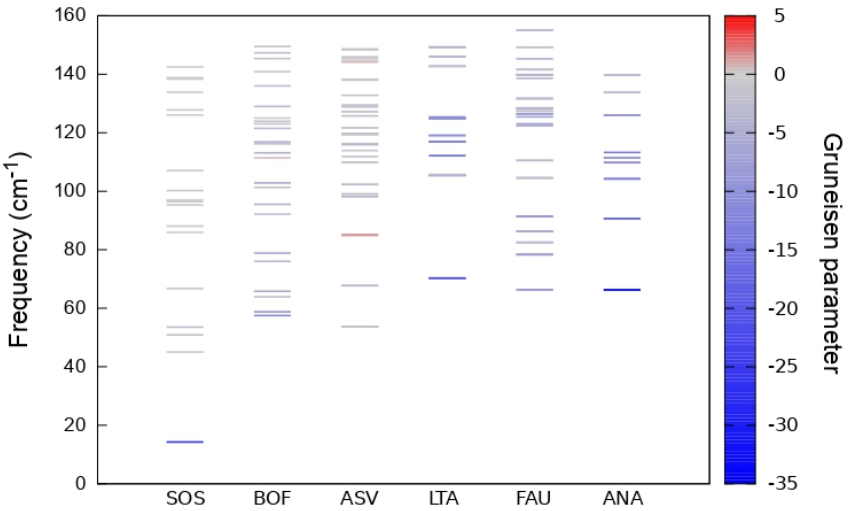


FIGURE 2.14: Representation of the mode Grüneisen parameters of 6 zeolites (SOS, BOF, ASV, LTA, FAU, ANA) for all the vibration modes under 160 $cm^{-1}$. Each line represent a vibration mode (possibly degenerate due to symmetry), and its color indicates the value of the mode Grüneisen parameter. The color scale is represented on the right.

Comparing the global Grüneisen parameter in fig. 2.12 with the thermal expansion, a

broad correlation can be observed between these two parameters over the temperature range studied. As observed previously for the thermal expansion, all the values of Grüneisen parameters are negative. It shows that the Grüneisen model for studying the coupling between the variations of volume and the vibrational properties is reasonably valid for systems such as zeolites. This validity is further confirmed with fig. 2.14 where I represented the mode Grüneisen parameters for the low-frequency vibration modes of several zeolites. We can clearly see from this representation that the negative global Grüneisen parameter is only influenced by a small number of strongly negative mode Grüneisen parameters associated with low-frequency vibrations. This was observed on the 6 frameworks presented here but confirmed on another number of frameworks. The modes associated with strongly negative Grüneisen parameters are found to have frequencies below 130 $cm^{-1}$, confirming the microscopic mechanism behind NTE as being linked to soft vibration modes involving coupled rotations of corner-sharing tetrahedra.

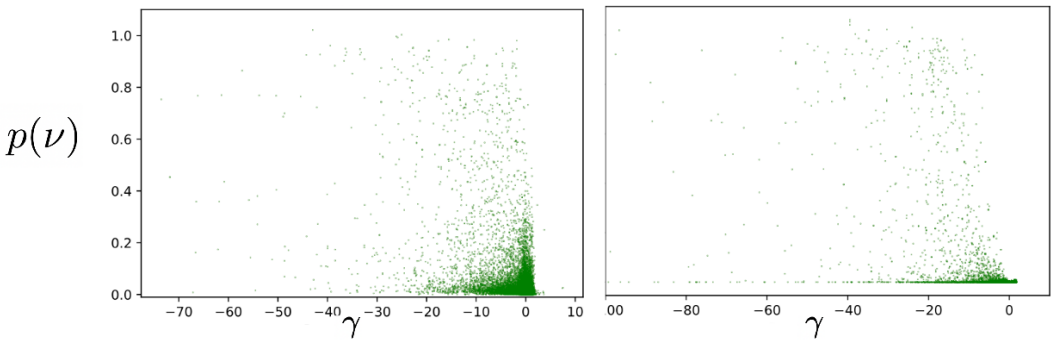


FIGURE 2.15: Frequencies for each modes of each zeolites against their corresponding mode Grüneisen parameters. The left panel corresponds to the results obtained with DFT and the right panel corresponds to results obtained with GULP.

During this study, I had the opportunity to work with Quentin Gueroult, a student realising an internship at the university of Oxford under the supervision of Pr. Andrew Goodwin. He was interested in quantifying the flexibility of zeolites using different approaches. As flexibility is often associated with anomalous behavior such as the negative thermal expansion, our respective works overlapped in some parts. Quentin calculated then the same 120 pure-silica zeolites that I was able to converge using DFT. He used for this the General Utility Lattice Program (GULP)[188], a program that can perform a variety of tasks related to three dimensional solids. From this software, he was able to obtain the phonon frequencies, eigenvectors and mode Grüneisen parameters of all the zeolites. He used for his calculations a potential developped by Catlow et al.[189]. In fig. 2.15 is represented the Grüneisen parameters calculated with GULP by Quentin along with my results obtained with DFT. It appeared that the behavior observed is similar with both methods: large negative Grüneisen are observed only at low frequencies. Interatomic potentials are known to be less accurate than DFT methods. Such similar results are then encouraging for the study of much larger databases as interatomic potentials are cheaper than DFT calculations.

One of the main work of Quentin was the determination of the "RUMminess" and its link with the flexibility. The global RUMminess can be thought of a measure of how much RUMs weigh in a structure. It is an indicator of how "RUMmy" a structure is. RUMminess is determined from the RUM character of the vibration modes.

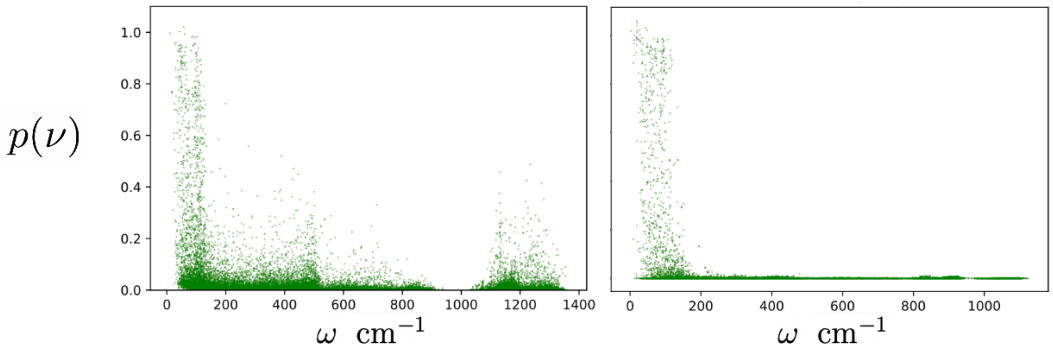


FIGURE 2.16: RUM characters of each mode of each zeolite against their corresponding frequencies. In both panels, RUM characters were calculated by Quentin with CRUSH. Left panel and right panel correspond respectively to frequencies calculated with DFT and GULP.

RUM character can be calculated on each modes and represents how close it is from a true RUM. Quentin calculated the RUM character $\rho(v)$ of each modes for all the different zeolites. He compared the results with the DFT calculations. He obtained these results using the CRUSH program, which is a lattice dynamical program that can calculate the frequencies and eigenvectors of RUMs. In fig. 2.16, is represented the RUM character of all the vibration modes compared to their frequencies of vibration obtained with CRUSH. The same results are also represented for my DFT calculations. One can see that high values of RUM characters are present only in low frequency modes in both cases. This is a logical result as RUMs are low energy vibrations involving rotations of polyhedra. However, the DFT also showed lower values of RUM character at higher frequencies while CRUSH results exhibits almost no character above 200 cm$^{-1}$. Some vibration modes at higher frequencies seem then to keep a certain amount of RUM character. This was not detected by CRUSH which means that some informations are lacking compared to the DFT.

After determining the RUM character of all the vibration modes, Quentin assigned a value of RUMminess to all the structures. We wondered then if the RUMminess score could have a link with the properties of zeolites. Indeed RUMs are one of the source of NTE. In theory, zeolites with a highest RUMminess score should show a larger NTE. fig. 2.17 represents the RUMminess score along with the thermal expansion coefficient and the bulk modulus. We observed in both cases that there was no correlation between RUMminess and the properties of interest. This suggests that RUMs are not the main origin of anomalous behavior such as NTE: other driving forces may be at play in these systems. It can be seen also that there are no clear differences between the RUMminess of experimentally feasible frameworks and the theoretical ones. Therefore, RUMminess does not constitute a condition for the synthesizability of pure silica zeolites.

I only showed for now the thermal expansion in volume, which is negative for all the zeolites studied. However, when looking closely to the linear thermal expansion, some anisotropy appears among the frameworks studied. Although several frameworks studied have a cubic symmetry (12 out of 120), many others have a lower symmetry allowing anisotropy to arise. In particular, I identified structures in which some directions of space have positive linear thermal expansion — while others have a larger negative linear coefficient, contributing to the overall volumetric

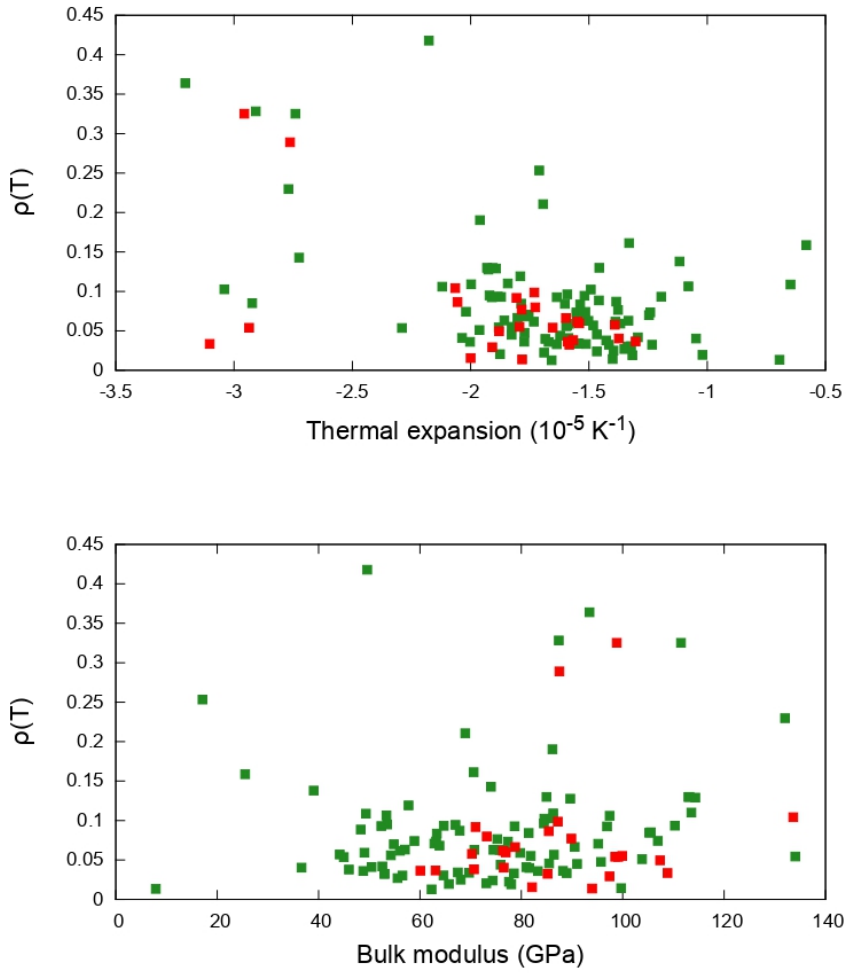


FIGURE 2.17: Thermal expansion and bulk modulus of pure-silica zeolites calculated with DFT compared to the "RUMminess" obtained with CRUSH.

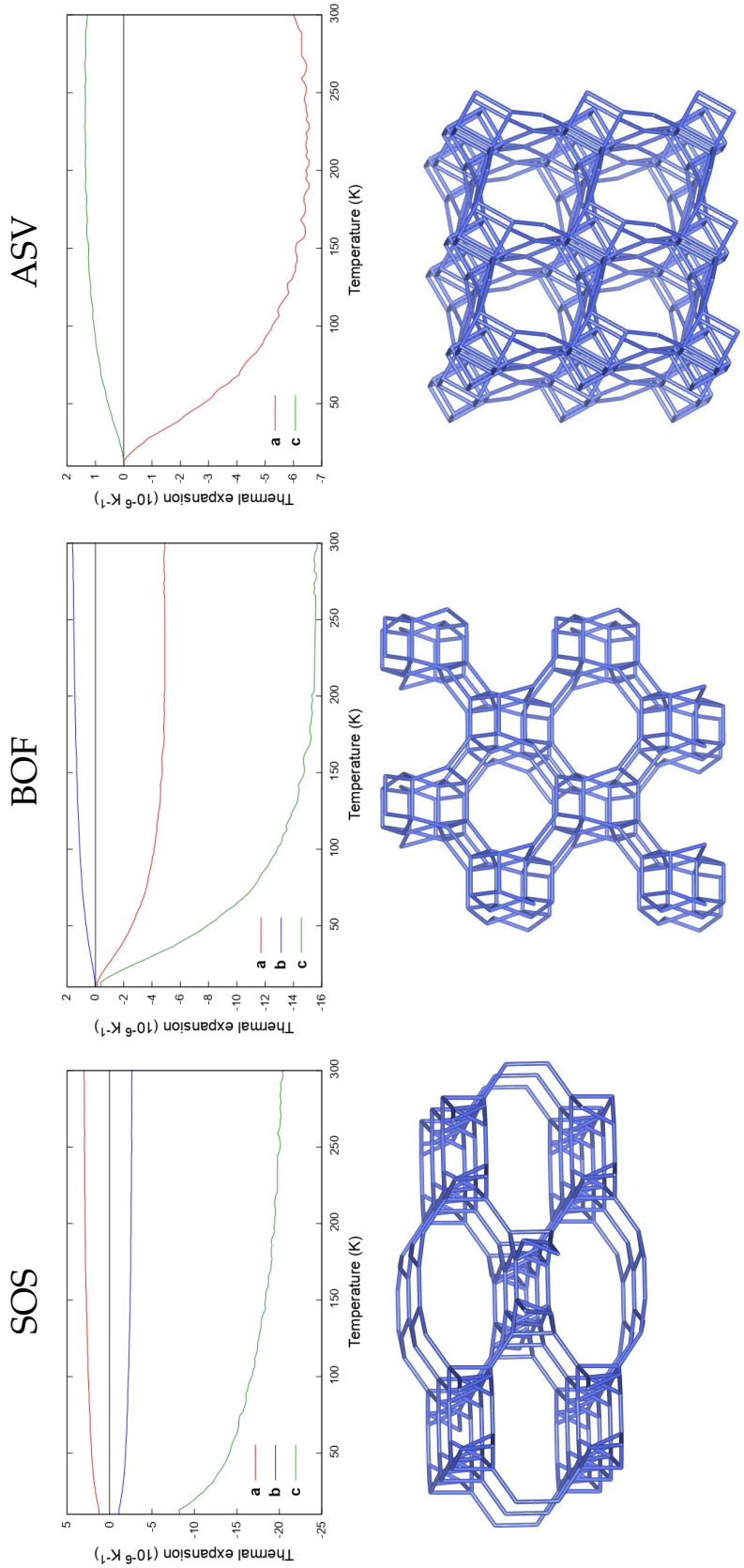


FIGURE 2.18: Evolution of the individual lattice parameters thermal expansion coefficients with respect to the temperature for the SOS, BOF, and ASV framework. The corresponding structures are represented under each graph.

NTE. Three examples of such cases are represented on fig. 2.18 where their structures are shown along with their linear thermal expansion for each cell parameter. The origin of anisotropy can be clearly seen on their framework structure. This gives rise to a very modest positive coefficient along a single crystallographic axis. The number of such highly anisotropic materials is small (9 out of 120), although all non-cubic frameworks feature some level of anisotropy. It is interesting that the SOS framework type was previously demonstrated to have highly anisotropic mechanical properties, showing a link between thermal and mechanical properties through topology[161].

### Thermodynamic quantities

While QHA calculations give access to the Helmholtz free energy, standard harmonic frequency calculations on the equilibrium structure allow to obtain several thermodynamic quantities, such as the vibrational entropy $S_{vib}$ and the constant-volume heat capacity $C_V$. Those are represented in fig. 2.19. Looking at the entropy, it can be seen that it varies very little among the structures considered with values between 37 and 43 J.mol$^{-1}$.K$^{-1}$ per $SiO_2$ unit, with the majority of frameworks around 39 J.mol$^{-1}$.K$^{-1}$. This is in stark contrast with the thermal expansion observed previously which showed a large range of values due to the effect of topology. Here, almost no effect of the topology is found on the vibrational entropy. This observation is to be put in relation with the open question of relative stability and experimental feasibility of zeolites[190, 191]. From the calculated values of vibrational entropy and especially their narrow distribution, it is straightforward to see that this quantity have almost no impact on the Helmholtz free energy ($F = U - TS$). This is illustrated on fig. 2.20 with the relative free energy $\Delta F$. At 300 K, it is completely dominated by the energetic contribution. The small impact of the entropy might change the stability ordering of some of the frameworks, but only the ones that are already energetically very close. This small impact comes from the fact that, while different zeolites have different low frequency modes, they are all widely populated at room temperature.

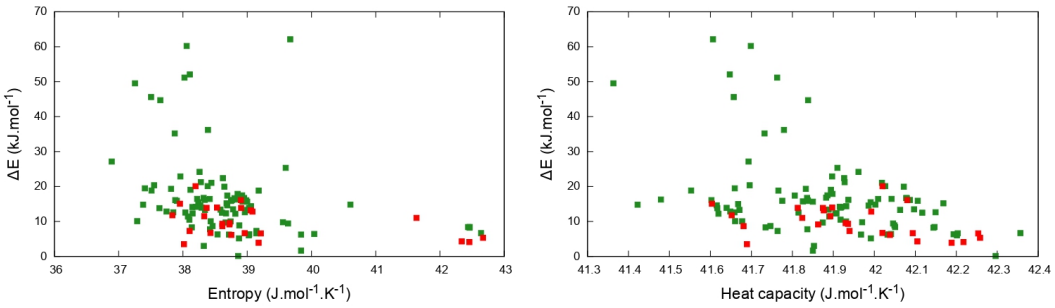


FIGURE 2.19: Relative energy with respect to $\alpha$-quartz ($\Delta E$ per $SiO_2$ unit) plotted against the vibrational entropy (top panel) and the heat capacity (bottom panel), both normalised per $SiO_2$ unit. Entropy and heat capacity values are given at 300 K. Green squares correspond to theoretical structures and red ones correspond to experimentally synthesized structures.

The same observation can be made for the heat capacity represented on fig. 2.19. The distribution of values is even more narrow than for the entropy. The influence of topology is so small on the heat capacity that it's almost imperceptible. Such a behavior is explained by the link between the vibration modes and these physical

properties: We saw previously that the thermal expansion is only dominated by a small number of low frequency modes which are very sensitive to the atomic arrangements. This is however not the case for the entropy and the heat capacity, which are averaged over all the modes of the structures and are therefore much less impacted by a change of topology.

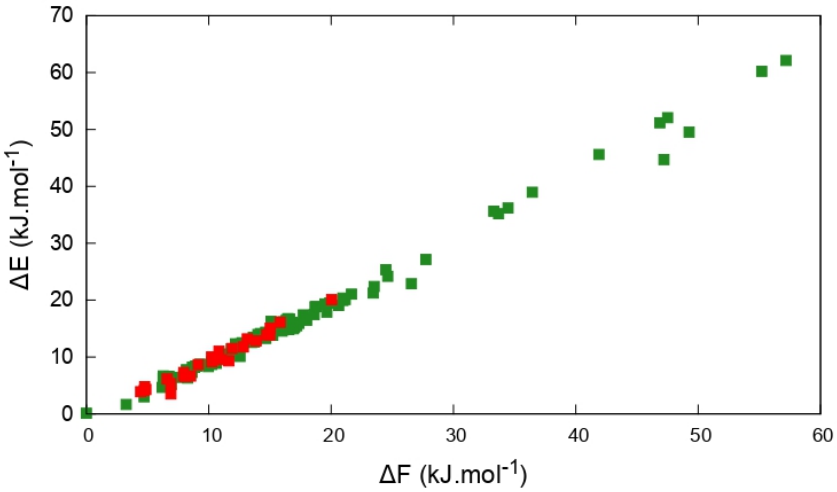


FIGURE 2.20: Relative energy with respect to $\alpha$-quartz ($\Delta E$ per $SiO_2$ unit) plotted against the free energy, also normalised per $SiO_2$ unit. Free energy is given at 300 K. Green squares correspond to theoretical structures and red ones correspond to experimentally synthesized $SiO_2$ structures.

### Volumetric properties

With the use of zeo++ software[192, 193], it was possible to compute characteristic data on the porosity of the structures considered. Several quantities could be obtained: the accessible volume, the accessible surface area, the largest free sphere and the largest included sphere. An illustration of the determination of surface area is shown on fig. 2.21. All of them are represented on fig. 2.22 against the relative energy with respect to the $\alpha$-quartz. These figures highlight two points: first, porosity is obviously omnipresent in zeolites as all frameworks feature some amount of accessible volume. Second, the pore shape of a given topology does not seem to affect its stability, as no correlation can be observed with the largest free sphere nor the largest included sphere. However, a weak correlation seems to appear in the case of the accessible surface area and the accessible volume. Although this could be simply due to the rather small data set used here, the topologies with smaller accessible surface area and volume seem to be more energetically stable.

Comparing now the relative energy of experimentally synthesised structures and theoretical ones, it appears that the characteristics of porosity are of no impact on the synthesizability. Indeed no differences between the experimental and theoretical structures can be observed. Both groups are really close to each other and no particular behavior nor correlation can be retrieved from these representations. Even when looking at the accessible surface area and volume, for which a weak correlation could be observed with the relative energy, no differences are observed. This shows that porosity is a common feature of zeolites and does not constitute a clear criterion for the experimental synthetic feasibility.

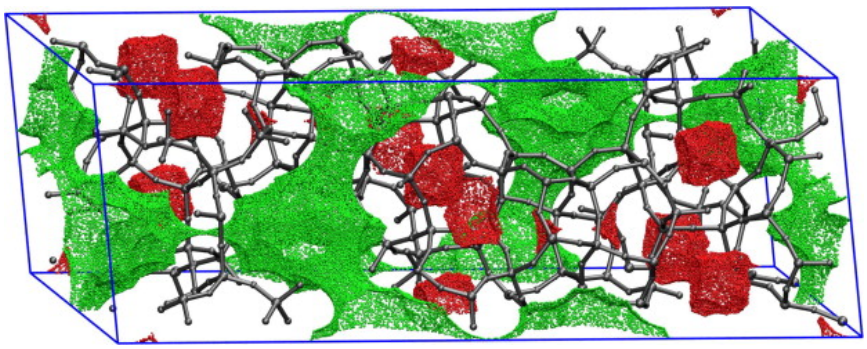

FIGURE 2.21: Sampled points on the surface of the DDR zeolite obtained with zeo++. Green and red points are respectively, accessible and inaccessible to a spherical probe of radius 3.2 Å. Taken from [192].

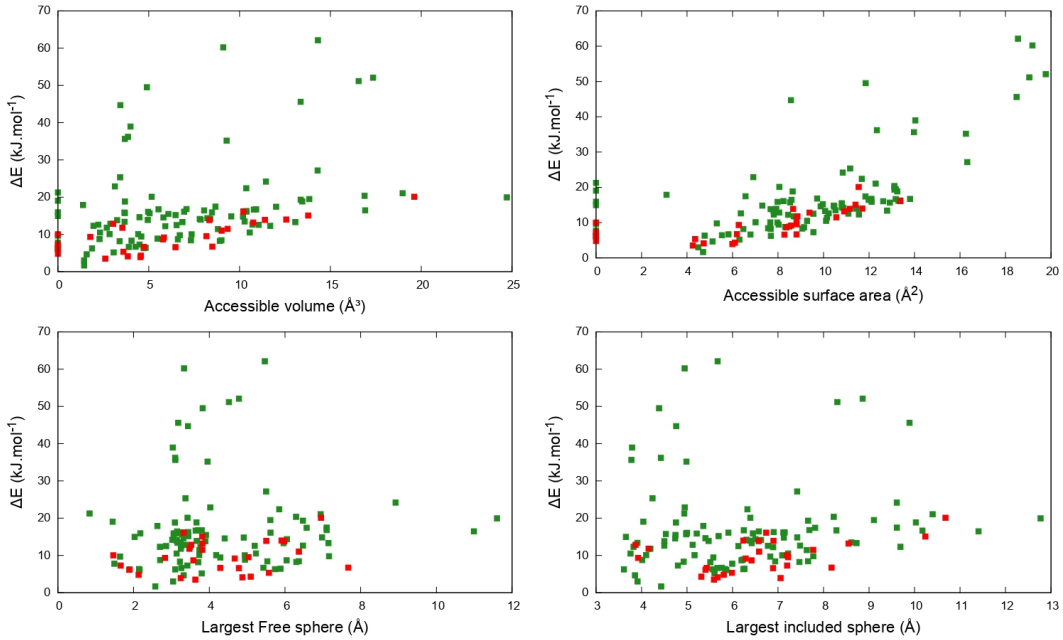


FIGURE 2.22: Relative energy with respect to $\alpha$-quartz ($\Delta E$ per $SiO_2$ unit) plotted against the accessible volume, the surface area and the largest included and free sphere. Accessible volume and surface area are also normalised per $SiO_2$ unit. Green squares correspond to theoretical structures and red ones correspond to synthesized $SiO_2$ structures.

### Mechanical properties

Previous computational studies from the group have shown that the analysis of mechanical properties of inorganic framework materials in general[194], and zeolites in particular[161], could provide valuable insight into the relationship between structure and physical properties. These properties are typically calculated through strain-stress relationships, whether it is at the quantum chemical (DFT) level[161, 195], using classical force fields[196], or through the application of machine learning methods[165, 195]. Whatever the accuracy of the interatomic description of the system, these mechanical properties were however systematically calculated in the linear elastic regime at the zero Kelvin limit[197, 198]. In particular, the temperature and pressure dependence of the elastic coefficients or other derived mechanical properties (elastic moduli, Poisson's ratio, etc.) have not been systematically studied. There is no data available in the literature on the range of values of those pressure and temperature dependence, nor on the impact of framework topology in a given class of materials.

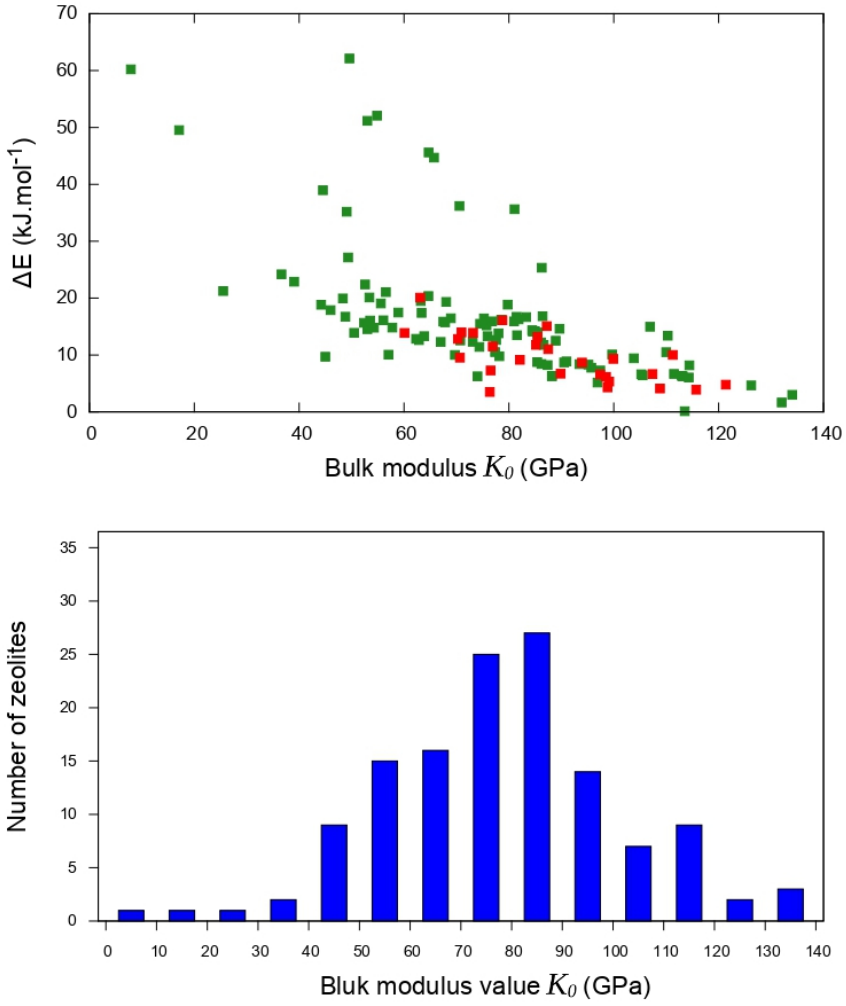


FIGURE 2.23: Top panel: Energy relative to $\alpha$-quartz ($\Delta E$) plotted against the bulk modulus $K_0$ obtained through the Birch–Murnaghan equation of state. Green squares correspond to theoretical structures and red ones correspond to synthesized structures. Bottom panel: distribution of the $K_0$ values for all the frameworks.

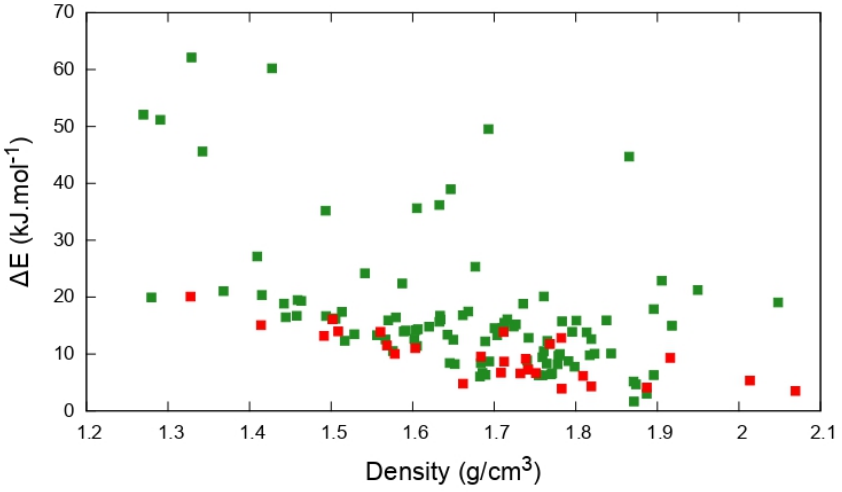


FIGURE 2.24: Relative energy with respect to $\alpha$-quartz ($\Delta E$ per $SiO_2$ unit) plotted against the density. Green squares correspond to theoretical structures and red ones correspond to experimentally synthesized structures

Through all the calculations done in this study, I obtained for the first time the bulk modulus for each zeolite structure considered as well as its pressure and temperature dependence. Values of bulk modulus for all the structures are represented in fig. 2.23 as a histogram and as a graph with respect to the relative energy. A large span of values is observed, from 8 GPa to 134 GPa with an average value of $\sim$78.5 GPa. Thus, like for the thermal expansion, a great impact of the topology on the bulk modulus is observed. Extreme values show a difference of one order of magnitude which is very large considering that the materials have the same composition. As expected, this property is thus strongly linked to the arrangement of the framework (topology). Looking at the graph of $K_0$ compared to $\Delta E$, a few remarks can be made: first we can observe a weak correlation between these two quantities, meaning that softer materials are less energetically stable. This result is intuitive and can be observed with the density as well fig. 2.24 as it is known to be linked with the relative energy. Secondly, it can be seen that experimentally feasible frameworks behave differently than the theoretical ones as they feature only high values of $K_0$ (above 60 GPa). Only theoretical frameworks were found with low values of bulk moduli. A high value of $K_0$ is then a first requirement for the synthesis of pure silica zeolites.

From the calculations performed, I also had access to the pressure and temperature derivatives of the bulk modulus (which we obtained respectively as a Birch-Murnaghan coefficient $K_0'$ and through $\delta = (1/K)(\partial K/\partial T)$). Comparing $K_0$ and its pressure derivative in fig. 2.25, it can be observed that there is again a large effect of topology on $K_0'$ which features both positive and negative values. Most frameworks are found in the $-5$ to $+5$ region but some values are found up until $-25$. It can be seen on the histogram of fig. 2.25 that a lot of zeolites show a negative value of $K_0'$. This means that a lot of pure silica frameworks exhibit an abnormal softening of the elastic modulus with compression — that can be linked to the mechanism already established for pressure-induced amorphization of zeolitic frameworks[199, 200]. Such a pressure-induced behavior has been previously identified in a small number of framework materials (including porous inorganic compounds)[201, 202, 203]. It

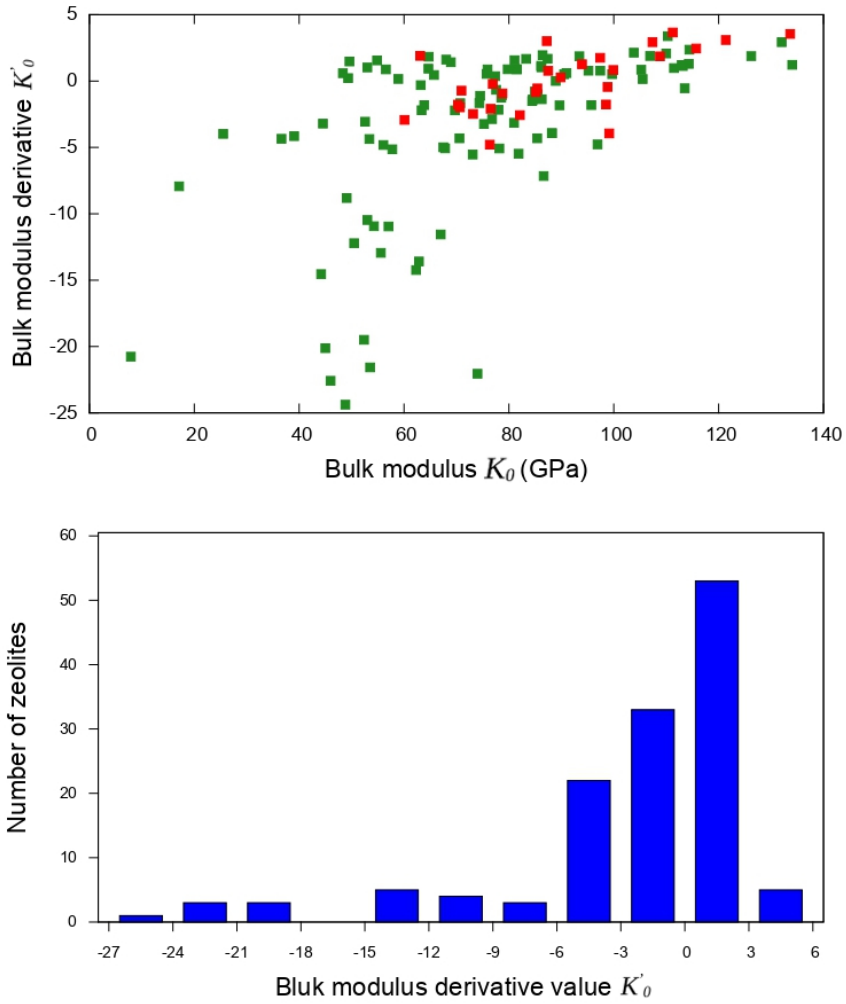


FIGURE 2.25: Top panel: Bulk modulus $K_0$ plotted against its pressure-derivative $K_0'$ for each zeolite. Green squares correspond to theoretical structures and red ones correspond to synthesized structures. Bottom panel: Distribution of the values of $K_0'$ for all the frameworks.

has even been suggested that pressure-induced softening could be generally linked to negative thermal expansion[204, 205], although we show here that it is not direct and systematic link: all $SiO_2$ frameworks studied in this work show NTE, but only 56% exhibit pressure-induced softening (and the vast majority of structures are within the $-5$ to $+3$ range).

In addition to this, it can be noted that among the frameworks showing a value of $K_0'$ below $-5$, no experimental structure can be found. This parameter appears then to be a synthesis criterion meaning that a too negative value would lead to instabilities.

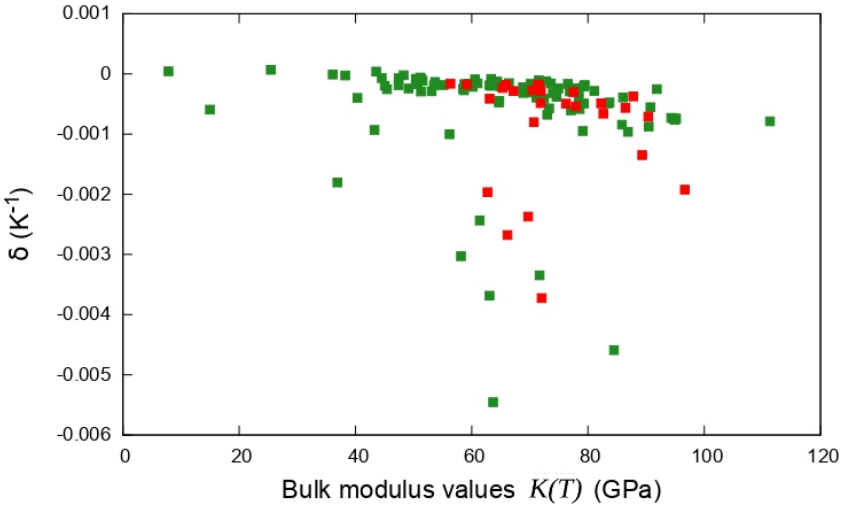


FIGURE 2.26: Temperature-dependent bulk modulus $K(T)$ compared to the temperature coefficient of elastic stiffness $\delta$ for all frameworks at 300 K. Green squares correspond to theoretical structures and red ones correspond to synthesized structures.

Finally, I also had the opportunity to investigate the temperature derivative of the bulk modulus. This quantity $\delta$ is represented in fig. 2.26. It appears clearly that, compared to the previous properties, the effect of topology is really small. Almost all the values are contained in the zero to $-10^{-3}$ $K^{-1}$. A few exceptions with a bit larger negative values can however be observed, among which both theoretical and experimental structures can be found.

## 2.3 Conclusion

The systematic study of a family of material is useful to deepen the understanding of a family's behavior. It is also a great tool for highlighting the specificities of a given property. Such studies are not experimentally feasible as it would require tremendous time and investment. Computational methods are more indicated for this type of work as it is cheaper and most of the time faster than experiments. However, setting up such studies is still a complex challenge as the exploration of a large number of frameworks leads to a high computational cost. This cost is mostly dependent on the chosen level of approximation as well as on the number of structures. Studies comprising large databases of structures usually resort to force field techniques as they keep a relatively low cost while ensuring a modest level of accuracy. However, calculating a limited number of frameworks allow to use methods with a higher accuracy. In the end, systematic studies are not common as they require a large

investment. They are nevertheless needed for the comprehension of systematic behaviors appearing in specific families of materials. In the case of pure-silica zeolites, only a few works of this type were performed. In the group, the systematic study of the mechanical properties of 121 zeolites revealed that synthesizable frameworks all had a relatively low elastic anisotropy as well as a low relative lattice energy. It allowed to guide the synthesis of new pure-silica zeolites by extending the feasibility criterion previously described in the literature. Other efforts were directed towards the construction of databases of theoretical structures of zeolites. The PCOD2 database contains around 600,000 structures and is useful for both synthesis of new frameworks and computational studies. However, the thermal properties of zeolites, known to show negative thermal expansion (NTE) as a common behavior, have not been systematically studied yet. The simulation of a large number of frameworks is necessary to attain a better understanding of both the behavior of the zeolite family and the NTE.

In this chapter, I investigated the thermal behavior of 120 pure silica zeolite frameworks. Static DFT calculations were used to optimise the geometry of each structure and to determine the thermal properties through the quasi-harmonic approximation. Preliminary calculations revealed a good agreement with the scarce experimental data available, comforting my choice of parameters. The thermal expansion obtained is negative for all the frameworks considered. The topology seems to play an important role in defining the amplitude of the thermal expansion as its values span over a large range. No differences could be observed between the theoretical and experimental structures, making it impossible to determine a synthesis condition. Investigating the mode Grüneisen parameter, I observed that the negative global Grüneisen is largely dominated by a small number of strongly negative mode Grüneisen parameter. However no link could be found between the properties calculated and the "RUMminess" determined by Quentin Gueroult with CRUSH. Observation of the thermodynamic quantities revealed that the effect of topology is not impacting the entropy nor the heat capacity. As a result, the value of vibrational entropy is very from one framework to another and has almost no impact on the Helmholtz free energy. Analysis of the volumetric properties led to the conclusion that the pore shape does not affect the stability of the framework. A weak correlation is observed in the case of the accessible volume and surface area, indicating that lower values of both features induces a higher stability. A large range a value was found for the bulk modulus showing an effect of topology even more stronger than in the case of thermal expansion. As no experimental frameworks exhibited a bulk modulus lower than 60 GPa, it could be hypothesized that a high value of bulk modulus is a requirement for the synthesizability of a framework. The pressure derivative of showed that a lot of the frameworks features a negative value, indicating an abnormal softening of the elastic modulus with compression. Also it was observed that experimental frameworks are found within the -5 to +5 range of value while theoretical frameworks could be find up until -25.

During this work, I created a database of DFT-calculated properties for 120 pure-silica zeolites. This database was made available online in the team's GitHub repository. While this database is quite small compared to others, I believe that the DFT-level of accuracy of the resulting properties can be useful in several domains. Such data can used as a basis for machine learning studies for example. Indeed, the efficiency of machine learning is strongly linked to the quality of the data. Therefore, using a database of DFT-calculated properties as training set would ensure a high quality of input data. They are however other challenges such as the size of the

database being small compared to the usual size of a training set or the representation — and differentiation — of zeolite structures that need to be addressed. The valorization of my data through machine learning was already envisioned before this study. Therefore I put in application this idea and tried to address the problematics mentioned previously, so as to determine if my database of DFT-calculated zeolites is suitable for such methods. This will be the main topic of the following chapter. Other possible development include the creation of a force field. Indeed, force fields development require a large amount of data calculated through DFT. Another interesting point would be to study the effect of composition. This could be achieved with the same methodology used in this chapter, considering the same topologies calculated here but with a different composition. This could also be done on a small number of frameworks on which the Si/Al ratio could be modified progressively in order to witness the effects on the properties, especially the thermal expansion.

# Chapter 3

# Machine learning for the prediciton of zeolites

Given the large number of DFT-calculated structures and properties of zeolites I had created during my PhD project, I decided to study the feasibility of a machine learning using this data set. Machine learning techniques are drawing more and more attention nowadays due to the increased possibility of obtaining and storing large amount of data. Depending on the method chosen, machine learning can be initiated without prior learning on a training set (unsupervised machine learning) or can need already known data in order to predict new ones (supervised machine learning). The latter requires then a set of clean and already prepared data. The data I obtained on pure silica zeolites in the previous chapter could form a suitable data set for machine learning purposes. Indeed, it is the first time in the literature that thermal properties were obtained systematically for zeolites using DFT. I believe then that the data acquired could be useful for the prediction of thermal properties of other zeolites or zeolite-like materials. However, my data set is smaller than those typically used in the field of machine learning. The aim of the following chapter is to assess the applicability of machine learning to the data previously calculated. The first part will define the concept of machine learning and explain its basic principles. The second part will present the results I have obtained.

The work described in this chapter has been published in the Journal of Physical Chemistry C[206].

## 3.1 Concept of machine learning

The concept of machine learning arose around 1960. Around that time, a psychologist named Frank Rosenblatt, inspired by the human nervous system, invented a machine for recognizing the letters of the alphabet[207, 208, 209]. He named his machine the "Perceptron". This became later the prototype of modern artificial neural networks. Several other groups focused on designing and testing learning recognition systems[210, 211, 212]. However, in 1969 Minsky and Papert published a book where they presented the limitations of the Perceptron[213]. They especially argued that some logical functions could not be handled by this system. This book slowed down the funding of AI research until the 1980s, even though more learning algorithms were still developed during this period.

The major turning point in the history of machine learning was during the first decade of the $21^{st}$ century. Several factors in that period led to the blooming of machine learning that we are experiencing nowadays. First, the accumulation of

data collected during routine, human activity and the birth of “Big data”. Due to the amount of data needed to be analyzed, new methods were invented. Compared to the past inventions originating from curiosity, it was this time motivated by necessity. The second important factor was the performance and accessibility of computer resources. At that time, the major companies of the field made breakthroughs in their respective domains. Google revealed in 2004 their MapReduce technology, improving the field of parallel computing. Nvidia, while in competition with AMD in the gaming market, also improved the performances of graphic cards and became the leader in the field of scientific computing, including machine learning applications. At the same time, RAM became more accessible as its price significantly decreased during this period. This led to the birth of several new types of databases, including NoSQL for example. Still nowadays, computers are becoming more and more efficient, allowing for more complex applications of machine learning with very large databases.

Overall, the idea of machine learning is recent (around 60 years) and its growth and intensive use is even more recent (last two decades). The development of machine learning was difficult, but was exponentially accelerated with the improvements on computer power. As of now, machine learning is experiencing its golden age, and is used in numerous studies every year.

### 3.1.1 General principle

In its most general definition, the term machine learning designates all the techniques or algorithms that generate other algorithms, so as to progressively improve the performance on a task. It is a generalized method, meaning that it does not require a specific type of data. Machine learning algorithms are not directly programmed to realize a specific task but rather to learn how to do it. This "learning" is evaluated using a user-defined function measuring the performance of the algorithm given a set of parameters.

Machine learning uses specific terms to designate some common elements of the field. Many of them will appear in this chapter. Most of them will be explained later, but a few are already introduced here:

**Model:** predicting object resulting from the training of an algorithm.

**Training:** process of feeding the algorithm with input data so it can learn the rules and/or the patterns.

**hyperparameter:** parameter controlling the training of the model.

**Feature (or descriptor):** input value that describes and differentiates the entries of the data set.

**Target:** designates the property that one aims to predict. Targets are not necessarily present in a data set.

The main idea behind machine learning is to identify patterns in data. In a sense, machine learning is really close to statistical modeling. However, there are some differences between them, the main one being that in statistical modeling, all of the data are used to fit the model. In machine learning, only a part of the data is used to "train" the model. The actual performance of the model is only known after testing it on a new set of data never seen during the training. This set is called the test set.

Statistical modeling is also more oriented towards the comprehension of the relationship between the data and the outcome variable, although it also allows to make predictions with limited accuracy. The main advantage of statistical models is thus the interpretability of the results. Machine learning techniques, on the other side, are more focused on making robust predictions and do not always care about the interpretability of the model (although it strongly depends on the chosen method). For example, neural networks usually show very accurate predictions but as a downside the results cannot be explained.

### 3.1.2 Data and descriptors

Machine learning uses data sets as input for training a model. Data sets consists of a group of data points, where each point (or "record") represents an entity that is to be studied. A data point is composed of two parts: a target value, which is the value we want to study, and a certain number of features or descriptors. Target values can be clear concepts (for example house price, a family's income, etc) or abstract concepts (attractivity, quality, etc). They are associated with descriptors also called features that are an ensemble of values describing the entity to study. They can be of three different types. Categorical, meaning they take values that fall into categories like "apartment" or "house". Ordinal, which take values in order (for example, the history of renovation of a house). Numerical, which take actual values like the surface or the number of rooms. There are no rules as to which type one should be using. It is completely dependent on the application. Data sets can contain features of only one type or can combine the three different types. The crucial point is to ensure that the features allow for the description of the entities and allow the model to differentiate them. It is also necessary that the features share some kind of link with the target property. As long as this condition is satisfied, the data set can be applied for machine learning purposes. As an example, if we take the example of housing and aim to predict its price, we can imagine a simple combination of features: the type of housing (categorical), the location (categorical), the surface (numerical), the date of construction (numerical) and the number of rooms (numerical). Depending mostly on the size of the data set, there could be many more features that could be included.

Aside from the choice of features, there are other aspects of the data set that need to be addressed. Among them, the quality of the data is important. Indeed, the model learns from the data and its performance is thus conditioned by the quality of features and target values. Ill-defined points of the data sets are often dealt with prior to realizing the machine learning study. An example of such points are the missing values in the data set, which need to be removed as they can introduce errors. Another point to pay attention to is the so-called "bias". Bias in machine learning is defined as the phenomenon of observing results that are systematically prejudiced due to faulty assumptions. While bias is not desired, it is omnipresent in machine learning. Indeed, even the choice of a cost function constitutes a bias in itself, although it is considered as a productive bias. A model without some degree of bias would do not better on a task than if the result was chosen at random. However introducing too much bias in a model can cause prejudiced results and reduce its performance on more general tasks. Indeed, according to the "No Free Lunch theorem"[215], if an algorithm is particularly adapted to solving a certain type of problem, then it will perform worse on the remaining average of problems compared to the general-purpose

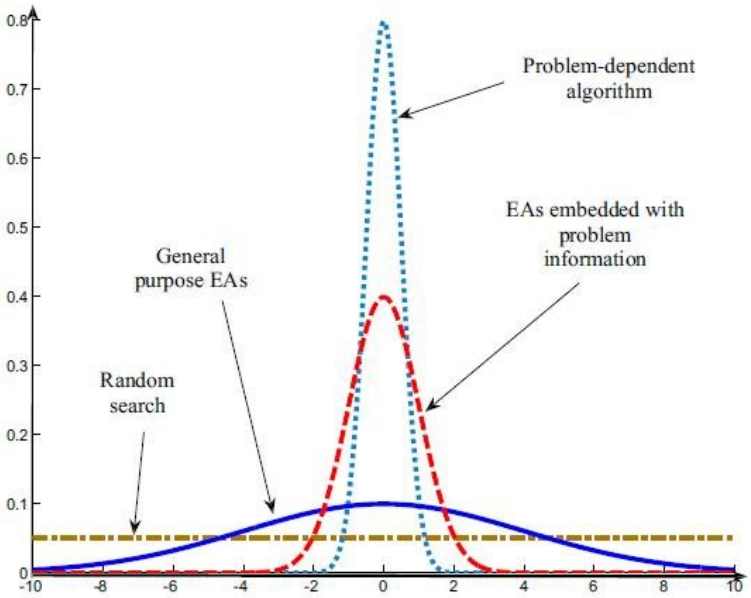


FIGURE 3.1: Illustration of the No Free Lunch theorem. Taken from [214].

algorithm. This principle is illustrated on fig. 3.1 Therefore, (un)intentionally introducing more bias induces a decrease of generalization. It is possible to estimate the degree of bias with tools such as the feature importances for example, which will be used later in this chapter.

### 3.1.3 Supervised and unsupervised machine learning

In machine learning, there are two large families of techniques: supervised and unsupervised learning. The main difference between them is that one requires data as well as target values to train on before being able to output results (supervised), while the other directly takes raw data without target values and analyze them without prior training (unsupervised). Each family has its applications: supervised learning is more oriented towards the prediction of properties while unsupervised learning is more efficient for identifying patterns in data. I will describe shortly one example of unsupervised learning. However, I will not linger on this family of techniques as it was not used in this study and will focus more on supervised learning.

As said previously, unsupervised learning does not need any target data to train on. The principle is to analyze the patterns of a given set of data. There are several methods such as clustering and dimensionality reduction. Clustering, for example, is used to identify the subgroups in a given data set without prior knowledge on the properties they might have. In practice, it determines the similarity of the points in the data set. With this similarity metric, it can then choose which points are similar enough to be a part of a subgroup and which points are too different and need to be in different subgroups. The human does not intervene in this process, making it a pure algorithmic decision. Overall, clustering is a technique used to understand the structure of a data set by grouping similar observations together. This concept of similarity can be defined through various algorithms such as density-based or k-means algorithms. Clustering is used in many fields: it can be used to detect

fraud in several domain by highlighting outliers, or can be used for recommendation (for example, proposing a music similar to what a user was listening to). It can also have various applications in organic and inorganic chemistry. It was used for example to characterize different diamond nanoparticles and their most relevant properties[216].

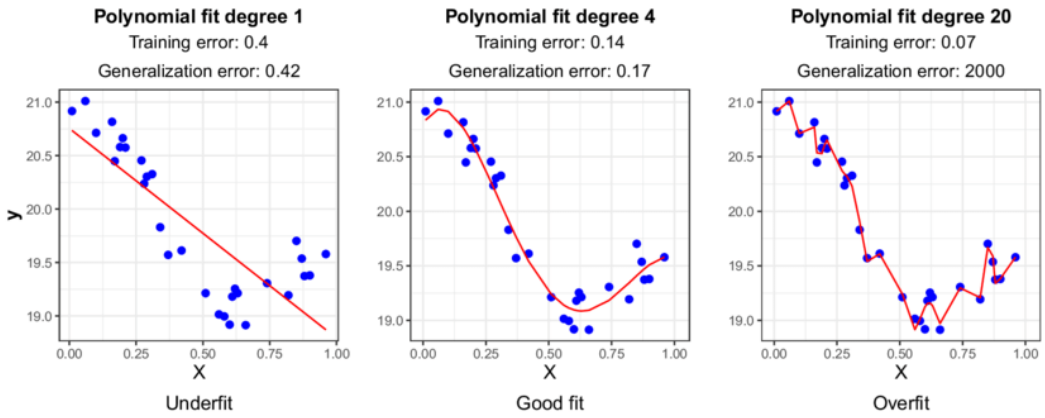


FIGURE 3.2: Examples of underfitting and overfitting for a given set of data. Data points are represented in blue and red lines corresponds to the models. Taken from [217].

In contrast with unsupervised learning, supervised learning requires target data to train on. The principle is to give the algorithm a set of "training data" containing both input values (also called features or descriptors) as well as output values (also called target values). The algorithm will learn the general rules linking input and output values. The performance of the resulting model is tested with a test set, being generally a small portion of the training data left aside (i.e. not used for the training). This process is called training the model. Once this training is done, the newly created model should be able to predict the output values of a new set of unseen input features. This is the general principle of supervised learning. In this family of techniques, there are two important aspects to take into account: model fitting and performance measure.

When training the model, ensuring a good fit is of primary importance. The model needs to be flexible enough to capture the complexity of the data, yet robust to noise in the input features. Under-fitting and over-fitting are two behaviors that need to be investigated and addressed in order to make sure that the training process is successful. Both cases are represented on fig. 3.2. Under-fitting occurs when a model is not able to predict the target values of the training set. In this case, it means that the algorithm did not learn the general rules linking features and target values. In the general case, a model should be able to predict values close to the observed ones on the training set. Under-fitting originates usually from badly-tuned hyperparameters. In opposition with this behavior, over-fitting corresponds to a model predicting well the values of the training set but performing poorly on the test set. This behavior means that the algorithm has been trained too much on the training set and the resulting model is not able to apply the rules learned to other unseen data. This is a problem of generalization: a model should be able to generalize beyond the training set. It can be due to the training set being too small compared to the test set. It can also come from the hyperparameters allowing the algorithm to learn more than what it should.

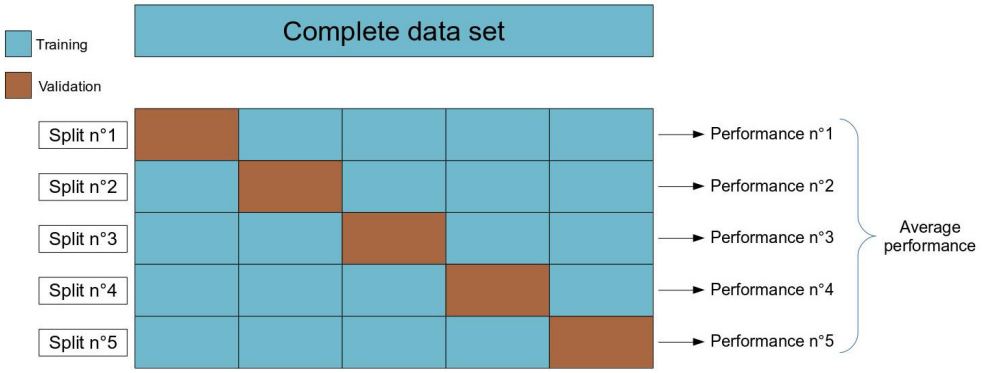


FIGURE 3.3: Schematic of the principle of cross-validation considering 5 fold.

The second important aspect of machine learning is the performance measure and the model selection. The principle is to measure how good or bad a model is, allowing then to choose which model is the best. There are several options for the measure of performance depending on the type of models. In my case, I used the root mean squared error (RMSE) as it is a common tool for measuring the performance of regression models. Other common measures include the mean absolute error or the mean squared error. Classification models, however, need different measures usually. I can be a simple accuracy score (percentage of correct predictions) or other methods derived from the confusion matrix. All these tools are used to assess the performance of a model and guide its selection. The usual way of selecting a model is as follow: the entire data set is separated in one large and two small data sets. The large one is the training set and the two others are the validation and test sets. The training set is used to create several models and the validation set, along with the performance measure, is used to select the best model. Finally, the test set is used to determine the generalization error. This is the standard way of creating a machine learning model for a prediction purpose. However, when the data set is not large enough, separating the data in smaller sets can be complicated and reduce the quality of the training. In such case, the use of cross-validation can be helpful. This method is depicted on fig. 3.3 with the example of a 5-fold cross-validation. The principle is to randomly split the initial set in k number of fold. For each fold, k-1 folds will be used to train the model and the remaining part will be used as a validation set. The performance measure of this method is the average of each simulation measure.

Similar to unsupervised machine learning, supervised learning comprises several techniques. Such methods include random forests, k-nearest neighbours or kernel methods. In my case, I used the gradient boosting technique. Gradient boosting was built around the idea of decision trees. Decision trees is a simple decision-making algorithm that finds a result by answering simple questions. However, this is prone to overfitting and is usually not used by itself. Gradient boosting regression is an alternative that uses decision trees but build them as weak predictors. It creates trees with a low depth (= a small number of decisions) and adjust the performance by adding more and more trees during the training process. The hyperparamters control what the gradient boosting algorithm is allowed to build, whether it be the maximum depth of a tree or the number of leaves. Fine-tuning these parameters is very important in order to avoid overfitting.

### 3.1.4 Machine learning in chemistry

The demonstration of the Schrödinger equation gifted the scientific community with a new powerful tool for the study of structure/properties relations. It was followed soon after with the development of quantum mechanics and led later to the birth of computational chemistry. Given the the exponential improvement of computer power, the use of calculations at different levels (force fields, Hartree-Fock, DFT, etc) has been increasing in the field of chemistry. Screening of large numbers of structures is now becoming more and more common, but is limited by the chosen level of approximation. The use of AI-based methods, such as machine learning, can improve the protocols already existing and can allow to work with extensively large numbers of structures within a reasonable timescale.

Over the years, in response to the needs of the scientific community, databases of materials grew exponentially. This also contributed to the blooming of machine learning in chemistry. The Inorganic Crystal Structure Database (ICSD)[218] is one of the oldest and most known database for crystal structures as it started collecting data in 1913. Around 6,000 structures are added every year to this database, containing already more than 200,000 structures in total. Only structures meeting the quality requirements are included in the database, making it a suitable base for machine learning applications because it is curated. More specialised databases also appeared such as CoREMOF[219], which is a database of metal organic frameworks last updated in 2019 and containing around 14,000 materials. New kind of databases, containing different entries than materials, were also created. It is the case of the Reticular Chemistry Structure Resource (RCSR)[220], which is a database of topologies. It stores the 3-letter name codes along with their respective nets. It is a useful database for the design of new structures. Molecular chemistry has also seen a rapid development of databases. ChemSpider[221], for example, is a database of molecules containing now more than 100 million structures. However, while some of these examples can be used in machine learning with only a minor treatment of the data itselves (as is the case with the ICSD which follows a strict quality control before adding a new entry, ensuring a clean database), this is not the case for all. Many of them contains some errors, which forces to clean the data before using it. Making large curated databases that can be used without prior treatment of the data is still an open challenge.

As described previously, the entries of a data set contain descriptors used to represent the input structures (as well as target values in the case of supervised learning). Therefore, one key aspect of machine learning in chemistry is the representation of materials. This is an important factor as the quality of the description directly conditions the performance of the model. Descriptors should allow to differentiate properly the different materials while giving an accurate and suitable representation of the structures. This is not straightforward and has been the object of many researches. The main issue in crystalline materials is that the conventional representation (unit cell) uses translation vectors and fractional coordinates. This representation cannot be used directly in machine learning as it can be represented in an infinite number of ways, resulting in different predictions. Chemical systems need descriptors that do not variate with a different referential. This field is still very active, and some solutions are proposed to palliate this issue. For example, the Coulomb matrix can be a good descriptor as it contains informations on atomic nuclear repulsion and the potential energy of free atoms[222]. Other valid representations include the ones based on radial distribution functions[223] or Voronoi tessellations[224].

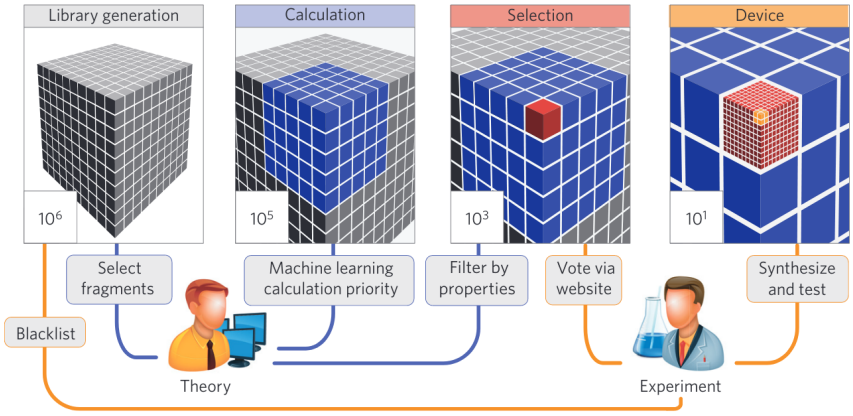


FIGURE 3.4: Diagram of the methodology employed in ref [1]. The cubes represent the size of the chemical space considered at any given stage of the process. Taken from [1].

The use of machine learning in chemistry can take several forms. One possibility is the prediction of chemical synthesis route. The discovery of new chemical synthesis routes is a very complex challenge. For a target compound, several routes could be possible and, within the same route, every step has a lot of possible parameters. Guiding the chemical synthesis with AI-based algorithms represents then a great application of machine learning in chemistry. It would allow to save a lot of time and resources. Such techniques are also useful for the discovery and design of new materials. A recent study showed that the combination of simulation techniques and machine learning could help the discovery of new light-emitting systems[1]. Their methodology is depicted in fig. 3.4 and is a quite typical example. This large screening of over 1.6 million molecules was only possible thanks to the relatively low cost of machine learning compared to DFT simulations or experiments. Simulation techniques usually take a long time (although it depends on the level of approximation) and it is one of their major limitations along with the computer resources. Machine learning can improve this computational workflow as it can learn from simulated data in order to output results with, theoretically, the same accuracy. Other ways of improving simulations have been investigated: machine learning was used to try to produce a universal density functional[225, 226], to optimise exchange and correlation functionals[227, 228] or to create adaptive basis sets[229].

Machine learning is also of use for predicting properties of materials. By learning on a specific type or family of materials, models can predict a handful of properties on similar materials. This is the main idea of this chapter as I tried to determine if my data set of calculated thermal properties is suitable for systematic prediction using machine learning. This was not the first time this type of application was studied in the group. Indeed, in 2017, Evans and Coudert used a gradient boosting regressor model to predict the mechanical properties of zeolites[165]. The model was trained on 121 structures and used to determine the bulk modulus $K$ and the shear modulus $G$ of around 600,000 zeolites. The results they obtained are shown in fig. 3.5. They proved that, for a given composition, geometrical descriptors of a structure such as local, structural and porosity related features can lead to accurate predictions of the mechanical properties. Later, Gaillac et al. expanded this study by looking at the anisotropy of mechanical properties[195]. It resulted in a model capable of predicting the auxeticity and Poisson's ratio for more than 1,000 zeolites.

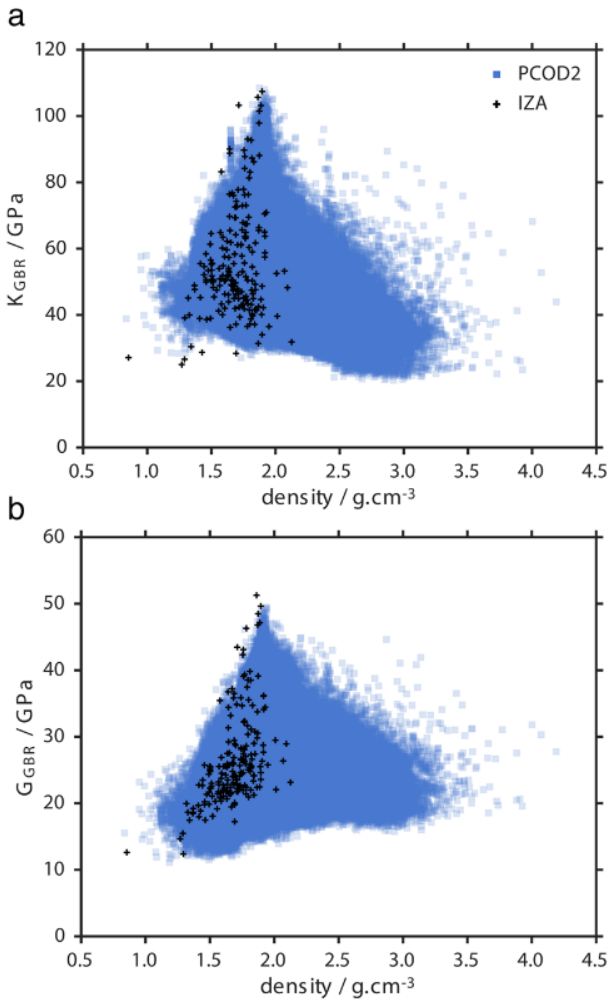


FIGURE 3.5: Bulk modulus $K$ (a) and shear modulus $G$ (b) of the PCOD2 database (blue) and recognized zeolite frameworks from the IZA database (black) as predicted by the GBR model of Evans and Coudert. Taken from [165].

In both cases, machine learning proved to be efficient for the prediction of properties and allows to speed up the discovery of new materials or new properties in known materials. In this chapter, I studied the feasibility of such a methodology for thermal properties. Considering a comparable level of accuracy, these properties are usually more complex to obtain compared to the mechanical properties as they require more calculations at the DFT level. The possibility of using machine learning would then be of great help for large scale screening of thermal properties.

## 3.2 Machine learning study

### 3.2.1 Model parameters

For this work, I used the structures and properties of pure silica zeolites calculated in the previous chapter. 120 topologies were then considered for the data set. Please refer to chapter 2 for a complete description on how these data were obtained. I also used another database containing hypothetical zeolite structures, the PCOD2 database, which will be detailed in the dedicated part of this chapter.

#### Descriptors

As described previously, it is necessary for supervised machine learning to define so-called "descriptors", which are used to represent the relevant features of the data. After the choice of a training data set, the next step concerns the choice of descriptors. As the subject of this study are materials, it is important to choose features that reflect their structure and that will be able to differentiate between the different materials. Descriptors on the composition would be meaningless for example, as all the compositions of the frameworks are identical in my case ($SiO_2$). However, structural features such as cell parameters or bond length and angles would constitute good descriptors. In addition, the descriptors chosen must also have some kind of link with the target values (here the thermal expansion or bulk modulus for example) and must be independent or arbitrary choices such as axes for examples. Considering these conditions, I identified four different types of descriptors which I used for machine learning. For each of them, I chose a total number of 12 descriptors. All of the descriptors chosen for each type are listed in table 3.1. The following of this part will describe each type of descriptor and explain how they were obtained.

*Ad hoc geometric descriptors:* As zeolites are composed of corner-sharing [$SiO_4$] tetrahedra, it is intuitive to describe the structure with metrics describing those tetrahedra. This means that simple measures such as Si–O distances or Si–O–Si angles could be used. The advantage is that these descriptors are readily available. Indeed the computational cost of obtaining them is almost neglectable. It was further demonstrated in a large number of works, including some of the earlier studies on structure–property relationships in zeolitic frameworks,[230, 231] that Si–O bond length and Si–O–Si angles are of great importance for understanding the physical and chemical properties of this family of materials.

In a prior work of the group, it was shown that such *ad hoc* geometric descriptors — designed from the chemical intuition and knowledge about the systems at hand — can be used in supervised machine learning for the prediction of mechanical properties[165, 195]. Therefore, I chose to use this type of descriptor for the first type. I retrieved all the angles and bond distances from the DFT-optimized structures using the pymatgen python package. I then used different statistical features

| Geometric descriptor | Unit |
|---|---|
| Si–O mean | Å |
| Si–O harmonic mean | Å |
| Si–O geometric mean | Å |
| Si–O minimum value | Å |
| Si–O maximum value | Å |
| Si–O variance | Å |
| Si–O–Si mean | ° |
| Si–O–Si harmonic mean | ° |
| Si–O–Si geometric mean | ° |
| Si–O–Si minimum value | ° |
| Si–O–Si maximum value | ° |
| Si–O–Si variance | ° |

| Volumetric descriptor | Unit |
|---|---|
| Accessible volume (AV) | $Å^3$ |
| Non accessible volume (NAV) | $Å^3$ |
| AV + NAV | $Å^3$ |
| Accessible surface area (ASA) | $Å^2$ |
| Non accessible surface area (NASA) | $Å^2$ |
| ASA + NASA | $Å^2$ |
| Density | $g.cm^{-3}$ |
| Volume | $Å^3$ |
| Largest free sphere | Å |
| Largest included sphere | Å |
| Number of channels contributing to ASA | - |
| Number of channels contributing to AV | - |

| SOAP descriptors | Unit |
|---|---|
| Principal components of the SOAP analysis from the first to the 12th (PCAn, n=1, 2, ... ,12) | - |

| Topological descriptors | Unit |
|---|---|
| Number of neighbours in the successive spheres of coordination from the 2nd to the 13th (COORDn, n=2, 3, ..., 13) | - |

TABLE 3.1: Complete list of features used for each type of descriptor. All the values of volume and surface area are normalized per $SiO_2$ unit.

as descriptors: different means, variance, extremal values, etc. For the complete list of features please refer to table 3.1. The python script used for retrieving bond distances and angles is available at `https://github.com/fxcoudert/citable-data`.

*Topological descriptors:* From the previous chapter, one conclusion that could be made is that the framework topology plays an important role on thermal and mechanical properties of zeolites. Indeed, large intervals of values could be observed for structures with identical compositions. Given this result, it is logical to think that descriptors using the topology could be a good alternative of describing the frameworks. Thus I decided to use these descriptors as the second type for my study.

There are several possibilities to represent the topology of a structure, but I chose the coordination sequence which is the simplest one and very intuitive. The descriptors were then the number of neighbours in the successive coordination spheres of the Si atom. As all Si are not necessarily equivalent, I retrieved these data for all Si atoms and averaged the results. Considering the four-connected nature of zeolites, the first coordination sphere contains always 4 neighbours and is thus pointless as a descriptor, as it does not differentiate the different frameworks. Therefore, I used the coordination spheres from the second to the $13^{th}$. Information on topology such as the number of neighbours were obtained from the optimized zeolite structures using the CrystalNets julia package[232]. This package designed for the identification and

manipulation of crystal nets representation and topology has been developed in the group by Lionel Zoubritzky and made available on Github at https://github.com/coudertlab/CrystalNets.jl.

*Volumetric descriptors:* Zeolites are nanoporous materials whose main applications rely on their porosity. The porosity (volume and shape) is different from one topology to another which means that metrics on the volume could be another possibility for differentiating the structures. In addition, the target properties present some links with the porosity. Indeed, as described earlier, the quasi-harmonic approximation we used to determine the thermal properties of zeolites is considering the harmonic expression of the Helmholtz free energy, to which we added the vibrational part of the energy (which depends on the volume). Therefore, metrics on volume check also all the requirements of descriptors. This is why I chose to include them in my study.

I retrieved the quantities on volume and porosity from the DFT-minimized structures using Zeo++ software package[192, 193]. The computation of these properties is not demanding in terms of resources. However, it can be complicated in terms of time. For 120 structures, the time needed is fairly short and anecdotic (around 5 minutes). But for a large database such as the Deem database containing around 600,000 structures, the time needed is much longer (roughly one month). This is why I did not used the volumetric descriptors for studying the Deem database. The surface area, accessible volume and volume being dependent of the choice of unit cell, I normalized them per number of $SiO_2$ units. For consistency, I used as descriptor the calculated (DFT-optimized) crystallographic density, instead of the topological density reported by the IZA. This choice is not crucial to the conclusions of this work, however, as the reported and calculated density values are almost identical.

*Smooth Overlap of Atomic Positions (SOAP):* The SOAP is an encoding method introduced by Bartók and co-workers,[233, 234] which is a descriptor of local geometry, describing the environment around a given point (usually an atomic position). By projecting the local geometry onto orthonormal basis functions based on spherical harmonics, it is invariant by rotation and permutation of atoms. It has been used in particular to determine the similarity of two neighbourhood environments, and to identify features that differentiate molecular structures from one another,[235, 236] including in the specific case of zeolites.[237] It has also been used to encode atomic environments for machine learning inter-atomic potentials, due to its powerful and rich material representation.[238] Finally, it is also used as descriptor for regression tasks, namely the prediction of physical or chemical properties.[239, 240]

One important parameter of this method is the cut-off, representing the distance until which all the environment is included in the description. Initial tests showed that a suitable value for zeolitic systems is around 6 Å, as this distance includes the nearest and next-nearest neighbours, including then characteristics of both Si–O distances and Si–O–Si angles. This was verified by computing the SOAP descriptors for different values of cut-off and running machine learning predictions based on these different sets. The results are shown in fig. 3.6. I observed that the root mean squared error (RMSE) is decreasing until 3 Å where the accuracy seems to no longer be affected by the increase of cut-off. Considering that the time needed to compute SOAP features is still acceptable at 6 Å(a few minutes), I confirmed that this was a good choice for the cut-off. Finally, in order to reduce the large dimensionality of the SOAP descriptors and bring them to a comparable set to other descriptors, I used the Principal Component Analysis (PCA) technique, which determines through an

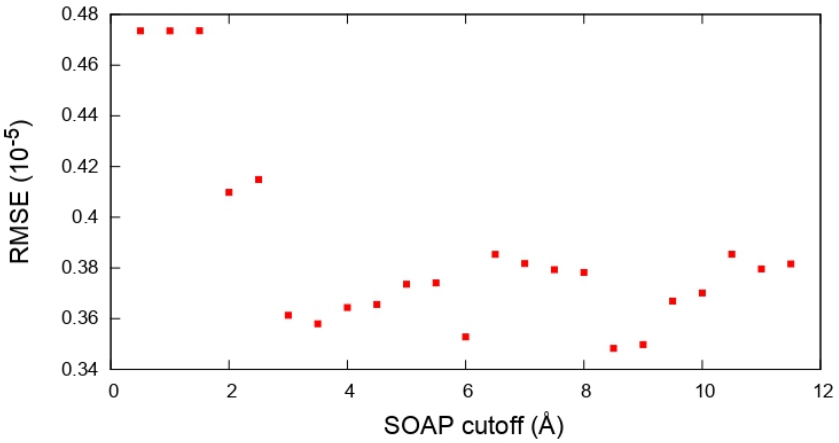


FIGURE 3.6: Cross-validated RMSE obtained with the SOAP descriptors using different cut-off values.

algorithm the most important components and projects the data on them — reducing the dimensionality while keeping as much variation as possible. I chose here the first 12 components from the PCA analysis, making then possible the direct comparison with other types of descriptor.

### Algorithm

Now that the data sets and descriptors are defined, the last step is to define the algorithm and the hyperparameters. I chose to use a gradient boosting regression (GBR)[241, 242] for the algorithm, as it has been proven to be quite robust and efficient for small datasets, like in my case[243]. This algorithm is a stage-wise additive model which trains decision trees that are built in a greedy fashion to minimize the loss function, which I chose to be a least squares function in my case. This methodology has been already applied in the group for the prediction of physical properties in dense and porous frameworks, including mechanical stiffness,[165] and anisotropic elastic properties such as negative Poisson's ratio.[195].

| Parameter | Value |
|---|---|
| Number of boosting stages | 250 [a] / 500 [b] |
| Learning rate | 0.01 |
| Minimum samples split | 2 |
| Maximum depth | 2 |
| Minimum samples leaf | 2 |
| Subsample | 0.4 |
| max features | square root of total features |
| loss function | least squares |

[a] Used for the geometric descriptors only to avoid over-fitting observed in the learning curves.

[b] Used for all other descriptors.

TABLE 3.2: Hyperparameters for the gradient boosting regression.

I used the GBR implementation from the `scikit-learn` Python package[244]. I used a 3-fold cross-validation procedure as implemented in Sci-kit learn package which I

repeated 50 times (with different random seeds) in order to obtain relevant accuracy scores and errors by averaging them over all the simulations. To choose and validate the hyperparameters, I used cross-validation and chose as a measure of accuracy the root mean squared error (RMSE). In particular, I focused on the impact of the number of boosting stages, which — although GBR is generally said to be fairly robust to overfitting — I found to be an important hyperparameter. The learning curves for this hyperparameter, for the prediction of thermal expansion coefficient based on different sets of descriptors, are displayed in fig. 3.7. Except for the *ad hoc* geometric descriptors, all the curves show a similar learning rate, and reach the final accuracy at around 500 boosting stages. However, in the case of geometric descriptors, it can clearly be seen that over-fitting is occurring, as the RMSE on the test set starts to increase at around 250 boosting stages while it decreasing in the training set. Therefore, I decided to use 250 boosting stages for this particular set of descriptors. For the other sets, this phenomenon is not happening. This is why I chose to use a higher number of decision trees (500). The final set of hyperparameters used for this study is reported in table 3.2.

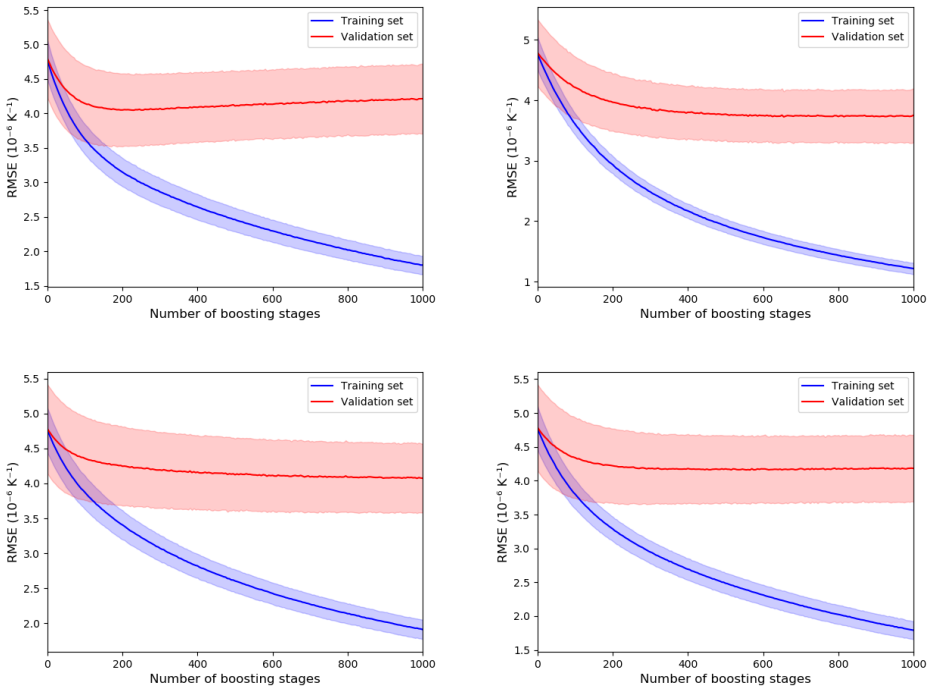


FIGURE 3.7: Learning curves of the thermal expansion using all the types of descriptors for the thermal expansion. Geometric descriptors, SOAP descriptors, topological descriptors and volumetric descriptors correspond respectively to the upper left, upper right, bottom left and bottom right panel. Lines correspond to the mean RMSE values and shaded surfaces correspond to the standard deviation.

### 3.2.2 Results

#### Machine learning model based on geometric descriptors

I started my study by using the *ad hoc* set of geometric descriptors as such metrics are known to be useful for describing zeolite structures. These simple geometric descriptors were already used in the past for the prediction of mechanical

properties[165, 195]. Here, I applied them this time to the thermal properties. As detailed in the previous section, I used a cross-validation strategy along with the hyperparameters defined in table 3.2. The result of a GBR trained on the geometric descriptors for the thermal properties is displayed on fig. 3.8. It can be seen that there is overall a good prediction of the thermal expansion as most values between $-2$ and $-1x10^{-5}$ $K^{-1}$ are well concentrated. As the majority of the data set is found within this interval, these results were expected. However, the points located outside of this range present a large deviation. Whether it is on the side of very negative values of thermal expansion or on the other side, they both stray away from the dashed line representing a perfect prediction. It seems that there are outlier materials, and that there are too few values on the extreme sides of the data set to allow for the model to train efficiently on them. This is why these points are exhibiting such high errors.

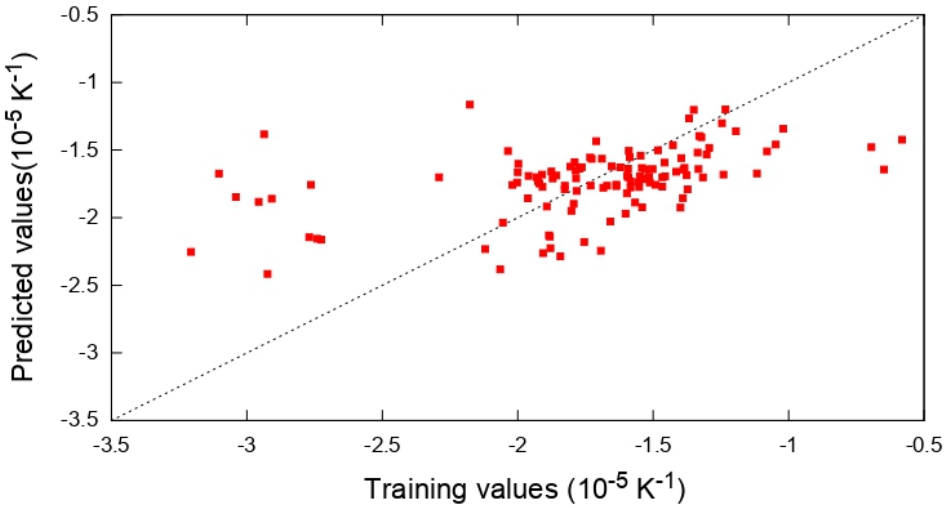


FIGURE 3.8: GBR cross-validation prediction of the thermal expansion $\alpha$, based on geometric descriptors.

As the score function, I chose to use the root mean squared error (RMSE) to quantify the quality of each model. Calculated on this prediction, I obtained a value of 4.24 $10^{-6}$ $K^{-1}$ corresponding to an error of around 20%. Considering the small data set that I am using, I believe that it is a reasonable error. This degree of error is also observed in other ML models based on simple geometric descriptors and predicting macroscopic physical properties[194]. A large part of the error observed comes from the outlier materials mentioned earlier. Indeed, removing the points with $\alpha < -2.5$ $10^{-5}$ $K^{-1}$ leads to a decrease of the RMSE to a value of 2.56 $10^{-6}$ $K^{-1}$, improving the predictive power. There is a clear separation of the outliers from the rest of the data set. To try to find the microscopic origin of this behavior, I investigated these isolated frameworks. As examples, I represented the thermal expansion along with the bulk modulus and the accessible volume on fig. 3.9 and highlighted the outlier materials in blue. As it can be seen, while their thermal expansion deviates away from the average value, the bulk modulus and accessible volume do not show any differences compared to the rest of the frameworks. Investigating their structure visually did not reveal any particularity either. Therefore, I believe that the specific behavior observed may be unphysical, and find its root in high order terms, which are not included under the quasi-harmonic approximation and thus not taken into account in my DFT calculations.

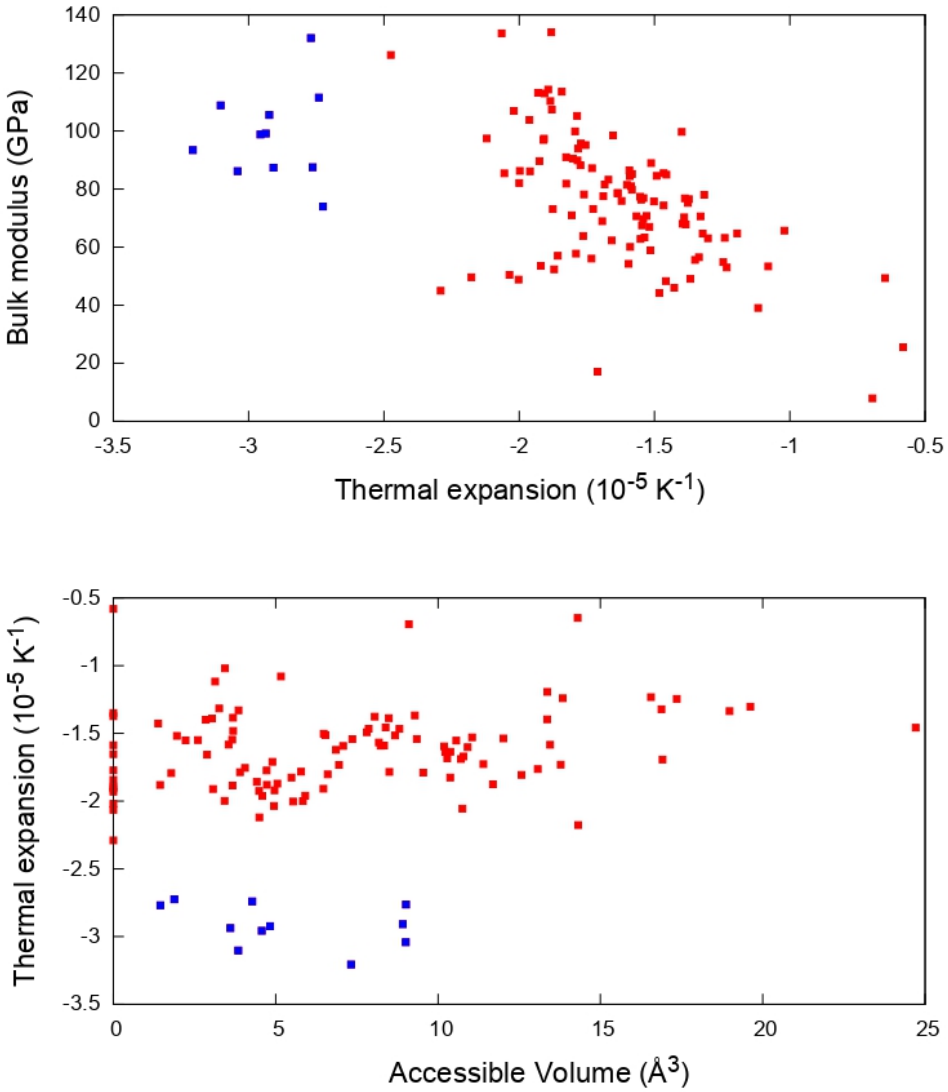


FIGURE 3.9: Top panel: thermal expansion compared to the bulk modulus value. Bottom panel: thermal expansion compared to the accessible volume. In both representations, outlier materials (as defined in the text) are shown in blue.

It is also important to note that the deviation observed finds its origin not only from the reduced size of the data set, but also from the noise of the training data itself. Indeed, the DFT calculations realised to create this data set come with a certain degree of uncertainty. QHA, used to calculate thermal properties, is a rather long and difficult process. In the scope of a systematic study, it is necessary to fix a certain set of parameters for all the frameworks (ranges of temperature, volume expansion, number of points in numerical derivatives, etc). Such parameters are then satisfying for a high number of frameworks but can be not adapted in some extreme cases that differ a lot from the "average framework". More fine-tuned parameters for each zeolite could have resulted in a better accuracy for the calculated properties, but was not feasible in a systematic approach. However, it can be noted that, in this study, the main interest is to evaluate the feasibility of the ML models and to compare different descriptors, for a physical property (thermal expansion) that has never been studied in framework materials at that scale before. Thus, the deviation observed does not hinder the purpose of this study.

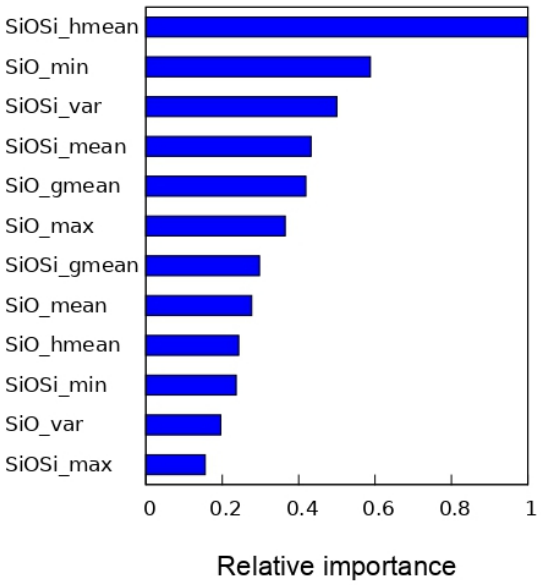


FIGURE 3.10: Relative importance of each descriptor for the GBR model.

One important tool in machine learning is the feature importance, which can be accessed in some (but not all) machine learning algorithms. This metric can inform on the relative importance of each descriptor in the training process, and sometimes allow to diagnose over-fitting cases. On fig. 3.10, I reported the relative importance of each descriptor for the prediction of thermal expansion. It can be seen that the first descriptor in terms of importance is related to the Si–O–Si angles (harmonic mean of angle values). Looking at the order of the rest of the descriptors, the statistics on angles seem to be of higher importance for the prediction of thermal expansion than the Si–O distances. This confirms the physical intuition, because thermal expansion is dominated by low-frequency vibration modes, which typically involve tetrahedral rotations of $SiO_4$ units and Si–O–Si angle bending. This is also confirmed with the partial dependence plots depicted in fig. 3.11. This type of representation is used to see how a chosen property (here the thermal expansion) responds as a function of some specific descriptors. It appears clearly that the Si–O–Si angles exhibits a very

marked, near linear dependence with the predicted thermal expansion while the Si–O distances show a lower amplitude as well as a monotonic trend. Knowing that the minimal Si–O distance was found to be the second most important feature when training the model, it highlights the fact that angles have much more importance than distances in the prediction of the thermal expansion.

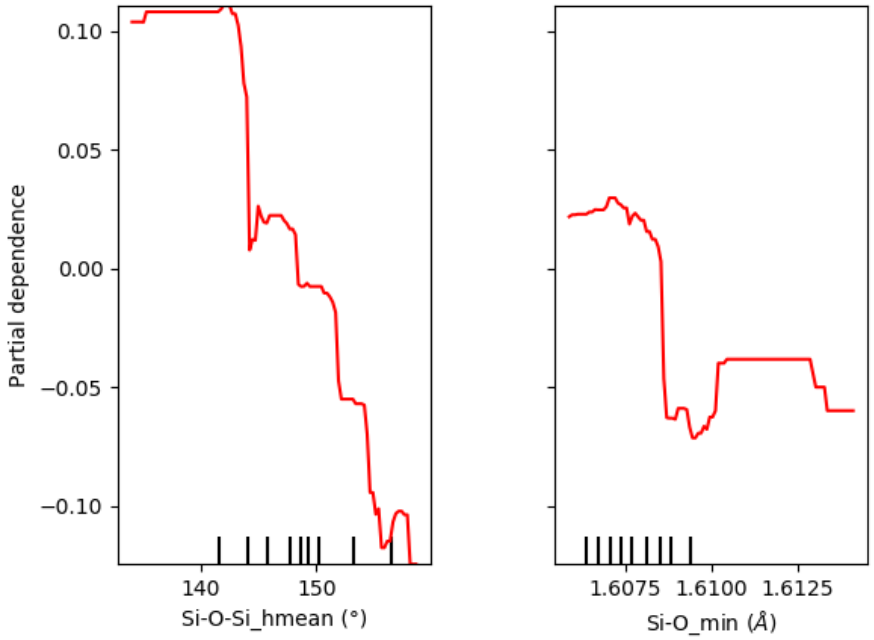


FIGURE 3.11: Partial dependence plots of the harmonic mean of Si–O–Si angles (left panel) and minimal value of Si–O distances (right panel)

**Comparison with other descriptors**

For the next step, I decided to try predicting the thermal expansion based on the other sets of descriptors. I used the exact same methodology (except for the number of boosting stages as described before) and obtained the results shown in fig. 3.12. I observed that, among all the different sets, the principal components of SOAP features seems to be the most performant in the prediction of thermal expansion, with a RMSE value of 3.75 $10^{-6}$ $K^{-1}$. This value is even smaller than the RMSE obtained previously with the *ad hoc* geometric descriptors. This is due to the fact that the SOAP features contain the informations on angles and distances while also including more complex effects, resulting in a thorough description of the local geometry. In contrary with this result, it appears that the volumetric and topological descriptors are similar to the geometric ones in terms of prediction. They exhibit values of RMSE of 4.16.$10^{-6}$ $K^{-1}$ and 4.13 $10^{-6}$ $K^{-1}$ respectively.

To further compare the different descriptors, I then performed predictions on another property: the bulk modulus $K_0$. I maintained all the parameters and obtained the results shown in fig. 3.13. The same observation on descriptors' performance can be made here: principal components of SOAP features are the most efficient with a RMSE value of 15.9 GPa, lower than for the topological and volumetric descriptors (RMSE of 17.6 GPa and 20.9 GPa respectively). SOAP features seems then better in

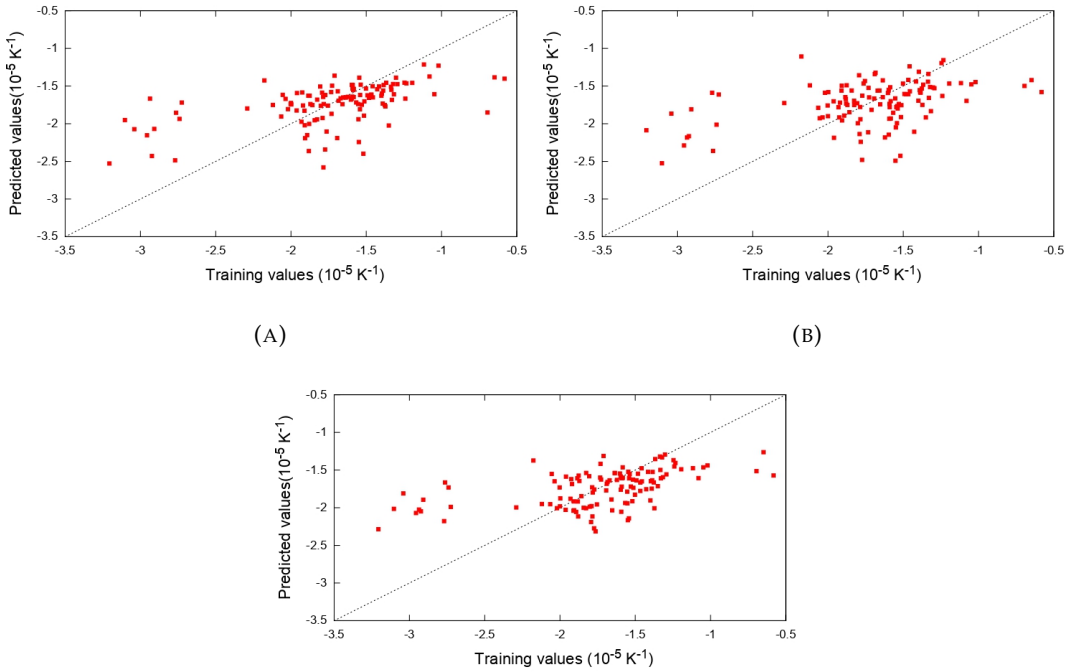


FIGURE 3.12: GBR cross-validated predictions on the thermal expansion using three different types of descriptors: (a) SOAP (RMSE = 3.75.$10^{-6}$ K$^{-1}$), (b) Volumetric (RMSE = 4.16.$10^{-6}$ K$^{-1}$) and (c) Topological descriptors (RMSE = 4.13.$10^{-6}$ K$^{-1}$).

the general case as they obtained the lowest error on both the thermal expansion and the bulk modulus. It can also be observed on the representations that volumetric descriptors result in a really poor prediction of bulk modulus, as the cloud of points strays away from the perfect prediction represented by the dashed line and tends to form an horizontal line. This is an interesting and counter-intuitive result as one could have thought that density and porosity-related metrics were directly linked to the stiffness of a material. Here it shows that this is not always the case and that porosity-related metrics alone are not sufficient to efficiently predict mechanical properties. It can be noted, however, that *ad hoc* geometric descriptors exhibit a good accuracy (RMSE = 16.0 GPa), very close to the accuracy obtained with SOAP features. This shows once more that Si–O–Si angles are crucial in determining the properties of zeolites, as the difference in prediction between *ad hoc* geometric descriptors and principal components of SOAP features is really small (despite SOAP features containing more informations).

Up until now, I compared the different sets of descriptors independently. However, it is also interesting to compare them when they are combined in the same model. As I chose to build models with a total number of 12 descriptors, I selected the three best descriptors of each type and combined them to form a new set of 12 descriptors. The summary of this new set is given in table 3.3. As before, I maintained the same hyperparameters after confirming through learning curves that no over-fitting was occurring. The cross-validated prediction of the thermal expansion as well as the relative feature importances of this new set of descriptors are represented on fig. 3.14. It can be observed that, while the accuracy is slightly better than before with a RMSE of 3.64.$10^{-6}$ K$^{-1}$, the improvement is anecdotic and mainly located on the low values

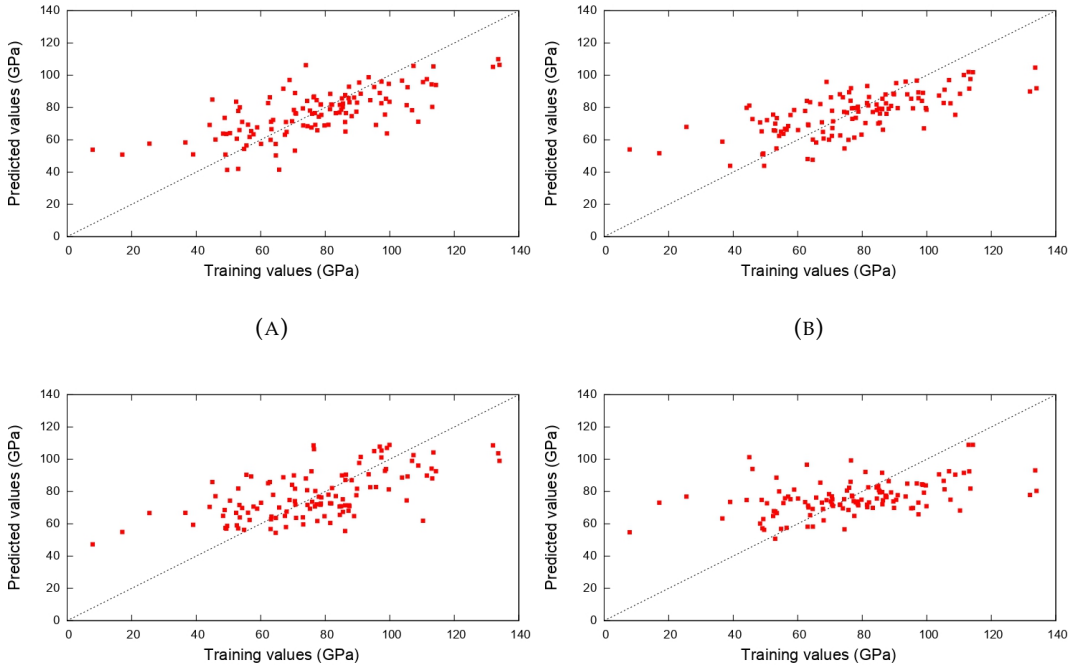


FIGURE 3.13: GBR cross-validated predictions on the bulk modulus $K_0$ using the four different types of descriptors: (a) Geometric (RMSE = 16 GPa), (b) SOAP (RMSE = 15.9 GPa), (c) Topological descriptors (RMSE = 17.6 GPa) and (d) Volumetric (RMSE = 20.9 GPa).

of thermal expansion — many of which were outliers in the previous models.

| Descriptor | Unit |
|---|---|
| Si-O-Si harmonic mean | ° |
| Si-O mean | Å |
| Si-O-Si variance | ° |
| Density | $g.cm^{-3}$ |
| Volume | $Å^3$ |
| Accessible + non accessible surface area | $Å^2$ |
| Principal component 1 | - |
| Principal component 2 | - |
| Principal component 3 | - |
| Number of neighbours in second sphere of coordination | - |
| Number of neighbours in third sphere of coordination | - |
| Number of neighbours in fourth sphere of coordination | - |

TABLE 3.3: List of descriptors used when mixing different types of descriptors in the model. 3 descriptors of each type was used for a total of 12 descriptors.

The relative feature importances revealed that the second principal component of SOAP features is the most important descriptor in the prediction of thermal expansion. This was already the case when training with only SOAP features. The dimensionality of SOAP features needed to be reduced in order to be exploitable, and

compared to smaller descriptor sets. Dimensionality reduction techniques usually causes a certain loss of information. However, the feature importance proves that this loss does not overly impact the descriptive potential of these features. Principal components of SOAP features represent then highly competitive descriptors of a structure, weighing even more than Si–O–Si angles in the training process. Partial dependence representations of the second principal component of SOAP features as well as Si–O–Si harmonic mean are shown in fig. 3.15. It can be seen on this figure that their dependence is quite similar. They both show a strong linear dependence, while all the other descriptors used exhibit a weak or even null dependence. As said previously, Si–O–Si angles are one of the major descriptors to include for the determination of thermal properties in zeolites, and I believe that is also a crucial feature for the prediction of other properties.

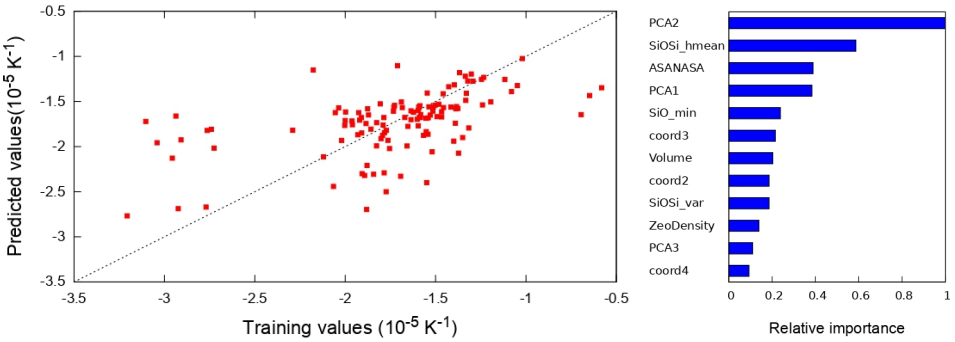


FIGURE 3.14: GBR cross-validated prediction of the thermal expansion using the four different types of descriptors combined along with the relative feature importances.

### Applying ML models to non DFT-optimised structures

Up until now, I predicted the properties of zeolite structures that had been previously optimized by DFT relaxation. So far, the predictions proved that a machine learning application of this database is possible with reasonable accuracy. However, while it is interesting, I was considering for now that the structures for which the properties needed to be predicted were DFT-optimised, just like the database I created. This need of DFT-level accuracy causes a serious obstacle to potential high-throughput screening studies. Indeed, while DFT optimisations are relatively cheap compared to the quasi-harmonic calculations of thermal properties, they are not scalable to the size of available hypothetical databases of zeolitic materials (which can contain hundreds of thousands of structures). Therefore, the applicability of this model to structures with a lower accuracy level needed to be tested.

Therefore I decided to check how a model based on non DFT-optimised structures would perform in the prediction of the thermal and mechanical properties obtained in DFT. This particular model will give then an indication of the sensitivity of the ML model to the accuracy of the geometries used as input. In practice, I used a new data set of structures retrieved from the IZA database (available at `http://www.iza-structure.org/databases/`) and containing the exact same 120 topologies used up until now. These structures were obtained by a distance least-squares refinement technique (DLS76 developed by Baerlocher and coworkers in 1977) and I left them as such without optimising them. I used the same set of hyperparameters

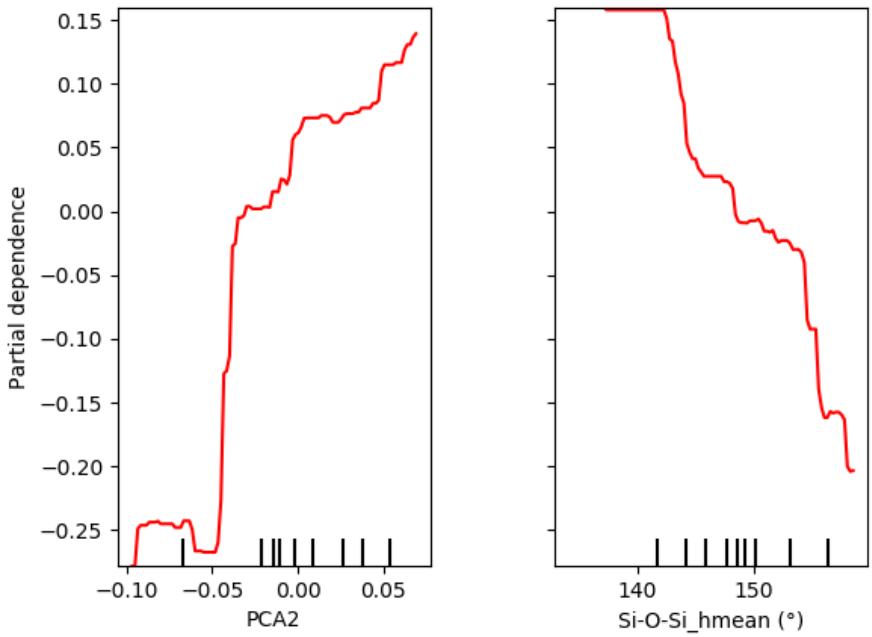


FIGURE 3.15: Partial dependence plots of the (left panel) second principal component of SOAP features and (right panel) Si–O–Si harmonic mean.

as well as the same combination of descriptors given in table 3.3, as it was proven that the results were slightly better than with just one type of descriptor. All the features required were retrieved on the non-optimised structures following the same procedure described in the computational methods.

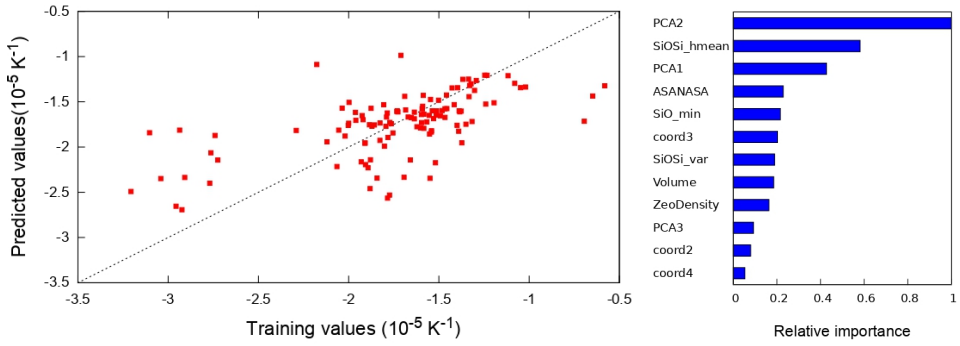


FIGURE 3.16: GBR cross-validated prediction of the thermal expansion using the non DFT-optimised structures along with the relative feature importances.

The cross-validated prediction of thermal expansion using the non DFT-optimized data set of structures is represented on fig. 3.16 along with corresponding relative feature importances. The RMSE obtained for this new model is $3.63.10^{-6}$ $K^{-1}$ which is very close from the value of $3.64.10^{-6}$ $K^{-1}$ obtained in the previous section. Both predictions also have identical behaviors, as we can see that there are outliers for

low values of thermal expansion in this case as well. It shows that the non DFT-optimised structures are not so different from the structures I optimised. As evidenced by the close values of RMSE, the small differences of geometry do not alter the performance of the model. This idea is also supported by the relative feature importances displayed in fig. 3.16. Indeed, the importance order of the descriptors is quite similar as before. The second principal component of SOAP features is again the most important descriptor, followed by the harmonic mean of angles. This is due to the fact that structures from the IZA database and DFT-optimised structures are close to each other, resulting in close descriptors as well. Therefore, this validates the prediction of physical properties from structures optimised at a level lower than DFT (for example, force field optimised structures). This represents a time-saving conclusion, as it makes it possible to investigate very large-scale zeolitic data sets, such as the PCOD2 database created by Deem et al.[245, 246].

#### Application to large data sets

The previous section proved that the ML model is not too sensitive to the geometries used as input. It resulted then that large data sets obtained with lower level of accuracy than DFT can be used with this model. Therefore, I decided to apply this model to a large data set of zeolites: the PCOD2 database created by Michael Deem. This set contains around 600,000 hypothetical pure silica zeolite structures, obtained from combining Monte-Carlo simulations, simulated annealing, and structure refinement using a classical force field (the Sanders–Leslie–Catlow (SLC) interatomic pair potential[189]). This database constitutes a great tool for the machine learning-based exploration of new synthesizable structures, or the identification of candidate zeolites with targeted properties. It was used in the group to identify new frameworks with auxetic behavior[195]. Here, the objective is to compare the data set of DFT-calculated zeolites with the full database of hypothetical structures and ultimately, to try to predict the distribution of thermal expansion and bulk modulus of the PCOD2 database.

First of all, I determined the different types of descriptors for the whole PCOD2 database, following the same methodology used up until now. Considering the size of this database, this step required a significant time to complete. For around 600,000 structures, the computational effort is as follows (timing reported for nonparallel, single-CPU calculations): a couple of days for the bond distances and angles, one week for the SOAP features and around two weeks for the topological descriptors. However, the calculation of volumetric descriptors was not possible. Many of the systems in the database are too large and the time needed for calculating them was excessively long. Results for other descriptors, such as the Si–O–Si angles and Si–O distances, are reported on fig. 3.17 for both my data set and the PCOD2 database. As my database is much smaller, both occurrences were normalised in both cases. The first observation that can be made, is that the span of values of both Si–O–Si angles and Si–O distances is very close for my data set and the PCOD2 database. This means that the geometrical descriptors of both databases will show the same distribution. This is an encouraging result as it means that a model trained on these descriptors should be generalizable without extrapolation. What can be seen also is that Si–O distances do not vary much, keeping values between 1.60 and 1.62 Å. This confirms the previous conclusions that distances do not impact the predictions as much as the angles. Indeed, in contrary with the distances, Si–O–Si angles show a larger range of values with a total amplitude of around 20°. This explains the

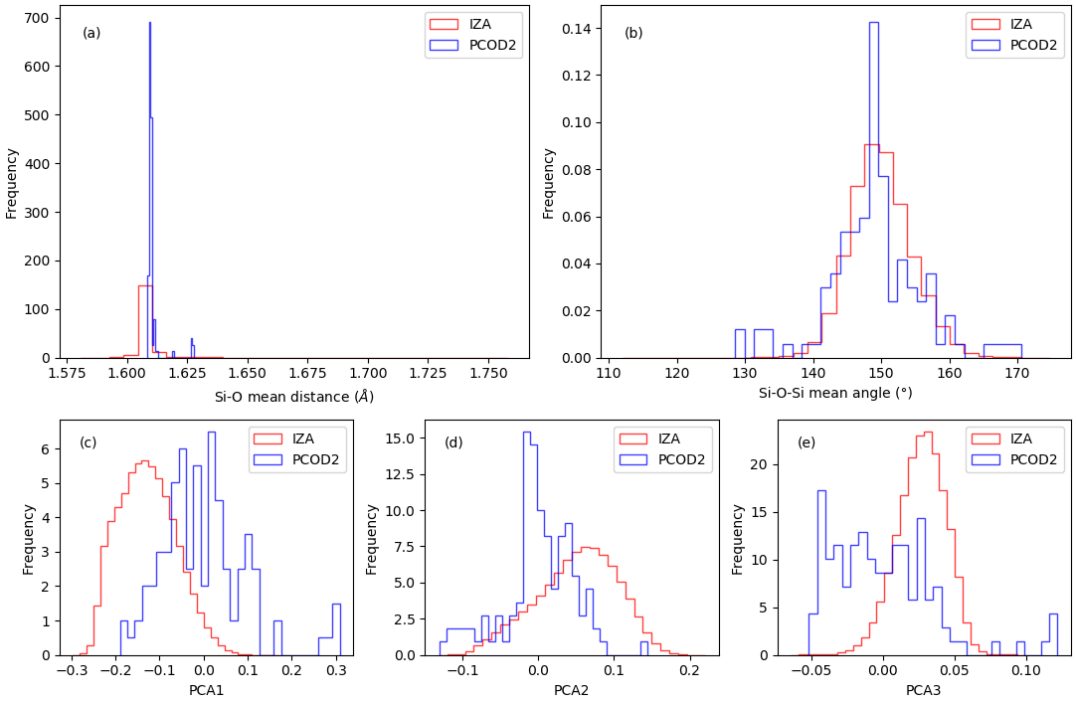


FIGURE 3.17: Distribution of descriptor values of our data set compared to the Deem database. Descriptors represented are: (a) Si–O mean distance; (b) Si–O–Si mean angle; (c), (d) and (e) first three components of principal component analysis of SOAP features. Due to the difference in data set size, frequencies have been normalized for both data set.

importance of these angles in the predictions, as they account for a lot of the diversity between the different structures.

The distributions for the three principal components of SOAP features for both databases are also represented on the bottom panels of fig. 3.17. In contrast with the geometrical descriptors, the distribution of the principal components of my database seems to be shifted compared to the case of PCOD2. Two conclusions can be drawn from this result. First, it means that the geometric diversity in the PCOD2 database is larger than in the DFT-calculated data set. My set of structures do not represent all of the possible environments included in this database of theoretical structure. The second conclusion is that some of the structures from the PCOD2 database could be outside of the realm of "feasible zeolite structure". This is an interesting new take on the question of experimental feasibility of frameworks, a question which was already addressed in the past in the group[161]. This subject is still pursued in the group, although it is outside of the scope of the current study. I can only conclude, for now, that a prediction on the PCOD2 database using the principal components of SOAP features as descriptors would probably result in a poor prediction.

Since SOAP features are out of the question, I decided to use the geometrical descriptors to try to predict the thermal expansion and bulk modulus of the database of theoretical zeolite structures. I used the same hyperparameters described earlier with my data set of DFT-calculated zeolite as the training set. The same set of geometrical descriptors, including statistics on Si–O distances and Si–O–Si angles, was used (see table 3.1. The results of the prediction were plotted as a heatmap, using a gradient of color as the third dimension. This is shown in fig. 3.18. Some general

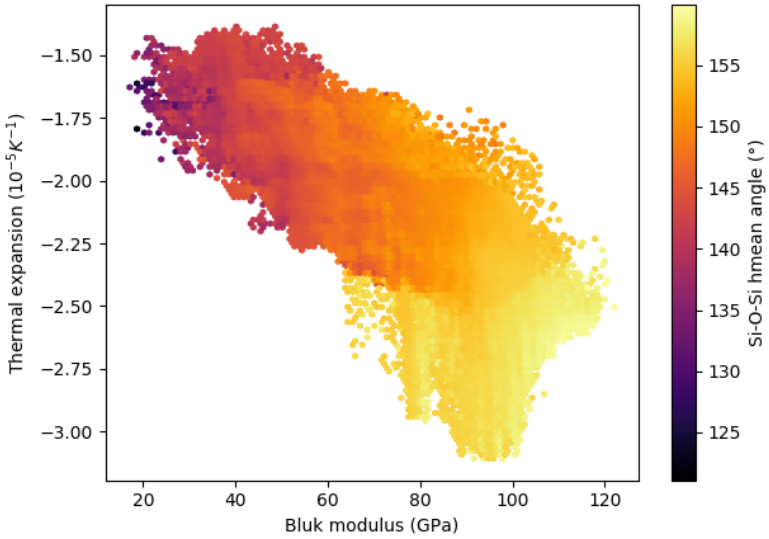


FIGURE 3.18: GBR prediction of the PCOD2 database on the thermal expansion and bulk modulus using the geometrical descriptors and our data set as training set.

trends can observed from this figure, such as the statistical correlation between the thermal expansion and bulk modulus. Structures with a higher bulk modulus tend to show a larger (negative) value of thermal expansion. This behavior was already slightly appearing in my data set, as evidenced by the fig. 3.9. It also seems that both quantities are linked to the same geometrical feature, namely, the Si–O–Si angle. As already explained previously in this chapter, angles were found to be one of the most important descriptor in the prediction of properties of zeolites. Here, it appears that materials with a higher value of angles exhibit both a higher bulk modulus and a larger negative thermal expansion. This correlation between mechanical and thermal behavior through a relatively simple geometric feature is an interesting new development, and would have to be confirmed — for example, through systematic calculations of representative structures within the PCOD2 database.

### 3.2.3 Conclusion

Machine learning techniques are popular and widely spread nowadays. They are used in chemistry for the prediction of properties, hypothetical structures or to predict chemical synthesis route. The field of computational chemistry was one of the first to recognise the potential of such methods and it can be now commonly implemented in a scientific methodology. One of the most important challenge when using these techniques in chemistry is the representation of molecules and structures. Indeed, the conventional representation using fractional coordinates and translation vectors is very common in chemistry, but it is not suitable for machine learning. The representation must be invariant with the referential. Several solutions were proposed, such as the use of Coulomb matrices for example. This domain is still active and is of great importance as the quality of the representation (along with the quality of data) conditions the quality of the prediction. In the case of zeolites, large databases of structures have been created. The International Zeolitic Association database contains the structures of all the accepted topologies of zeolites. The

PCOD2 database contains around 600,000 hypothetical structures of pure silica zeolites. In my case, I created a database of properties of zeolites calculated through static DFT calculations. DFT is a high accuracy method leading generally to better results than force fields, which are usually used for the creation of databases as it requires a limited computational cost. However, the use of DFT prevented me from calculating a large number of structures. The 121 simulated zeolites represent a relatively small database when compared to other databases of zeolites. This can be an obstacle to its use in machine learning as a training set. I then investigated this possibility and estimated the performance of a model trained on my database.

As I was using supervised machine learning, I needed to define descriptors to represent the structures of my database. I chose to use different types of descriptors as it would allow me to compare their performance. I chose *ad hoc* geometric descriptors using statistics on Si–O bonds and Si–O–Si angles, volumetric descriptors such as accessible volume or surface area, topological descriptors using the number of neighbours in the successive coordination spheres, and descriptors arising from the principal components analysis of Smooth Overlap of Atomic Positions (SOAP) features. The GBR model trained on the geometric descriptors resulted in a good prediction of the thermal expansion with a low value of RMSE. The analysis of the relative feature importance and partial dependence revealed that statistics on angles are decisive in the prediction of thermal expansion compared to distances. In the case of thermal expansion, it is simple to link the low frequency vibration modes involving tetrahedral rotations of $SiO_4$ with the Si–O–Si angles. Comparing different models trained on different descriptors highlighted the efficiency of SOAP features as it showed the best accuracy. This trend was also observed in the prediction of the bulk modulus, although *ad hoc* geometric descriptors resulted in an accuracy close to SOAP features. Combining the best descriptors of each type into one model (containing the same number as the previous ones) showed an improvement of the accuracy. Relative feature importance showed that the SOAP features are clearly the most important parameters along with the Si–O–Si angles. By applying a model trained on non DFT-optimised structures and predicting the DFT calculated properties, I proved that the training set of structures does not need to reach the same level of accuracy. Indeed, structures of the IZA database, obtained by a distance least-squares refinement, performed identically in the prediction of thermal expansion compared to the DFT structures. This is an important and time-saving result as it shows that lower accuracy techniques can lead to the same results. I finally applied my model to the database of theoretical zeolites PCOD2. Comparison of the descriptors showed that, while SOAP features were the most important descriptors, the values between my database and PCOD2 differ. In contrast, *ad hoc* geometric descriptors were relatively similar, leading me to choose to perform the prediction with a model based on those descriptors. Results revealed that a structure with a higher bulk modulus tend to have a larger negative thermal expansion.

Through this work, I proved that the prediction of the thermal expansion and bulk modulus was possible using my database of DFT-calculated zeolites as a training set. As shown, it can be used to predict the properties for large databases of theoretical structures and is thus able to guide the synthesis towards materials with specific desired characteristics. It has been proved also that theoretical structures can be optimized with techniques requiring a relatively low computational cost without losing the quality of prediction. This represents a time-saving conclusion allowing for the consideration of larger databases. The range of applications of machine learning in chemistry is vast but requires more and more data of higher accuracy in order to

improve the predictions' performance. This database of zeolites alone can be used in various machine learning studies or could complete already existing databases of similar accuracy.

# Chapter 4

# Collaborations

As the subject of my PhD is quite broad both in terms of phenomena and concerning the materials chosen, I was able to dive into different projects with systems that stray away from the zeolite structures studied up until now. Such projects usually originate from the group's collaboration wih experimentalists. Indeed simulation is a powerful tool to explain phenomena that cannot be fully understood with experiments alone, providing a microscopic view of the systems and a direct link to their macroscopic properties. The reverse is also true as experiments are guidelines for simulations. It's the combination of both approaches that usually leads to the birth of new knowledge, especially on very complex systems. Therefore simulations and experiments are often combined as part of collaborations of different groups of research. During my PhD, I had the opportunity to collaborate with experimentalists in three such studies. I studied their systems to try to explain certain phenomena that they were observing. Despite having different objectives, each collaboration was done on MOFs materials, looking at their structure and properties.

This chapter will describe these different collaborations. Each section will talk about a different project with a different laboratory. It will start each time with an introduction of the project and will be followed by the results. Some results obtained by collaborators, especially experimental results, will be mentioned and described in this chapter. It will be specified directly when such results were not obtained by me.

## 4.1 The flexibility of MUF-15

### 4.1.1 Introduction

This project was realised in collaboration with the group of Professor Shane Telfer from the University of Massey (New Zealand). He contacted our team because he had trouble determining the structure and flexibility of a material.

He was studying a MOF material that was previously synthesized in his group [247]. This MOF called MUF-15 (**M**assey **U**niversity **F**ramework) was originally studied for the separation of ethane and ethylene. The structure of this material is depicted on fig. 4.1. It is composed of isophtalate organic linkers and contains cobalt-based metal nodes of formula $Co_6(\mu_3\text{-}OH)_2(RCO_2)_{10}$. This particular arrangement leads to the presence of cages that can be clearly seen on the right panel of fig. 4.1. MUF-15 features a high pore volume, which results in the highest $C_2H_6$ uptake among the top-performing ethane-selective MOFs. This MOF features relatively small pores suitable for the adsorption of small gas molecules such as ethane. Moreover, these pores are decorated with benzene rings facilitating the contact between the guest

molecules and the framework walls. They found that MUF-15 exhibits a great selectivity of ethane over ethylene while having a high uptake. An example of adsorption isotherms of MUF-15 is shown on fig. 4.2. The Massey group attributed this to the particular pore architecture of their material and to the absence of regions of high polarity such as open metal sites. Besides its good uptake and selectivity, this material also stands out by its availability. Indeed it is readily obtainable from cheap and simple reagents. Added to its nice adsorption properties, this makes this material an interesting choice for the separation of ethane and ethylene.

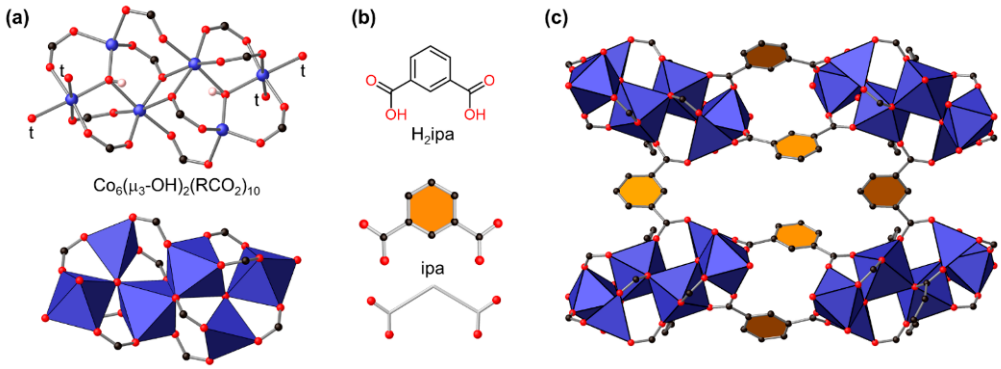


FIGURE 4.1: (a) Structure of the metal SBU (cobalt = dark blue, carbon = black, oxygen = red). Sites occupied by $H_2O$ ligands are marked with a "t". (b) Structure of the isophtalic acid. (c) Representation of the assembly between inorganic and organic SBUs. Taken from [247].

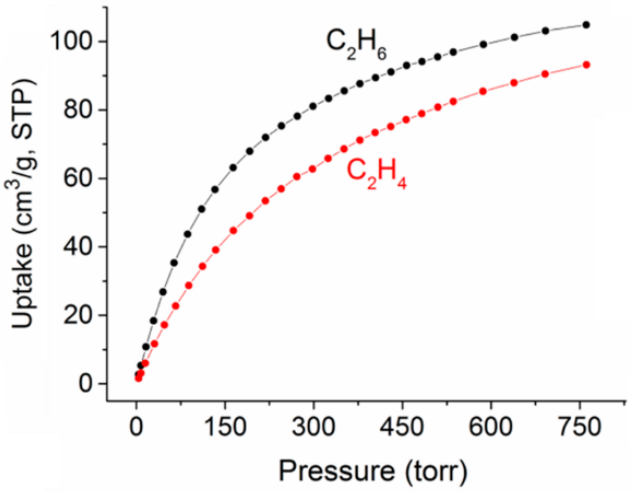


FIGURE 4.2: $C_2H_6$ and $C_2H_4$ adsorption isotherms of MUF-15 at 293 K. Taken from [247].

After working with this "parent" structure, they decided to synthesize derivatives of the MUF-15 framework by changing the nature of the organic linker. They kept using isophtalate but added different functional groups to it. They were able to synthesize derivatives with –F, –Br, $–CH_3$, –OH, $–NO_2$ and –OMe. Among all these derivatives, they were particularly interested in the –OMe compound as, in contrary with all the others, this one appeared to show some flexibility under 1 bar of pressure (which was their limit in terms of measurement). They were able to identify this flexible behavior with the isotherm of Ar at 87 K, which is represented on fig. 4.3. A step is clearly visible on the adsorption isotherm of Ar indicating that the structure

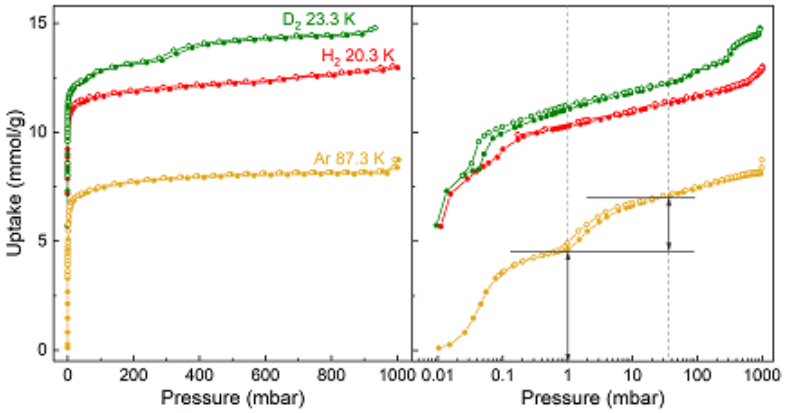


FIGURE 4.3: $D_2$, $H_2$ and Ar adsorption isotherms of MUF-15–OMe at low temperatures. Taken from [248].

undergoes some changes towards an "open form" of the initial structure. Pr. Telfer and his group tried to identify this open form but the XRD experiments did not allow to conclude on the structure. This flexibility was only observed under 1 bar on the –OMe compound. However, it is expected that the parent form and all its derivatives may show the same kind of flexible behavior, albeit possibly at higher pressure.

The goal of the simulations was to find the mechanism behind the gate opening of the -OMe compound as well as the open form of the structure and to study this MOF's behavior under pressure.

### 4.1.2 Calculation parameters

I started from CIF files of the parent MUF-15 and the –OMe derivative that Pr. Telfer shared with me. These files were obtained from XRD experiments. All the calculations were realised at the DFT level with the CRYSTAL17 code[171, 170]. All-electron gaussian basis sets of double-$\zeta$ valence with polarization were used for all the atoms. All these basis sets can be found online in the CRYSTAL basis set library (`https://www.crystal.unito.it/basis-sets.php`). The exchange and correlation were treated at the hybrid level using the PBEsol0 functional[174]. Considering the large size of the system, a unique $k$-point for the sampling of the Brillouin zone was enough to converge the properties of interest.

Convergence criteria for the geometry optimization were kept at their default values: a maximum of 0.00045 a.u. on atomic displacements during one optimization step and 0.0018 a.u. on forces. Once the structures were converged, I used the Zeo++ software package[192, 193] to calculate the accessible volume and surface area based on geometric considerations. Representative input files for the calculations are available online at `https://github.com/fxcoudert/citable-data`.

### 4.1.3 Results

I first started by optimizing the geometry of both the parent MUF-15 structure and the –OMe derivative. After a few attempts, the optimization of the MUF-15 structure went rather smoothly within reasonable computational time. In contrary, the optimization of the –OMe compound (for which the experimental structure was not

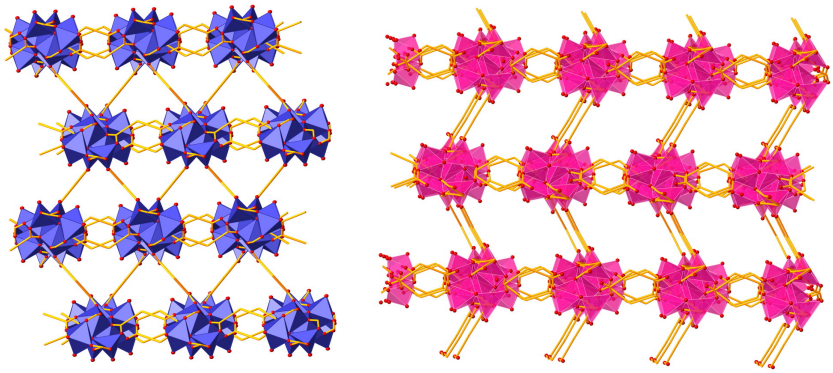

FIGURE 4.4: Left side: Representation of the packing inside the MUF-15 parent structure. Right side: Representation of the packing inside the MUF-15-OMe structure.

available) was much trickier. I first attempted to add –OMe functional groups to the parent structure, which failed quickly. Indeed, the resulting structure could not converge at all. As this material is pretty recent and complex, there was still some doubts on the exact structure. Several arrangements and symmetries were plausible which did not help in my study as the simulation requires the right structure. I made a few other attempts with slightly different possible structural arrangements that Pr. Telfer shared with me. Unfortunately none of the structures I tried to optimize could converge.

Later during the project, Pr. Telfer and his group were able to obtain the single crystal X-ray diffraction structure of the –OMe derivative they synthesized. Its structure is shown in fig. 4.4 along with the structure of the parent MUF-15. It appeared that while their structure is very close, adding –OMe functional groups induces some structural changes. It has then been determined that the –OMe derivative and MUF-15 are topological isomers, rather than strictly isoreticular. Therefore I took the new structure and restarted the optimization from the beginning. However, the same problems appeared on this new configuration and no improvement on the convergence was observed. I concluded that the lower symmetry of the –OMe compound as well as its additional degrees of freedom (rotations of methyl groups) are preventing the calculations to converge within the accuracy required. Given the fact that the parent MUF-15 is structurally close and that it is expected to show some flexibility – although at higher pressure – we decided to study this structure instead. Pr. Telfer was also interested in the flexible behavior of the parent structure.

To understand the opening mechanism, I followed a methodology previously used in exploring the flexibility of novel (or hypothetical) compounds[249]. First I started by applying negative isotropic pressure on the system to mimic the opening of the framework under a guest-induced adsorption stress. This is illustrated on fig. 4.5. It is shown here that the cell parameters increase constantly until -1.6 GPa. An increase of around 0.3 Å is observed for both $a$ and $c$ while the $b$ parameter exhibits an increase of around 2.3 Å. This shows that this framework has indeed some potential for flexibility. It also demonstrates the anisotropy of the structure with the $b$ direction being by far the softest. This is understandable as the bridging linkers are oriented along this direction. Therefore the opening mechanism is most certainly happening in the $b$ direction. Increasing further the pressure at around -1.7 GPa leads to an even more pronounced increase of the $b$ cell parameter of more than 1.3 Å over 0.1 GPa.

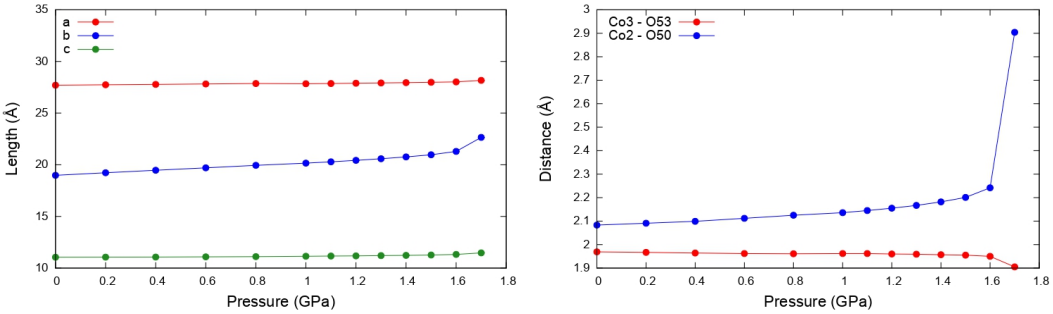


FIGURE 4.5: Left side: Evolution of the cell parameters of the MUF-15 with respect to the pressure. Right side: Evolution of the Co–O distances between the organic linker and the inorganic node with respect to the pressure.

I investigated the evolution of the bond distances inside this material and noted a similar increase for one of the Co–O bond between the linker and the metal node (see fig. 4.5). This increase of around 1 Å at -1.7 GPa indicates the breaking of the bond. This was confirmed by looking directly at the structures which are shown in fig. 4.6: we can see that one oxygen of the linker (shown by a red arrow) detaches from the metal node.

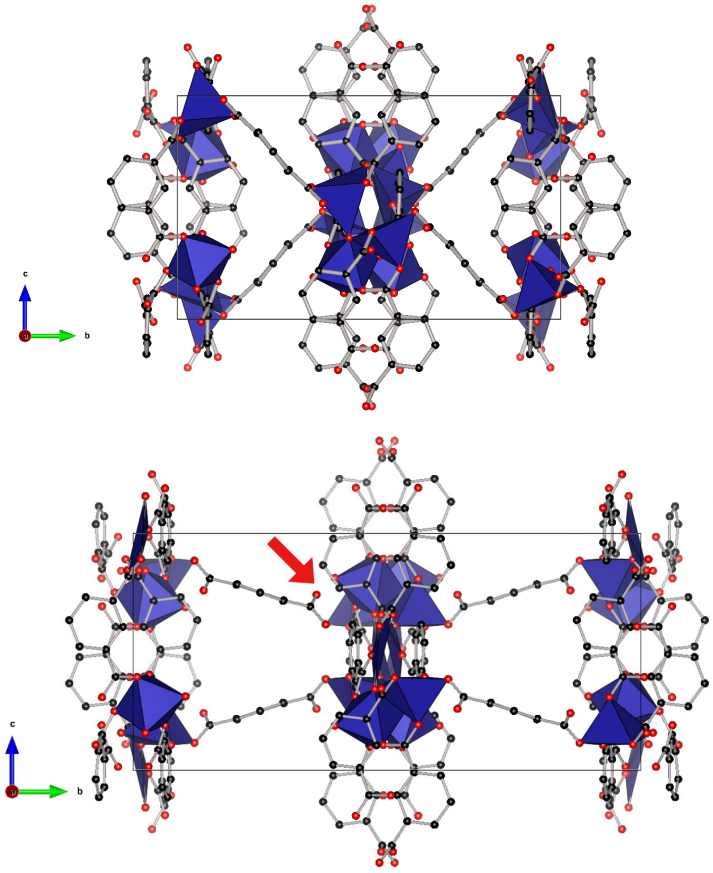

FIGURE 4.6: Top: Structure of the MUF-15 at the equilibrium from DFT calculation. Bottom: structure of the MUF-15 at -1.7 GPa.

As this phenomenon is occurring quite abruptly, constant pressure calculations were not sensitive enough to investigate structural changes happening at around -1.7 GPa.

I therefore switched to performing constant volume calculations in order to have a better control on the opening process. Constant volume calculations consist in optimizing the atomic positions and cell parameters with only a constraint on the total unit cell volume. I calculated several structures around the volume of interest and confirmed the microscopic mechanism shown in fig. 4.6. First, as the volume increases, the linkers orient in the *b* direction and flatten in the *ab* plane. This is followed by the cleavage of one of the Co–O coordination bonds aligned in the *b* direction. This explains the strong anisotropy of the mechanical properties.

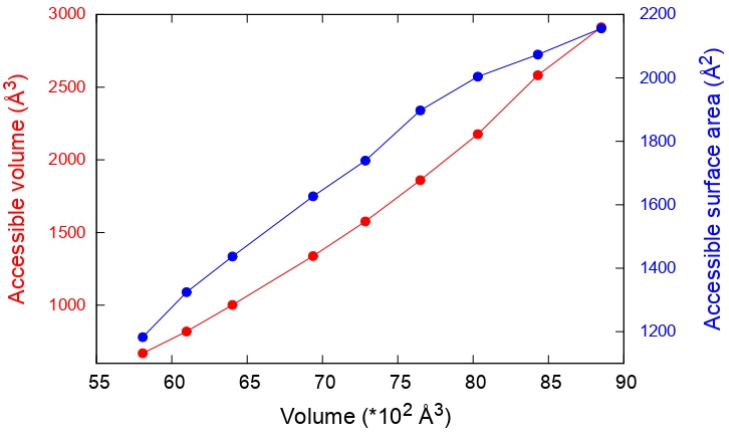


FIGURE 4.7: Accessible volume and surface area plotted against the unit cell volume.

In order to confirm that the structural changes observed were directly connected with a modification of the pore network, I conducted porosity analysis with the Zeo++ software package[192, 193]. I retrieved from this the accessible volume and surface area and represented them in fig. 4.7. It can be clearly seen that both of these quantities are increasing with the unit cell volume. The parent structure at equilibrium has an accessible volume of 669 $Å^3$ and a surface area of 1183 $Å^2$. These values are lower than the experimental measurements, which is consistent with the smaller unit cell of the energy-minimized structure. Upon increasing the unit cell volume from 5807 $Å^3$ to 7648 $Å^3$, the accessible volume and surface area both increases to 1860 $Å^3$ and 1898 $Å^2$ respectively. As a result, the framework flexibility appears to be strongly linked to the drastic change of pore network. It demonstrates that the flexibility is associated to (and triggered by) the adsorption of guest molecules with sufficient host-guest interaction strength.

The flexibility being proven to be associated to the adsorption of guest molecules, the last point of the study was to clarify the accessibility of this opening process. This was realised by calculating the energy penalty associated. As the simulations were realised on a "perfect" structure with no adsorbed molecules, it was expected that the non-deformed phase was more stable than the opened one. I found an energy difference of +54 kJ/mol per Co atom. This is comparable to "breathing" energies observed for other porous MOFs, and fully consistent with materials where flexibility can be triggered by adsorption. I represented in fig. 4.8 the total energy in the case of a framework filled with guest molecules and in the case of an empty framework. As it is not possible to introduce guest molecules in the calculation, their contribution to the total energy was considered to be the internal pressure. At each volume, in the case of the framework filled, the contribution of pressure was removed from the total energy in order to mimic the presence of molecules in the pores. It appears that

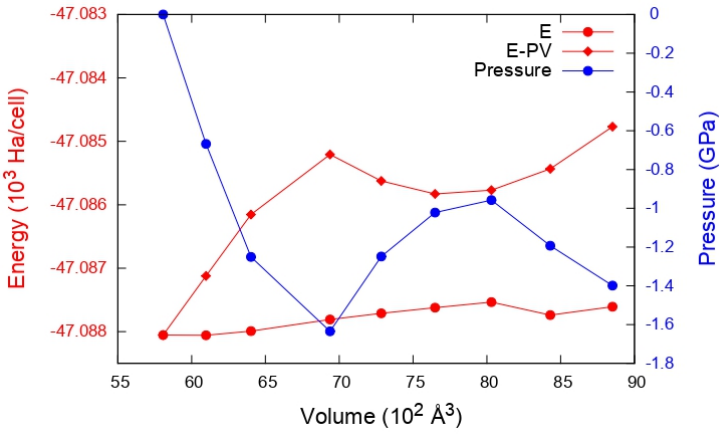


FIGURE 4.8: Energy $E$, pressure $P$ and energy considering the void filled ($E - PV$) with respect to the volume.

the open form of the structure observed earlier corresponds to a metastable phase. Indeed a local minimum of the energy is observed when the contribution of pressure is omitted, indicating that the adsorption of molecules can stabilize this metastable phase. The mechanism observed for MUF-15 appears to be generic, although the details of the energetic balance involved in the opening process will depend on the microscopic nature of each structure: this means that while all members of the MUF-15 family are expected to have the potential to be flexible, flexing will be triggered under different conditions (temperature, pressure, adsorption) in different materials.

### 4.1.4 Conclusion

With DFT-based simulation techniques, I set to investigate the behavior of the MUF-15 framework and its -OMe derivative. The complexity of the -OMe structure made the geometry optimization impossible and forced me to study only the parent MUF-15 structure. Their structures are nonetheless quite close which makes it possible to transpose some of the results obtained with the parent structure. I then optimized the structure of MUF-15 at different pressures and noticed a strong anisotropy of this framework with the $b$ direction being bar far the softest. I also saw an abrupt change of cell parameters and Co–O bond distances at around -1.7 GPa indicating the breaking of a Co–O bond. This was confirmed by looking directly at the structure at -1.7 GPa. However, this phenomenon occurring very quickly, the constant-pressure calculations were not sensible enough to allow a close look at the opening mechanism. I therefore switched to constant-volume calculations, which confirmed the microscopic mechanism: First, as the volume increases, the linkers orient in the $b$ direction and flatten in the $ab$ plane. This is followed by cleavage of one of the Co–O coordination bonds aligned in the $b$ direction. Porosity analysis with the Zeo++ software package revealed an increase of the accessible volume and surface area with respect to the volume. This showed that the flexibility is coupled to the change of pore network and that the adsorption of guest molecules can trigger the opening of the framework. The energy penalty of such opening process was found to be reasonable compared to "breathing" energies of other MOFs. Studying the energy revealed a local minimum, indicating the existence of a metastable phase. This metastable phase corresponds to the open structure observed at high pressure.

The type of calculations performed in this work allows to investigate the flexibility of novel or unknown materials. It is a way of studying the effect of pressure on a structure without resorting to methods with a higher computational cost. While it cannot simulate the adsorption of gases, it mimics the process of adsorption by introducing a negative external pressure. Molecular dynamics simulation would have been preferred to simulate this process, however this was not possible due to the high computational cost. This work was published in Chemistry of Materials[248].

## 4.2 The role of the node-linker bond in the compression of UiO-66

### 4.2.1 Introduction

The second project I realised in collaboration with experimentalists was with Louis Redfern from Pr. Omar Farha's team in the university of Northwestern in Evanston (USA). They contacted our team because they wanted to add computational insights on their system.

They were working at the time on a widely-studied MOF called UiO-66. Its structure consists of 12-connected $Zr_6(\mu_3 - O)_4(\mu_3 - OH)_4$ nodes joined by linear 1,4-benzenedicarboxylic acid linkers (also called BDC) that assemble into the fcu topology. A representation of the structure is depicted in fig. 4.9. As it can be seen on the figure, the UiO-66 features two different cages within its structure: a large octahedral one (around 12 Å) and a smaller tetrahedral one (around 7.5 Å). UiO-66 has a remarkable chemical and thermal stability[54]. It also shows a higher resistance to mechanical stress compared to other common MOFs[250]. However, the presence of missing linkers defects can drastically impact the mechanical properties[251, 252]. Prior to collaborating with us, Pr. Farha's team studied the effect of the linker's length and determined that a systematic extension of the organic linkers leads to a decrease in the bulk modulus[253]. This is shown on 4.10. They also noticed that the void fraction was at play in the mechanical properties, as the same correlation could be observed there.

They noticed in the literature that, while UiO-66 has been extensively studied, there are fewer works focusing on changing the composition of the MOF[255, 256]. This motivated them to investigate the behavior of this MOF with different metal nodes and to study the role of the metal-carboxylate bond in the compression of UiO-66. They determined that the only variable in their experiment was the coordination bond. They synthesized a series of three MOFs M-UiO-66 (M = Zr, Hf, Ce) and investigated their behavior under pressure. They observed that, while Zr and Hf analogues were quite similar, Ce-UiO-66 is found to be much more compressible. They measured values of 37.9 GPa and 37.0 GPa for the bulk moduli of Zr and Hf respectively whereas Ce exhibited a value of 16.9 GPa. Having kept all the variables of the synthesis constant, they attributed this difference to the differences in the bond between the metal node and the linker.

They determined the bulk moduli values from a second-order Birch Murnaghan equation of state using their PXRD measurements under pressure. Their results are shown on fig. 4.11. It can be clearly seen that Ce-UiO-66 compresses much more readily than its two other analogues. Alternatively, they noticed a slight change of slope at around 0.4 GPa for Ce. They attributed this discontinuity to a possible structural change or a partial amorphization, but could not conclude on it since it

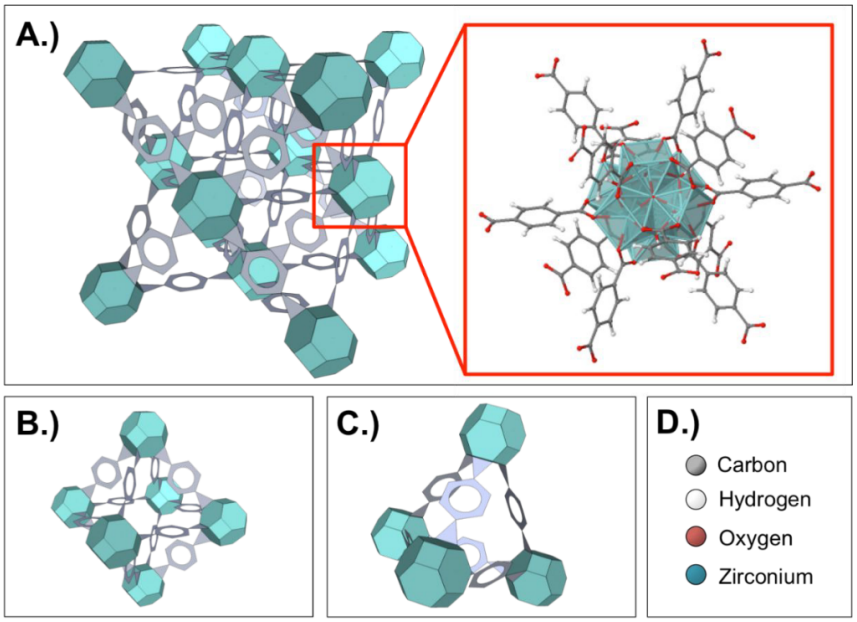


FIGURE 4.9: Representation of the UiO-66 structure. (a) The face-center-cubic unit cell of UiO-66. Metal nodes are represented in light blue and linkers are represented in grey. (b) Structure of the octahedral cage. (c) Structure of the tetrahedral cage. (d) Color scheme for the atomic representation. Taken from [254].

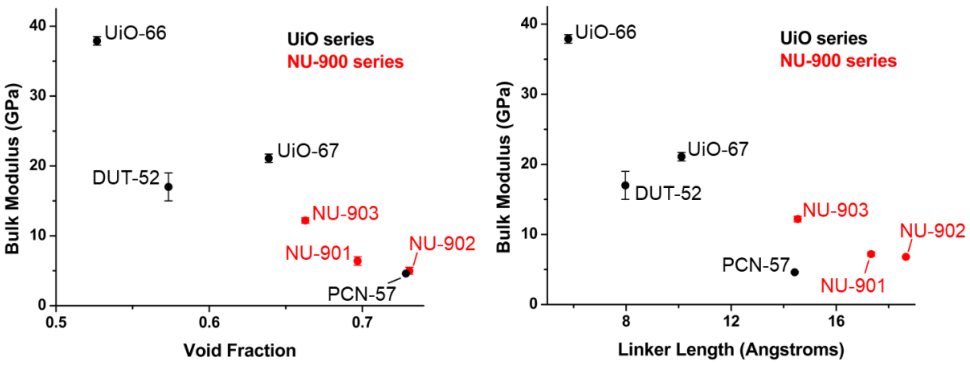


FIGURE 4.10: Effect of (left panel) void fraction and (right panel) linker length on the bulk modulus of two series of MOF (UiO and Nu-900). Taken from [253].

would require further characterization. Although it is not possible to probe directly the bond between the metal and the linker with Raman spectroscopy, pressure-dependent Raman measurements allowed them to study the carboxylate group of the linker. They drew a particular attention to the C–C stretch between the aromatic core of the linker and the carboxylate carbon and the O–C–O symmetric stretch of the carboxylate to provide insight into the metal-linker bond strength.

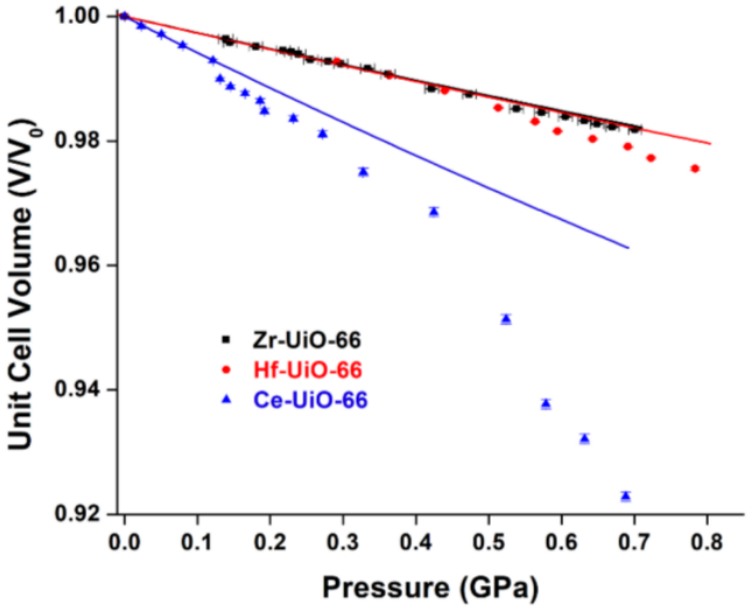


FIGURE 4.11: Unit cell compression of M-UiO-66 for M = Zr, Hf, Ce. Solid curves represent the second-order Birch-Murnaghan equation of state used to determine the bulk modulus of each MOF. Taken from [257].

They first observed that the peaks of interest broaden with increasing pressure, which is a common behavior known in high-pressure Raman spectroscopy[258]. In the case of Zr-UiO-66 and Hf-UiO-66, they observed an hypsochromic shift (increase of frequency) of both C–C and O–C–O vibrations over the range of 0 to 1.2 GPa. However, the spectrum of Ce-UiO-66 was much more complex to analyse quantitatively, as the immediate broadening of the two peaks leads to a single broad signal. This significant change at low pressure may result from structural changes, which is in agreement with changes observed at 0.4 GPa with PXRD measurements.

They then approached us to perform simulations on this system. They wanted to further understand why the Zr and Hf systems behave differently than the Ce one, based on a microscopic picture of the system.

### 4.2.2 Calculation parameters

I started my calculations on CIF files provided by Pr. Farha's team and which were obtained through powder XRD experiments. All the calculations were performed at the DFT level with the CRYSTAL14 software package[259, 260]. I tested several functionals with and without D3 dispersion corrections. As already observed earlier in this thesis, the PBEsol0 hybrid functional without dispersion corrections is the most efficient. Therefore I decided to use this functional again. All-electron localized basis sets were used for all the atoms except for hafnium and cerium for which I opted for a pseudopotential approach. The acronyms of the basis sets used for

each atoms are given in table 4.1 and can be found on the CRYSTAL online library (https://www.crystal.unito.it/basis-sets.php). I tested the convergence of *k*-point sampling for each material and found that only a few number of *k*-points was needed. Therefore I used a shrinking factor of 2, which ensures the convergence of the properties of interest.

| Atom | Basis set |
|---|---|
| Zr | Zr_all_electron_dovesi_unpub |
| Hf | Hf_ECP_Stevens_411d31G_munoz_2007 |
| Ce | Ce_ECP_Meyer_2009 |
| C | C_6-31d1G_gatti_1994 |
| O | O_6-31d1_gatti_1994 |
| H | H_3-1p1G_gatti_1994 |

TABLE 4.1: List of the acronyms of the basis sets used for the simulations.

Stricter convergence criteria than the defaults proposed by the CRYSTAL14 code were used for geometry optimization (a maximum of 0.0005 au on atomic displacements during one optimization step and 0.0001 au on forces) because the aim was to study very fine variations of structure and vibrations. Representative input files for the calculations are available online at https://github.com/fxcoudert/citable-data.

### 4.2.3 Results

I started by optimising the geometries of all three materials. I made full use of symmetry and relaxed both the atomic positions and the cell parameters. Due to the high symmetry of these systems, the minimization was relatively quick and I was able to obtain the equilibrium structures within a short time. Comparing the Zr and Hf materials, I saw that the results of the optimisation and the calculated properties (that will be shown later) were completely similar between the 2 metals, as seen experimentally. Therefore, for the sake of clarity, I will sometimes only use one of the Hf and Zr compounds for the comparison with Ce-UiO-66. The reader can note, however, that all the trends observed for Zr apply to the Hf counterpart and vice versa. The cell parameters obtained are available on table 4.2. A good agreement can be observed between my calculations and the measured experimental cell parameters. Although a small discrepancy can be seen, it is completely reasonable and fits the accuracy of the parameters I chose. Following the optimization of the structures, I conducted frequency calculations in order to obtain the frequencies of vibrations of the C–C stretch and the O–C–O symmetric stretch. Again I found a good agreement with the experiment. For example, Zr C–C stretch was found to be around 1430 $cm^{-1}$ and I obtained a value of 1428 $cm^{-1}$, which is very close. The same goes for the O–C–O symmetric stretch for which I found a value of 1446 $cm^{-1}$, very close to the value of 1451 $cm^{-1}$ measured experimentally. The same agreement was observed for the two other compounds. Overall my simulations reproduced quite well the experimental findings and this comforted me in the parameters I chose.

The frequency calculations I performed also revealed an interesting point: the presence of a negative frequency in Ce-UiO-66. Negative frequencies usually denotes the instability of the structure of a material or molecule. This instability can be investigated by looking at the vibration mode related to the negative frequency. Here, the

| Hf-UiO-66 | | Ce-UiO-66 | |
|---|---|---|---|
| Expt. | Calc. | Expt. | Calc. |
| 20.7279 | 20.6268 | 21.488 | 21.3944 |

TABLE 4.2: Cell parameters comparison between the experiments[257] and the simulations. Values are given in Å.

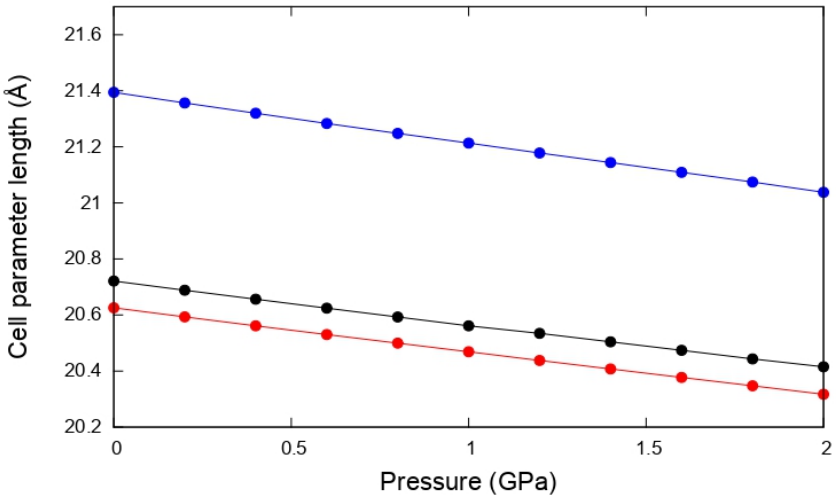


FIGURE 4.12: Evolution of the cell parameters of M-UiO-66 (M = Zr, Hf or Ce) with respect to the pressure.

instability observed is linked to the ligand bowing out of its average plane. These modes, also called "guitar string" modes, have previously been observed in UiO compounds and other MOFs[261]. They originate from the nature itself of DFT calculations: indeed DFT is conducted at 0 K. This means that low frequency vibrations are not activated energetically. Vibrations such as the ligand bowing out of its plane require low temperature to be active. At 0 K, the minimal-energy configuration for the ligand is out of plane, in one of the two possible symmetry-equivalent configurations. However, at finite temperature, thermal motion would enable the ligand to move back and forth resulting then in the averaged in-plane position, which is the high-symmetry structure observed in XRD. Therefore, the imaginary frequency obtained here is completely due to the nature of DFT calculations and does not imply any instability of the structure. Knowing this, and because I am not interested in the details of the "guitar string" effect, I decided to ignore this phenomenon and pursue the study.

I then conducted constant pressure calculations to investigate the evolution of the structure under pressure. I adopted a pressure range from 0 to 2 GPa. The evolution of the unique cell parameter is shown in fig. 4.12. As it can be seen, the cell parameter decreases linearly with the pressure in all three cases. Zr and Hf have a similar slope, which means that their compressibility is very close. Ce on the other hand shows a higher slope meaning that the Ce-based MOF is more compressible than the two others. However, the difference is not as marked as in the experiment. This is why I expected to find a less important difference between the bulk moduli values of Ce and its counterparts compared to the experiments. It can also be noticed on fig. 4.12 that the discontinuity observed experimentally on the volume of the Ce

compound at 0.4 GPa is not reproduced here. This discontinuity was attributed to structural changes occurring at this pressure. If the simulations do not reproduce this behavior, then it means that this phase transition involves a symmetry-breaking transition to a lower symmetry phase. This type of process cannot be reproduced by the calculations I am running. Simulating this kind of behavior would need a technique that allows symmetry-breaking processes. Molecular dynamics could be a method of choice for this, but it was not considered since reproducing the transition was not the point of the present study.

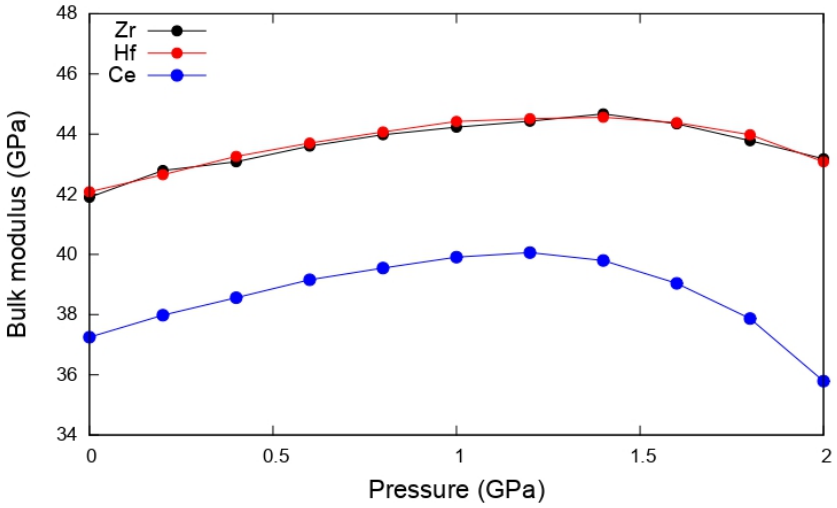


FIGURE 4.13: Evolution of the bulk modulus of M-UiO-66 (M = Zr, Hf or Ce) under pressure.

I also conducted linear response calculations in order to determine the stiffness tensor of each structure. This allowed me to derive the bulk modulus for each structure. From these calculations, I obtained a similar value of around 42 GPa for Zr and Hf compounds while Ce-UiO-66 showed a value of around 37 GPa. This difference, while not being as marked as with the experiments, correlates with the difference of compression measured experimentally and the evolution of the cell parameter observed previously: Ce-UiO-66 is more compressible than its Zr and Hf analogues. I confirmed this trend through the whole pressure range studied by determining the bulk moduli values for each pressure. These are displayed on fig. 4.13. This figure also confirmed that Ce-UiO-66 and its counterparts are really close in terms of behavior and differ mainly in terms of amplitude. Indeed we can see that the bulk modulus of all three materials undergoes the same variations albeit at a lower value for Ce.

Simulating structures under pressure also allowed me to probe the variations of specific structural features. Among all of them, I especially focused on structural features around the ligand. First, I observed that O–C–O angles and dihedral angles between ligands were identical in all cases, shifting by around 1° in each case. The different bonds in the materials are decreasing which is a logical result of the compression. However, adopting the cell parameter as a reference leads to interesting new conclusions. I represented $(l/l_0)/(a/a_0)$ curves for C–C, C–O and M–O bonds with respect to the pressure ($l$ and $a$ correspond respectively to the bond length and lattice parameter at the associated pressure and $l_0$ and $a_0$ are the zero pressure values). This can be seen on fig. 4.14. I omitted the Hf compound in the figure for clarity as the results were similar to the Zr analogue. This representation allows

to compare the variation of the bond distances with the deformation of the cell. It appears that the C–O bond decreases more slowly than the cell parameter in both cases $[(l/l_0)/(a/a_0) > 1]$ whereas the C–C bond decreases almost at the same rate $[(l/l_0)/(a/a_0) \approx 1]$. Moreover, it can be seen that the variation of the C–O bond is even slower in the case of Ce-UiO-66, which supports the slower variation of the O–C–O symmetric stretch for Ce. This shows that the carboxylate group is a key feature when investigating the compression of this material. The last bond I focused on was the M–O bond as it is the link between the metal node and the organic linker. The plot revealed that this bond decreases faster than the lattice parameter. It is even more pronounced in the case of Ce, which I ascribed to the orbitals of Ce being more diffuse than those of Zr, leading then to a slightly higher flexibility of the Ce–carboxylate bond. All these calculations were static (realised at 0 K) but I believe that, due to the relatively low thermal expansion of UiO-66[262], the observations made here are also valid at room temperature.

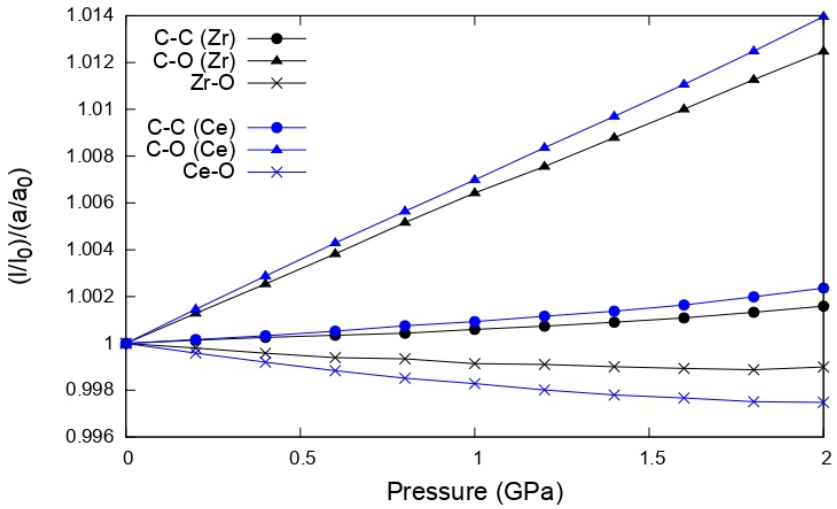


FIGURE 4.14: $(l/l_0)/(a/a_0)$ plot for C–C, C–O and M–O bonds of Zr-UiO-66 and Ce-UiO-66. Black lines and blue lines correspond respectively to Zr and Ce. Circles, triangles and crosses correspond respectively to the C–C, C–O and M–O bonds. $(l/l_0)/(a/a_0)$ is dimensionless.

During our collaboration, we also discussed the possibility of $Ce^{3+}$ forming in the framework, as $Ce^{4+}$ has a high potential of reduction. These $Ce^{3+}$ cations could incorporate into the structure accompanied by an additional proton on one of the $\mu_3$-O groups present in the $Ce_6$ node. This motivated Pr. Fahra's group to conduct XPS experiments, which revealed that the synthesised material contains about 10 % of $Ce^{3+}$ ions. This ratio corresponds roughly to 47 % of $Ce_6$ nodes containing at least one $Ce^{3+}$, assuming they are evenly distributed throughout the MOF. The presence of $Ce^{3+}$ within the framework may lead to distortions of the node structure, as $Ce^{3+}$ has a longer ionic radius then $Ce^{4+}$. This would influence the mechanical properties of the MOF. The existence of $Ce^{3+}$ species inside this material correlates with the observation made so far and explains the differences between experiments and simulations. Indeed the simulations were realised on a perfect structure containing 100 % of $Ce^{4+}$. I think that this is a further explanation of why the bulk modulus we calculated for Ce-UiO-66 is not as low as measured experimentally. Moreover, $Zr^{3+}$ and $Hf^{3+}$ are not likely to form which, considering how close our results are from the experiments for Zr-UiO-66 and Hf-UiO-66, supports the idea that $M^{3+}$ ions play

a role in the properties observed. It has been hypothesized then that the presence and proportion of $Ce^{3+}$ can influence the unique behavior observed in Ce-UiO-66.

### 4.2.4 Conclusion

With DFT-based methods, I investigated the different behaviors shown by the Zr-UiO-66 and Hf-UiO-66 compared to their analogue Ce-UiO-66. The optimisation of geometry succeeded rather smoothly and within reasonable time. A good agreement with the experiments was observed for both the lattice parameters ($\sim$1 Å of difference) and the frequencies of vibrations (less than 10 $cm^{-1}$ of difference). The frequency calculations revealed however a negative frequency in the case of Ce-UiO-66, which is usually an indicator of instability. This vibration mode appeared to be linked to the linker bowing out of its average plane. However, this negative frequency was not correlated to a potential instability: it was found to be linked to the particular nature of DFT being conducted at 0 K. This "guitar string" effect was of no interest in my case and did not constitute an instability of the material. Therefore, I continued the study without taking into account this phenomenon.

Constant pressure calculations allowed me to highlight the close behaviors between the three materials, albeit at different values for Ce. Lattice parameters were found to decrease linearly with the pressure and faster in the case of Ce which is in line with the lower bulk modulus measured experimentally. However the phase transition observed experimentally could not be reproduced here with static DFT calculations. Calculating the bulk modulus led to values of around 42 GPa for Zr and Hf and 37 GPa for Ce. While the difference is not as marked as in the experiments, it is still confirming the different behaviors between Ce and its two counterparts. The evolution of specific structural features could be probed using the constant pressure simulations. C–O bonds were found to decrease more slowly than the lattice parameter. The variation of the C–O bond is even slower in the case of Ce, which supports the slower variation of the O–C–O symmetric stretch for Ce. After some discussion with our team, Fahra's team decided to measure the possible presence of $Ce^{3+}$ in Ce-UiO-66. The presence of $Ce^{3+}$ would change the node structure by inducing some distortions and would impact the mechanical properties. With XPS measurements, they found that the material contains $\sim$ 10 % of $Ce^{3+}$. This is most certainly the reason why I obtained different values in my calculations of Ce-UiO-66 as my simulations were realised with a perfect structure containing 100 % of $Ce^{4+}$. It shows however that the presence and proportion of $Ce^{3+}$ could influence the behavior of this framework.

We discussed the possibility of performing simulations taking into account the percentage of reduced Ce ions. It would be a nice way of proving the hypothesis we formulated. To create $Ce^{3+}$ cations, it was necessary to remove some linkers in the structure. Even though we selected carefully the linkers to remove in order to keep the highest symmetry possible, the decrease of symmetry complicated a lot the convergence of the calculations. In the end, the calculations became too complex and could not succeed.

This work was published in Chemistry of Materials[257].

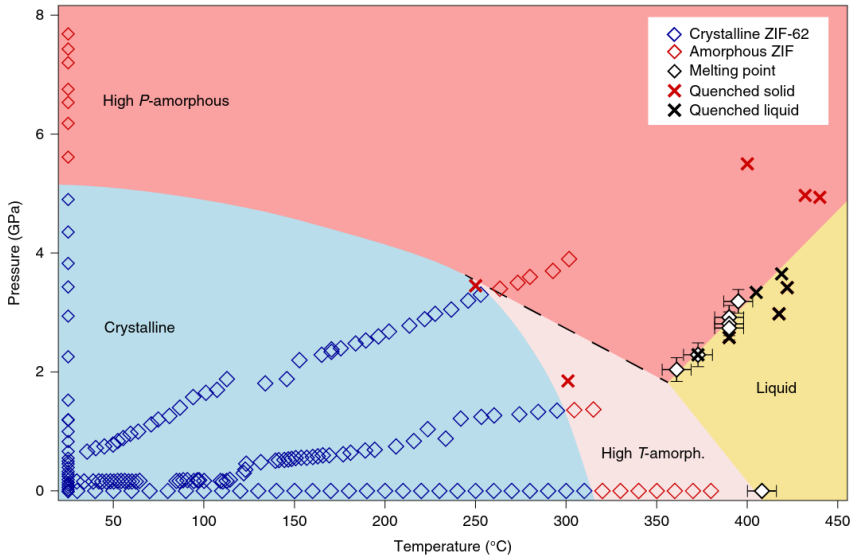


FIGURE 4.15: Experimentally derived P-T phase diagram for ZIF-62. Taken from [265].

## 4.3 Melting of hybrid organic-inorganic pervoskites

### 4.3.1 Introduction

Melting is a process often used in industry as it is a convenient way of shaping a material and is also necessary to create glasses. However, the melting of MOFs is a complicated topic of research. Indeed, MOFs tend to decompose before melting due to their rather low thermal stability compared to inorganic solids. Their instability is the principal obstacle for industrial applications. Nevertheless, recent efforts were made in the group to obtain and study MOFs as glasses or liquids in collaboration with experimentalists, especially the group of Pr. Thomas Bennett from the university of Cambridge. In 2017, Gaillac et al[263] studied a zeolitic imidazolate framework (ZIF) called ZIF-4 that experimentalists were able to melt. They demonstrated the retention of the chemical configuration, coordinative bonding and porosity. With their results, they introduced the new term "MOF liquid". Later, Gaillac, along with Pullumbi and Coudert[264], investigated the transferability of the mechanism they found previously to other ZIFs. They showed that the melting phenomenon is similar for ZIF-8 and ZIF-zni. They also found that ZIF-8 cannot melt as they observed a temperature of fusion above its temperature of decomposition. They suggested that the high porosity of ZIF-8 is the origin of this phenomenon. Thus, they hypothesized that denser materials such as ZIF-4 and ZIF-62 are more prone to melt. They further studied these systems by establishing the pressure-temperature phase diagram of ZIF-4 and ZIF-62 (shown in fig. 4.15 for ZIF-62) through combined in situ powder XRD and optical microscopy with thermodynamic considerations[265]. They proved that increasing pressure induces the lowering of the melting temperatures. This is an important discovery for industrial processes as it makes it easier to produce MOF glasses since the pressure needed is within the industrially achievable range.

To this day, our group is still closely collaborating with experimentalists on this topic of research. During my PhD, I also had the opportunity to tackle this theme

with a final collaboration (in the order presented in this thesis). This last collaboration was with Pr. Thomas Bennett's team. They were particularly interested to find new MOFs that can melt. Among the different possible MOFs, they recently focused on hybrid organic-inorganic perovskite (HOIP). Perovskites are materials that have a general formula of the form $ABX_3$ with A and B being ions (often of different sizes) and X another ion (often oxide) bonding with A and B. Examples include the well-known $CaTiO_3$ or $BaTiO_3$. Following the same rule, HOIP are materials with the $ABX_3$ structure. The difference with classical perovskites is that A is an organic cation, B a metal ion and X a bridging ligand. These types of material are used in ionic transport, luminescent or ferroelectric applications[266, 267, 268]. Examples of these are the hybrid lead halide perovskites, which are used in photovoltaic devices[269, 270]. Although the melting of MOFs presents advantages such as the possibility to shape the material easily and post-synthesis, the field of MOF science remains dominated by the crystalline state. A few efforts have nevertheless been made towards the controlled amorphization and the formation of glasses. Indeed the reversible pressure-induced amorphization of methylammonium lead halide perovskites was reported in the litterature as well as the melting of layered HOIP $(C_4H_9NH_3)_2MI_4$ (M = Ge, Sn, Pb)[271, 272, 273].

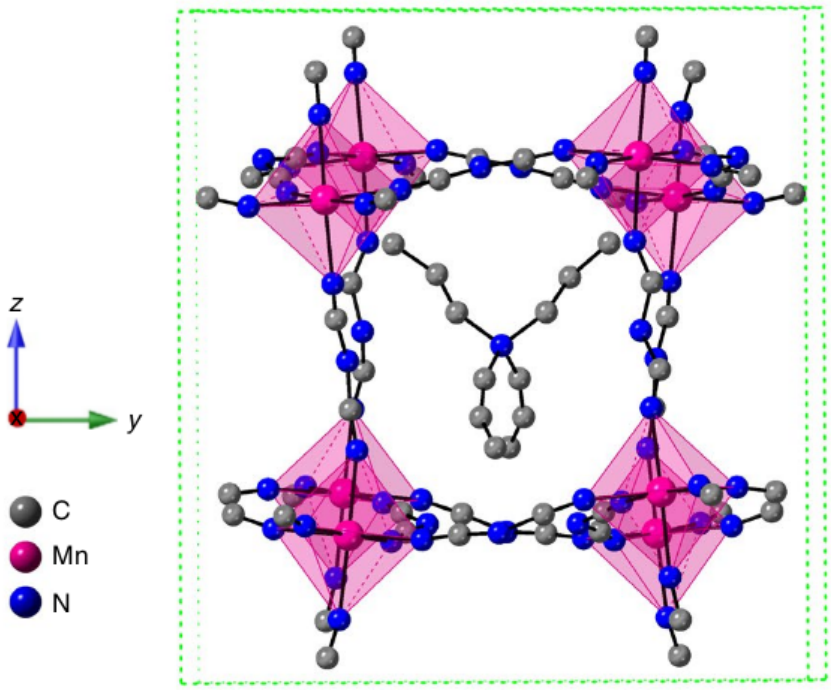


FIGURE 4.16: Unit cell representation of the $[TPrA][Mn(dca)_3]$. Hydrogen atoms are not shown. Taken from [274].

In the case of Pr. Bennett's team, they were interested in three particular hybrid perovskites whose structure is depicted on fig. 4.16. They all show the same structure and only differ with the nature of the metal ion. The general formula of the material is $[TPrA][M(dca)_3]$ with TPrA being the abbreviation of tetrapropylammonium $(CH_3CH_2CH_2)_4N^+$ and M = $Mn^{2+}$, $Fe^{2+}$, $Co^{2+}$. The Bennett group carried out different experimental measurements in order to find evidence of melting. They reported

the thermogravimetric analysis (TGA) of their solids which showed the decomposition at 281, 273 and 267°C for respectively Mn, Fe, and Co materials. This is comparable to previous observed temperature of decomposition for similar frameworks [275]. They also conducted differential scanning calorimetry (DSC) measurements (which are shown in fig. 4.17) along with optical images of [TPrA][Mn(dca)$_3$] before heating and after cooling. They identified strong endothermic peaks which they attributed to the melting of the frameworks. Melting temperatures were found to be of 271, 263 and 230°C for Mn, Fe, and Co respectively. This was found to follow the trend of ionic radii, as Co has the smallest ionic radii and Mn has the largest one. They found that these results follow the soft-hard acid-base theory, which says that a soft metal centre such as $Mn^{2+}$ would interact more favourably with the dca ligand leading then to a higher melting temperature.

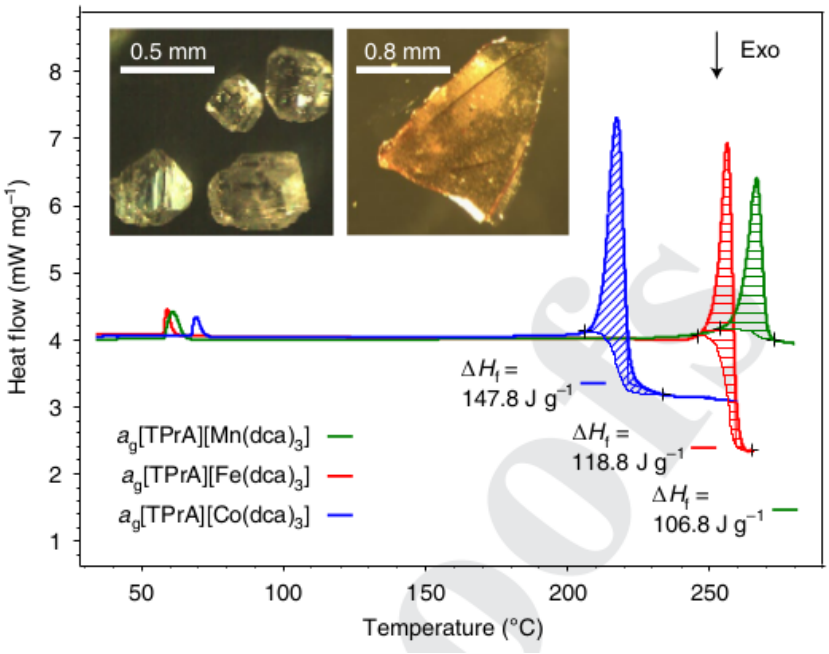


FIGURE 4.17: Differential scanning calorimetry measurements on [TPrA][M(dca)$_3$] (M = Mn, Fe, Co) along with optical images of [TPrA][Mn(dca)$_3$] before heating (left) and after cooling (right). Taken from [274].

They realised combined DSC-TGA experiment and witnessed a clear liquid-like behavior when heating above the melting temperature and cooling back to room temperature. This liquid behavior was also confirmed through XRD measurements. They determined the glass transition temperature ($T_g$) of all their samples through DSC and completed with additional methods to confirm the results. DSC led to $T_g$ values of 223, 225 and 125°C for Mn, Fe and Co respectively. Thermomechanical analysis (TMA) and frequency-dependent dynamic mechanical analysis (DMA) resulted in very close values compared to DSC. This confirmed the values obtained for the glass transition temperature. Other evidence of melting was found in the $^{13}$C NMR where the peak corresponding to the N–C–N carbon was broad in the melt-quenched material, whereas it was clearly defined in the crystalline material. A new peak was also observed at around 120 ppm and attributed to a free dca ligand, which highligths the decoordination of dca. Finally, pair distribution functions depicted in fig. 4.18 showed that the peaks at high *r* broaden and weaken after melt-quenching,

evidencing the loss of long-range order. Decoordination of dca ligand was also observed, as several peaks related to the dca split or broaden.

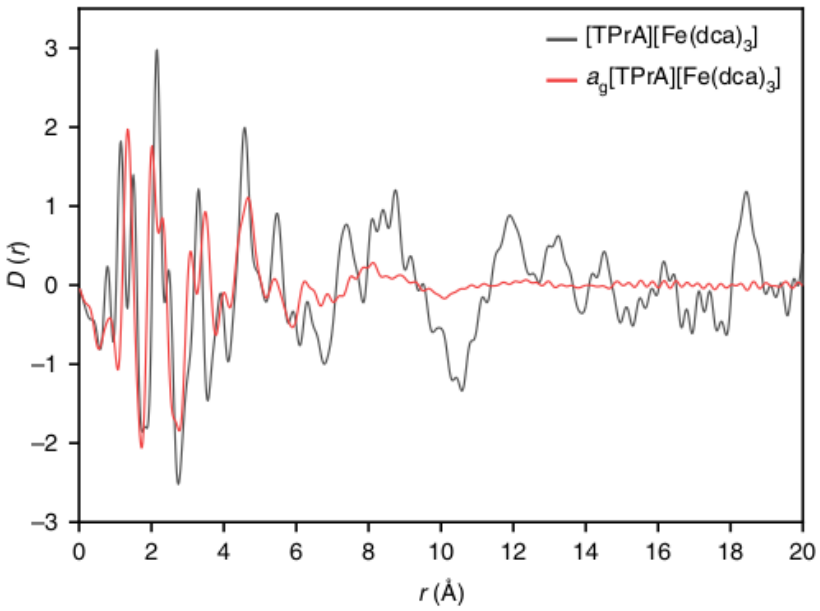


FIGURE 4.18: Partial distribution functions of [TPrA][Fe(dca)$_3$] and its glass analogue noted a$_g$[TPrA][Fe(dca)$_3$]. Taken from [274].

With their work, Pr. Bennett's team succeeded in proving the melting of their materials by bringing evidence of liquid-like behavior and measuring melting temperatures. However, they wanted to have better atomistic insight in the process of melting and asked us if we could perform simulations on their systems.

### 4.3.2 Calculation parameters

Due to the thermal nature of the behavior I am studying, I used ab initio MD to investigate the dynamics of this system. The hybrid Gaussian and plane wave method GPW as implemented in CP2K were used[276]. The simulations were performed in the constant-volume (N, V, T) ensemble with fixed size and shape of the unit cell. A time step of 0.5 fs was used in the MD runs, and the temperature was controlled by velocity rescaling[99]. The exchange-correlation energy was evaluated in the PBE approximation[277] and the dispersion interactions were treated at the DFT-D3 level[97]. The Quickstep module uses a multigrid system to map the basis functions onto. The default number of 4 different grids was used, along with a plane-wave cutoff for the electronic density of 600 Ry, and a relative cutoff of 40 Ry. Valence electrons were described by double-$\zeta$ valence polarized basis sets and norm-conserving Goedecker-Teter-Hutter pseudopotentials[278] all adapted for PBE (DZVP-GTH-PBE) for H, C and N or optimized for solids (DZV-GTH-PADE) in the case of Mn.

### 4.3.3 Results

Pr Bennett's team worked on three members of the [TPrA][M(dca)$_3$] materials family. However, it was not possible for me to study all of them as it would have required a lot more computational resources than what our group had acces to. Indeed, the melting process cannot be investigated with static DFT methods that were

used in the previous collaborations. Such process involving bond-breaking phenomena also prevents the use of classical force-fields. This problem must then be addressed with first principles molecular dynamics which is much more demanding in terms of CPU resources. This is why I chose to study only one of these compound. I therefore conducted my study only on the Mn compound. However, given that these materials have a completely similar structure and only differ in the nature of the metal, I believe that the general results observed here (in particular the mechanism) are transposable to [TPrA][Fe(dca)$_3$] and [TPrA][Co(dca)$_3$]. As the process I aimed to study was too slow, it would require a very long timescale to observe it, which is not something possible. Therefore, I chose to simulate the systems at temperatures far above the melting temperature in order to have a timescale of the order of the picosecond. This is a common trick to simulate processes that are too slow.

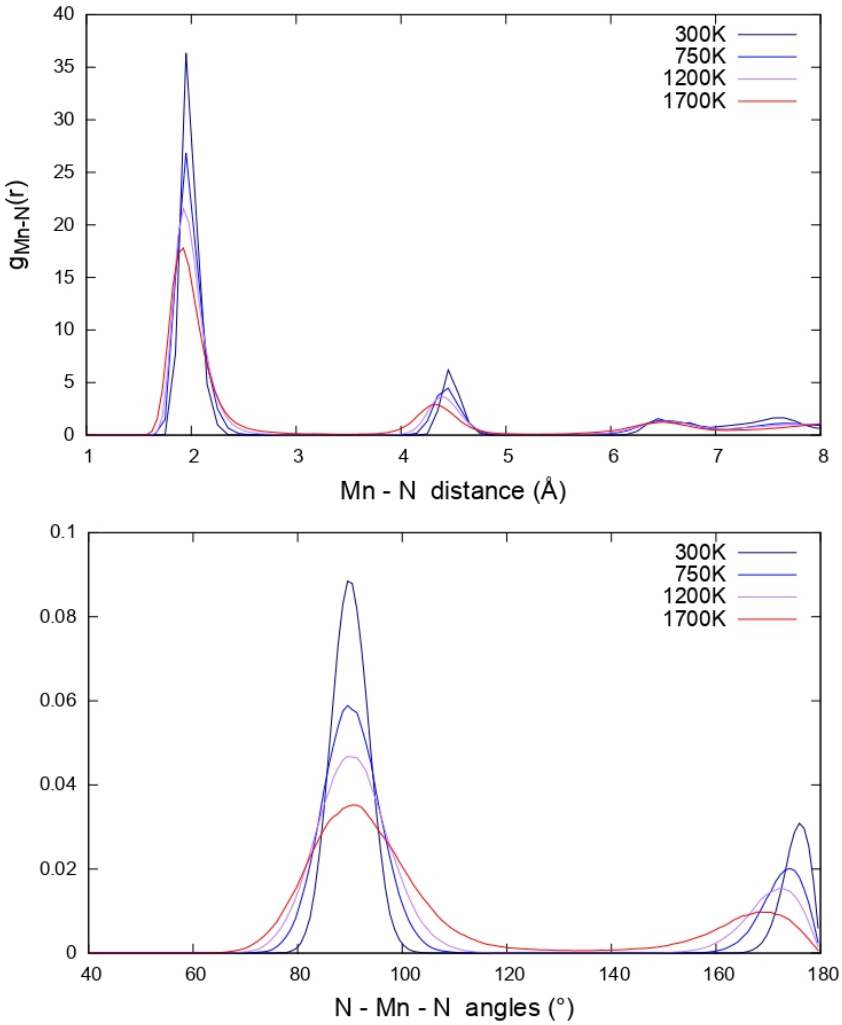


FIGURE 4.19: (left) Pair distribution function of Mn–N distances at 300, 750, 1200 and 1700 K and (right) distribution of N–Mn-N angles at 300, 750, 1200 and 1700 K.

I realised 4 simulations at constant temperatures of 300, 750, 1200 and 1700 K. The results displayed in this part were obtained from analyses of the trajectories of these simulations. I first started by extracting pair distribution functions (PDF). Mn–N

PDFs are represented on fig. 4.19. It can be seen that the peaks broaden with increasing temperatures, meaning that Mn–N distances are less well defined as the temperature grows. It can also be observed that the long-range order (above 6 Å), while not being well defined at 300 K, is completely lost at high temperatures. I calculated the generalized Lindemann ratio for the first two peaks of the Mn–N pair distribution functions. This ratio was defined by Lindemann in 1910 as a threshold criterion for the identification of melting behaviors. The values are reported in table 4.3. I found that the ratio calculated exceed the Lindemann ratio between 750 and 1200 K[279], indicating a liquid-like behavior at these temperature. The thermal fluctuations are also noticeable on the N–Mn–N bond angles that are represented also in fig. 4.19. I observed that the peak at 90° broadens with increasing temperature, translating the distortion of the perfect octahedral environment. This distortion can also be seen at 180° where the peak broadens in the same fashion.

| Temperature (K) | First peak (%) | Second peak (%) |
|---|---|---|
| 300 | 7.8 | 6.5 |
| 750 | 11.1 | 10.1 |
| 1200 | 13.9 | 13.4 |
| 1700 | 17.2 | 17.2 |

TABLE 4.3: Lindemann ratio of the first and second peak of the Mn–N pair distribution functions at the different temperatures studied.

From the simulated trajectories I was also able to extract the mean square displacement (MSD) of the atoms present in the structure. The total MSD is represented in fig. 4.20 for all the temperatures considered. This MSD shows that no large-scale motion of atoms is present at 300 K as the curve stays almost flat. The small variations of the order of 1 Å are due to framework vibrations. However, starting at 750 K, the MSD shows a monotonic increase with the simulation time, hinting at a linear behavior (although we only have limited statistics). This is indicative of a diffuse behavior evidencing the melting of the material. This diffuse behavior is also present at 1200 K and even stronger at 1700 K. Among the atoms of the structure, the Mn shows almost no motions. Most of the motions observed come from the TPrA cation which seems to need rather small thermal activation to move in the structure. The other contribution to the total MSD at high temperature comes from the dca ligand. Examining the trajectories revealed the breaking and reformation of Mn–N bonds, explaining the contribution of the dca linker in the total MSD. I also confirmed that the TPrA cation shows significant movement inside its cage.

The breaking and reformation of Mn–N bonds led me to calculate the free energy profiles based on the Mn–N PDFs. They are represented on fig. 4.21. These potentials of mean force were calculated as $F(r) = -kT \ln g(r)$. It can be seen on the figure that the breaking of Mn–N bonds is not happening at 300 and 750 K during our limited simulation time as the barrier shows discontinuities at these temperatures. However, this barrier appears at 1200 and 1700 K showing that the breaking and reformation of Mn–N bonds can happen at these temperatures. I estimated an energy barrier of 85 kJ.mol$^{-1}$ at 1200 K. This high energy barrier confirms that, as in conventional solids, melting occurs through rare barrier-crossing event. It can be noted that this energy decreases with increasing temperature, although it decreases slower than what would be expected. As a result, the process involved here is less impacted by temperature compared to porous MOFs. I confirmed the high energy

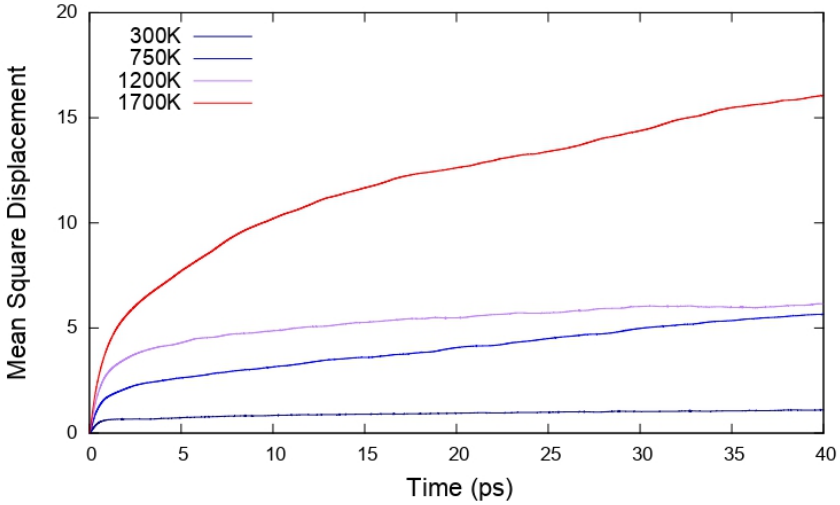


FIGURE 4.20: Total mean square displacement of [TPrA][Mn(dca)$_3$] at 300, 750, 1200 and 1700 K.

barrier with another approach considering defects. I defined the five-coordinated Mn ions as defects and calculated their enthalpy of formation through the probability equation: $n_d \propto exp(-\Delta_f H_d/RT)$. The distribution of coordination state of Mn atoms at different temperatures is represented in fig. 4.22. Five-coordinated M ions are only present at 1200 and 1700 K. I then calculated the enthalpy of formation between these two temperatures and obtained a value of $\Delta_f H_d \approx 91$ kJ.mol$^{-1}$. This value is very close to the one obtained through the potentials of mean force, confirming the process of melting as well as the energy barrier found earlier.

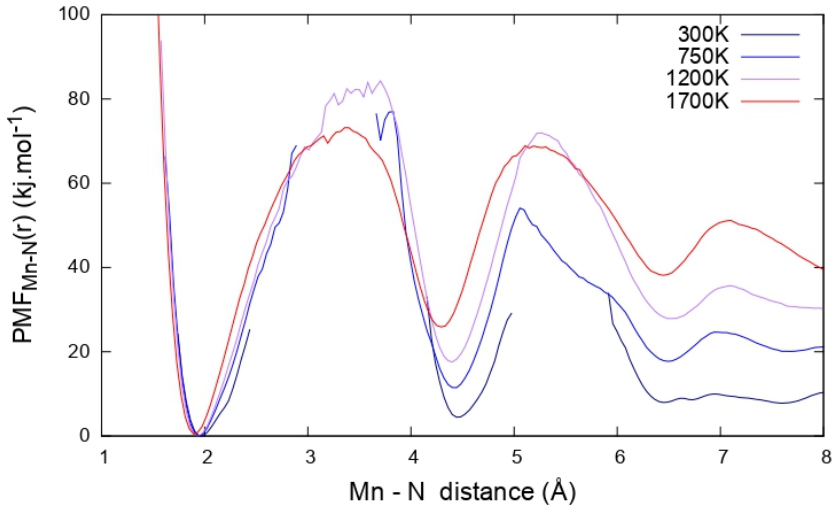


FIGURE 4.21: Potential of mean force along the Mn–N distance coordinate at 300, 750, 1200 and 1700 K. The discontinuities observed arise from the null values of the Mn–N PDFs.

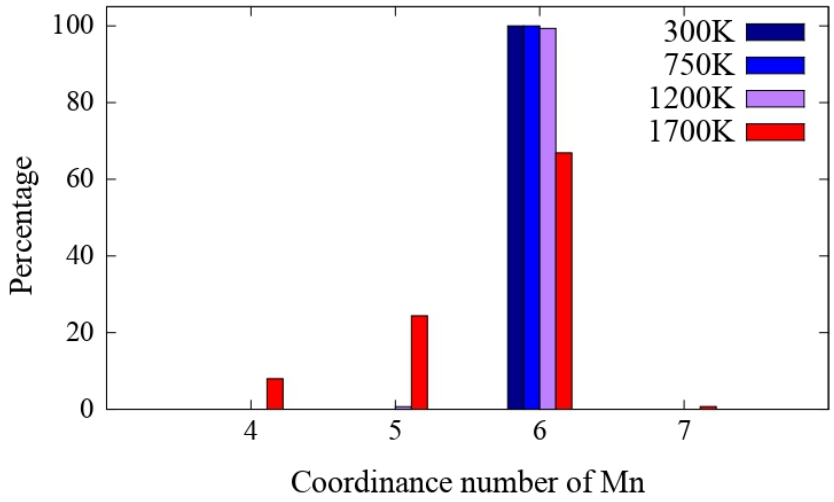


FIGURE 4.22: Percentage of coordinance number of Mn metal centres in all four AIMD trajectories.

### 4.3.4 Conclusion

During my collaboration with Pr. Bennett's team, I investigated the melting process of [TPrA][M(dca)$_3$] (M= Mn, Fe, Co). The nature of the process prevented me from using static DFT-based calculations as well as classical force fields. I therefore used first principles molecular dynamics in order to study this phenomenon. However, given the high computational cost of such method, I only chose to study the [TPrA][Mn(dca)$_3$]. I conducted molecular dynamic simulations in the NVT ensemble at temperatures of 300, 750, 1200, 1700 K. I used the resulting trajectories for the analysis of melting. I extracted the Mn–N pair distribution functions and observed that the peaks broaden and that the long range order fades with increasing temperature. The generalized Lindemann ratio calculated on the first two peaks of Mn–N PDFs was found to exceed the Lindemann ratio between 750 and 1200 K, indicating the melting of the original crystal structure. Distortions of the perfect octahedral environment was also observed as the temperature grows.

The total mean square displacement showed no motion of atoms at 300 K but revealed a diffuse behavior at higher temperatures. I found that TPrA and dca were the main contribution to those motions while Mn atoms were almost not moving. The MSD along with the exceeded lindemann ratio proved that the melting is occurring at 1200 and 1700 K. Examining the trajectories revealed the breaking and reformation of Mn–N bonds throughout the simulations. In order to obtain the energy barrier for such event to occur, I calculated the free energy profiles of the Mn–N PDFs. This resulted in a energy barrier of 85 kJ.mol$^{-1}$. I confirmed this value by calculating the enthalpy of formation of five-coordinated Mn ions considering them as defects. I obtained a value of 91 kJ.mol$^{-1}$, validating the value measured with the potentials of mean force.

This work was published in Nature Chemistry[274].

# Conclusion

In this thesis, I investigated the thermal behavior of a large number of pure silica zeolites through DFT calculations. The systematic study confirmed some hypotheses such as the recurrent presence of negative thermal expansion in zeolites and the fact that this property is driven by low vibration modes with large negative Grüneisen parameter. This work also created an operable database of DFT-optimised structures with their corresponding calculated properties. It has been then further exploited with the help of machine learning methods. I showed that my database was suitable for the prediction of thermal expansion. I used different types of descriptors to compare their performance and found that environment-describing methods such as SOAP were the most efficient. Combining the different types of descriptors, I successfully predicted the bulk modulus and thermal expansion of the PCOD2 database, which contains more than 600,000 structures.

I also worked in collaborations with experimentalists on a large variety of topics. I studied the opening process of MUF-15 and its -OMe analogue. While the calculations could not converge with the -OMe group, the mechanism was resolved for the parent structure. The linkers orient in the softer direction ($b$) and flatten in the $ab$ plane. This is then followed by the cleavage of one of the Co–O coordination bond. I also studied the properties of the Zr, Hf, and Ce analogues of UiO-66 and highlighted the differences observed experimentally with Ce. The discrepancy of values between the simulations and the experiments was linked to the presence of $Ce^{3+}$, which we could not take into account in the calculations. I finally investigated the melting mechanism of a Mn-based hybrid perovskite through molecular dynamics. It led to the conclusion that melting was occurring through rare barrier-crossing events, a process similar to conventional solids. All of these studies highlight the importance of combination between theory and experiments. When performed under the right conditions, experiments inform us about the real behaviors of materials. However, it is sometimes complex to give the origins of such behaviors as some informations are not accessible through experimental measurements. In those cases, computational methods can often offer another perspective. While simulations can only hope to reproduce the results observed experimentally, they can however probe effects that cannot be measured. Computational methods are then useful for adding insights into a mechanism or a phenomenon observed experimentally. The combination of theory and experiments is thus necessary for the establishment of new knowledge as they complete each other.

This work is in line with the research topics of the group. Indeed, the team is focused on two main topics: the first one is the adsorption of fluids in porous materials and the behavior of nano-confined phases. The second one is the study of behaviors and properties of MOFs and *soft porous crystals*. Through this work, I investigated the unusual thermal behavior of zeolites using simulation techniques commonly used in the group. I set up a methodology to conduct a large systematic study, which deepened the understanding of negative thermal expansion and its occurrence in

zeolites. Along with the previous mechanical studies and the previous methodologies of simulation developed in the group, the work shared in this thesis represents another step towards the multi-scale description of the complete behavior of *soft porous crystals*.

This thesis opens a number of questions and possible developments. First, the question of the effect of composition on the properties of zeolites. Indeed, here I studied the effect of topology by keeping the composition of all the structures to the exact same composition $SiO_2$. However, most synthesized zeolites are not pure silica and are often composed of other atoms such as aluminium. The question that can be asked here is: are the results found for pure silica zeolites transposable to isostructural zeolites with different compositions? Or in other words, are the properties calculated in this work only driven by the topology of the material, or is the composition playing a role too? The answer would come from the study of materials with different compositions and same topology. This could be done on a small number of frameworks, allowing to consider a large number of compositions and to perform a thorough analysis for each case. A systematic study could also be considered, although including the same number of frameworks compared to this work would result in an even higher computational cost. Indeed the same methodology used in this thesis would have to be performed on each different compositions of each structures. This limits the number of compositions that can be included, but would allow for systematic effects of composition to be investigated.

The systematic study of pure silica zeolites created a database of high accuracy optimized structures and their corresponding properties. Such database adds to the already existing resources for zeolites like the IZA database or the PCOD2 database of hypothetical structures. It is a tool that can complete the other databases or can be used for comparison. But it can also be exploited alone for various applications. One of the largest limitation to systematic studies is the computational cost. The only way to calculate more and more frameworks at the same time is by sacrificing the accuracy, proportionally to the number of entries to calculate. Force field (FF) methods are well indicated for this. FFs are most of the time developed and optimised with a high number of calculations on different materials. They aim to reproduce the same accuracy on which their development is based, provided that they are used on the same family of materials used for their optimization. Some FFs are oriented towards more general and thus less accurate applications while others are much more specific and accurate for a limited range of materials. The development of FFs requiring a lot of high accuracy calculations, the database calculated in this work is suitable for the creation of a high accuracy FF for zeolites. This can allow for larger systematic studies on this family of materials without the need of sacrificing accuracy. This development was considered in this thesis, but has been left out in favor of machine learning. It has been proved through this thesis that machine learning is another possible exploitation of the database.

# Résumé

Les matériaux nanoporeux ayant une surface spécifique élevée suscitent depuis longtemps un grand intérêt. Ils sont utilisés dans une grande variété d'applications telles que le stockage de gaz, la catalyse ou la détection. Les 15 dernières années ont vu naître de nouvelles classes de matériaux nanoporeux cristallins basés sur des liaisons plus faibles. Ces nouvelles classes comprennent les *metal-organic frameworks* (MOF) ou les *covalent organic frameworks*. Parmi eux, une nouvelle classe de matériaux a commencé à émerger, appelée "*stimuli-responsive materials*" ou "*soft porous crystals*". Ces matériaux présentent des réponses importantes ou anormales à une stimulation physique ou chimique externe. La modification de la structure induite par la réponse entraîne également, à son tour, une modification des propriétés physiques et chimiques, ce qui rend ces matériaux multifonctionnels. Les *stimuli-responsive materials* présentent alors une grande variété de comportements : adsorption négative, compressibilité linéaire négative, structures photoréactives ou expansion thermique négative, pour n'en citer que quelques-uns. Chacune de ces propriétés peut être utilisée dans différentes applications telles que les capteurs ou les actionneurs, ou dans l'ingénierie de matériaux composites aux propriétés mécaniques et thermiques ciblées, par exemple.

Au cours de ma thèse, j'ai mené des études sur deux familles de matériaux qui ont toutes deux un potentiel d'applications industrielles : les zéolithes, qui sont déjà utilisées dans l'industrie et sont présentes dans de nombreux exemples de la vie quotidienne, et les *metal-organic frameworks* (MOF), qui ont un grand potentiel pour les applications basées sur l'adsorption et pour lesquels des procédés industriels sont actuellement développés.

Les zéolithes sont une famille d'aluminosilicates naturels découverte au $18^{e}$ siècle qui a été largement étudiée jusqu'à présent. En raison de leur nature poreuse, les zéolites ont été étudiées pour leurs propriétés d'adsorption et leur utilisation potentielle comme catalyseurs et tamis moléculaires. Bien que leur adsorption ait été largement étudiée au cours des 50 dernières années, leurs autres caractéristiques, telles que leurs propriétés mécaniques ou thermiques, n'ont suscitées que peu d'intérêt. Si cette lacune tend aujourd'hui à être comblée pour les propriétés mécaniques, les propriétés thermiques, en revanche, n'ont pas reçu la même attention. Elles ont pourtant aussi un comportement "anormal" dans les zéolithes, puisque la plupart des structures présentent une expansion thermique négative, au point que ce soit considéré comme un comportement commun dans cette famille de matériaux. Très peu d'études ont été orientées vers les propriétés thermiques des zéolithes et le besoin d'une étude à grande échelle devient de plus en plus important. C'est pourquoi j'ai réalisé une étude systématique des propriétés thermiques des zéolithes afin de combler le manque de connaissances sur l'expansion thermique négative et sa présence dans la famille des zéolithes.

Les propriétés thermiques telles que l'expansion thermique ne sont pas accessibles par de simples calculs de phonons. En effet, l'approximation harmonique utilisée

pour réaliser des calculs de fréquences ne prend pas en compte les termes anharmoniques (comme les interactions phonon-phonon). De ce fait, cette approximation ne permet pas de mesurer certains phénomènes comme la conductivité thermique ou l'expansion thermique. Pour pouvoir obtenir l'expansion thermique, je me suis servi de l'approximation quasi-harmonique (QHA) qui consiste à introduire aux phonons une dépendence par rapport au volume :

$$F(V,T) = U_0(V) + F_{vib}(V,T) \tag{4.1}$$

Avec $F(V,T)$ l'énergie libre d'Helmholtz, $F_{vib}(V,T)$ sa partie vibrationnelle et $U_0(V)$ l'énergie interne à 0 $K$. Toutes ces grandeurs peuvent être obtenues par une simple optimisation de géométrie ou par des calculs vibratoires harmoniques. Les fréquences $\omega_k(V)$ doivent être déterminées à plusieurs volumes afin de déterminer leur dépendance vis-à-vis du volume. C'est pourquoi les calculs QHA nécessitent plusieurs calculs de fréquences à différents volumes fixes. Une fois cette information connue, on peut déterminer le volume à l'équilibre à chaque température souhaitée en minimisant eq. (4.1) par rapport au volume en maintenant la température constante. Ceci est généralement accompli par le bais d'une équation d'état. Dans mon cas, j'ai utilisé l'équation d'état de Birch-Murnaghan du troisième ordre. Une fois que l'on connait le volume à l'équilibre à n'importe quelle température, il est possible d'obtenir diverses propriétés comme par exemple l'expansion thermique :

$$\alpha_V(T) = \frac{1}{V(T)} \left( \frac{\partial V(T)}{\partial T} \right)_{P=0} \tag{4.2}$$

Cette méthode a été appliquée pour l'étude systématique à l'aide du code CRYSTAL17[171, 170]. Tous les calculs ont été effectués avec la théorie de la fonctionnelle de la densité. Des bases tout-électrons localisées ont été utilisés pour le silicium et l'oxygène. Elles peuvent être trouvées sur la bibliothèque en ligne de CRYSTAL17 avec les acronymes et les références suivantes : Si_88_31G*_nada_1996[172] pour le silicium et O_8_411_muscat_1999[173] pour l'oxygène. Après avoir testé plusieurs fonctionnelles, j'ai décidé de choisir la fonctionnelle hybride PBEsol0[174] pour le traitement de l'échange et de la corrélation et je n'ai pas utilisé de correction de dispersion. Des fichiers input représentatifs et les structures optimisées sont disponibles en ligne sur `https://github.com/fxcoudert/citable-data/`. Les structures de départ ont été tirées de la base de données de l'International Zeolitic Association (IZA) disponible à l'adresse `http://www.iza-structure.org/databases/`, en ne considérant que les structures purement $SiO_2$ entièrement ordonnées (soit un total de 242 au moment de l'étude). Les zéolithes dont la maille unitaire contient plus de 150 atomes n'ont pas été considérées. 134 structures ont réussies à converger dans les contraintes de temps imposées. Les calculs d'approximation quasi-harmonique ont été effectués avec l'algorithme automatisé QHA implémenté dans CRYSTAL17 : j'ai utilisé l'algorithme avec une gamme de déformation volumique de $-1.5\%$ à $3\%$ avec 4 volumes différents et j'ai calculé les propriétés thermiques sur l'intervalle de 10 K à 300 K. Parmi les 134 structures, 120 ont pu converger dans la plage de volume et de température.

Les calculs préliminaires sur 3 topologies ont révélés des résultats très proches des valeurs expérimentales. J'ai observé une valeur de $-7.3 \times 10^{-6}\ K^{-1}$ pour l'expansion thermque du réseau LTA comparé à $-7.4 \times 10^{-6}\ K^{-1}$ pour sa valeur expérimentale. Des résultats similaires ont été obtenus pour les deux autres, me confortant dans le

choix des paramètres. L'expansion thermique a donc été obtenue pour 120 structures et a été représentée dans la fig. 4.23. On peut constater que toutes les structures purement $SiO_2$ étudiées ici présentent de l'expansion thermique négative (NTE). Cela semble être un comportement commun dans la famille des zéolithes car de la NTE a déjà été observé dans de nombreuses études antérieures [104, 105, 151, 176, 178, 179, 180], à la fois expérimentales et théoriques. Le mécanisme derrière ce phénomène a été précédemment suggéré pour les zéolithes comme étant dû à la présence de "rigid unit modes" [127, 181] provenant de la nature dites "corner-sharing" des structures zéolithiques [182]. Je démontre ici que ce mécanisme est en fait générique, puisque les 120 structures étudiées montrent de la NTE dans leur forme pure $SiO_2$.

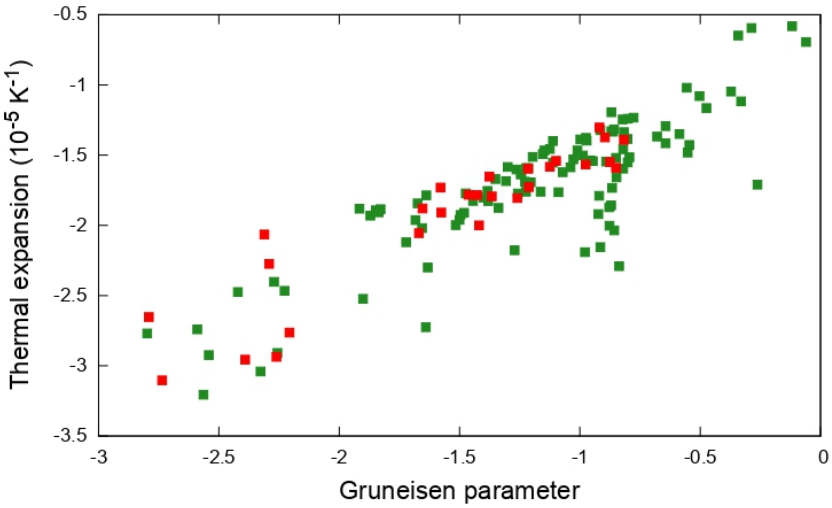


FIGURE 4.23: Coefficients d'expansion thermique volumétrique tracés en fonction du paramètre de Grüneisen moyen pour chaque zéolithe à 300 K. Les carrés verts correspondent aux structures zéolithiques hypothétiques et les rouges aux structures purement $SiO_2$ synthétisées expérimentalement.

L'intervalle des valeurs d'expansion thermique est assez large, ce qui montre que la topologie de la structure a un impact important sur cette propriété — même avec une composition chimique fixe, comme c'est le cas ici. Ceci avait déjà été suggéré par Astala et al.[186] sur un petit nombre de zéolithes et confirmé ici avec un ensemble de structures beaucoup plus important. Dans mes données, j'ai différencié les structures $SiO_2$ théoriques et celles qui ont été synthétisées expérimentalement sous une forme pure $SiO_2$. Ces dernières sont représentées respectivement en vert et en rouge dans la fig. 4.23. Le but de séparer les données de cette manière est de pouvoir déterminer des conditions de synthèse. Cependant, on peut voir sur la figure qu'aucune différence ne peut être observée entre les structures synthétisées et les structures théoriques. J'ai pu observer que les structures expérimentales couvrent presque tout l'intervalle de valeurs, avec peut-être une légère dominance vers les hautes valeurs de NTE, bien que cela puisse être dû à la taille relativement petite de l'échantillonnage. Ceci est en contraste avec les propriétés mécaniques, où il a été montré que la faisabilité expérimentale des structures pouvait être liée à leur stabilité mécanique[161].

En comparant le paramètre de Grüneisen global sur la fig. 4.23 avec l'expansion thermique, une large corrélation peut être observée entre ces deux paramètres sur l'intervalle de température étudié. Comme observé précédemment pour l'expansion

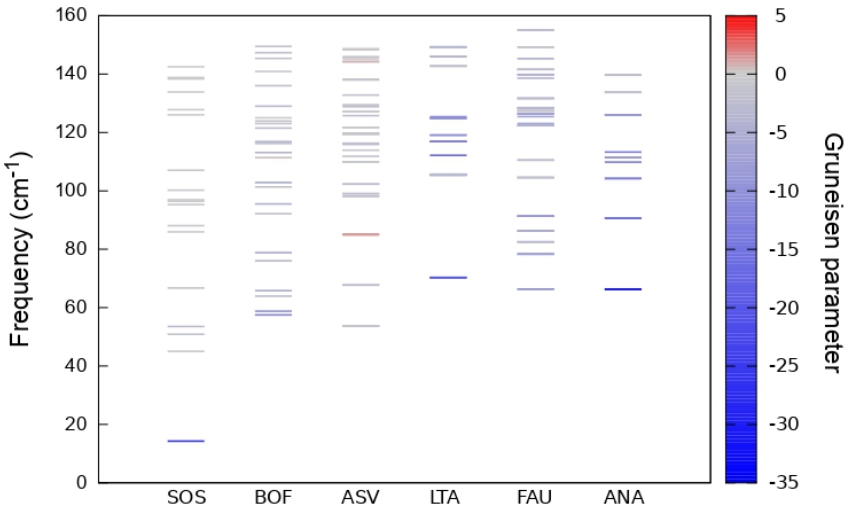


FIGURE 4.24: Représentation des paramètres de Grüneisen de 6 zéolithes (SOS, BOF, ASV, LTA, FAU, ANA) pour tous les modes de vibration sous 160 cm$^{-1}$. Chaque ligne représente un mode de vibration (éventuellement dégénéré en raison de la symétrie), et sa couleur indique la valeur du paramètre de Grüneisen. L'échelle de couleurs est représentée sur la droite.

thermique, toutes les valeurs des paramètres de Grüneisen sont négatives. Cela montre que le modèle de Grüneisen pour étudier le couplage entre les variations de volume et les propriétés vibratoires est raisonnablement valide pour des systèmes tels que les zéolithes. Cette validité est encore confirmée sur la fig. 4.24 où j'ai représenté les paramètres de Grüneisen des modes de vibration à basse fréquence pour plusieurs zéolithes. Cette représentation montre clairement que le paramètre de Grüneisen global négatif n'est influencé que par un petit nombre de modes ayant des paramètres de Grüneisen fortement négatifs associés aux vibrations à basse fréquence. Ceci a été observé sur les 6 topologies présentées ici mais confirmé sur un plus grand nombre de structures. Les modes ayant à des paramètres de Grüneisen fortement négatifs ont des fréquences inférieures à 130 cm$^{-1}$, ce qui confirme que le mécanisme microscopique de la NTE est lié à des modes de vibrations impliquant des rotations concertées de tétraèdres.

Les calculs d'approximation harmonique permettant d'accéder à certaines propriétés mécaniques, j'ai également eu l'opportunité de me pencher sur ces propriétés. Grâce à tous les calculs effectués dans cette étude, j'ai obtenu pour la première fois le module bulk pour chaque structure zéolithique considérée ainsi que sa dépendance en pression et en température. Les valeurs du module bulk pour toutes les structures sont représentées sur la fig. 4.25 par rapport à l'énergie relative (en fonction de l'$\alpha$-quartz). Un large éventail de valeurs est observé, de 8 GPa à 134 GPa avec une valeur moyenne de $\sim$78.5 GPa. Ainsi, comme pour l'expansion thermique, un impact important de la topologie sur le module bulk est observé. Les valeurs extrêmes montrent une différence d'un ordre de grandeur, ce qui est très grand lorsque l'on prend en compte le fait que les matériaux ont la même composition. Comme prévu, cette propriété est donc fortement liée à l'arrangement de la structure (topologie). En regardant le graphique de $K_0$ par rapport à $\Delta E$, quelques remarques peuvent être faites : tout d'abord nous pouvons observer une faible corrélation entre ces deux quantités, ce qui signifie que les matériaux plus mous sont moins stables énergétiquement. Ce

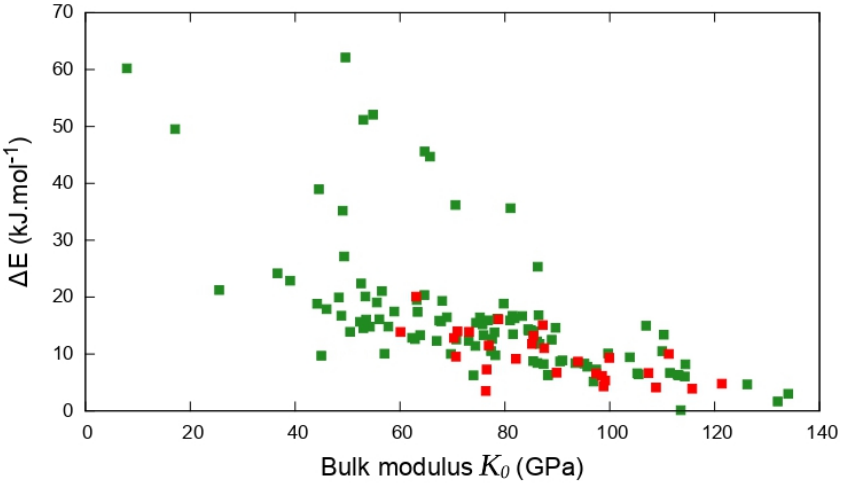


FIGURE 4.25: Énergie relative par rapport au quartz alpha ($\Delta E$) en fonction du module bulk $K_0$ obtenu à travers l'équation d'état de Birch-Murnaghan. Les carrés verts correspondent aux structures zéolithiques hypothétiques et les rouges aux structures purement $SiO_2$ synthétisées expérimentalement.

résultat est intuitif et peut être observé avec la densité également car on sait qu'elle est liée à l'énergie relative. Deuxièmement, on peut voir que les structures réalisables expérimentalement se comportent différemment des structures théoriques car elles ne présentent que des valeurs élevées de $K_0$ (au-dessus de 60 GPa). Seuls les zéolithes théoriques ont été trouvés avec de faibles valeurs de modules bulk. Une valeur élevée de $K_0$ est donc une première exigence pour la synthèse de zéolites purement $SiO_2$.

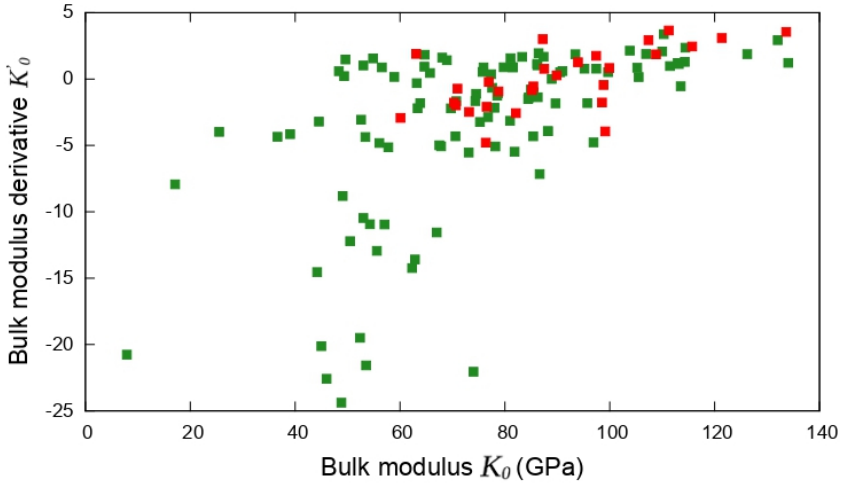


FIGURE 4.26: Module bulk $K_0$ en fonction de sa dérivée sous pression $K'_0$ pour chaque zéolithe. Les carrés verts correspondent aux structures zéolithiques hypothétiques et les rouges aux structures purement $SiO_2$ synthétisées expérimentalement.

À partir des calculs effectués, j'ai également eu accès aux dérivées en pression et en température du module bulk (qui a été obtenu respectivement sous la forme d'un coefficient de Birch-Murnaghan $K'_0$ et à travers l'équation $\delta = (1/K)(\partial K/\partial T)$). En

comparant $K_0$ et sa dérivée en pression sur la fig. 4.26, on peut observer qu'il y a à nouveau un effet important de la topologie sur $K'_0$ qui présente des valeurs positives et négatives. La plupart des structures sont trouvées dans la région de $-5$ à $+5$ mais quelques valeurs sont trouvées jusqu'à $-25$. On peut voir que beaucoup de zéolithes présentent une valeur négative de $K'_0$. Cela signifie que de nombreuses structures $SiO_2$ présentent un diminution anormal du module élastique avec la compression — qui peut être lié au mécanisme déjà établi de l'amorphisation induite par la pression sur les structures zéolithiques [199, 200]. Un tel comportement induit par la pression a déjà été identifié sur un petit nombre de matériaux (y compris des composés inorganiques poreux)[201, 202, 203]. Il a même été suggéré que le diminution induite par la pression pourrait être généralement lié à une expansion thermique négative [204, 205], bien que nous montrions ici qu'il ne s'agit pas d'un lien direct et systématique : toutes les structures $SiO_2$ étudiées dans ce travail montrent une NTE, mais seulement 56% d'entre elles présentent cette diminution (et la grande majorité des structures sont dans la gamme $-5$ à $+3$).

De plus, on peut noter que parmi les structures présentant une valeur de $K'_0$ inférieure à $-5$, aucune structure expérimentale n'est observée. Ce paramètre apparaît alors comme un critère de synthèse : une valeur trop négative conduirait à des instabilités.

L'ensemble des calculs réalisés dans ce travail a permis de créer une base de données de propriétés de zéolithes $SiO_2$. Cette base de données contient des structures optimisées par DFT ainsi que des propriétés calculées par l'approximation quasi-harmonique. Ces données peuvent être utilisées pour des études ultérieures avec différentes méthodes. Il est par exemple possible de développer des champs de force, car leur création nécessite une grande quantité de calculs DFT. Le *machine learning* (ML) peut également être envisagé pour la prédiction des propriétés thermiques d'autres zéolithes ou de matériaux proches. C'est l'application que nous voulions développer et nous avons donc essayé d'étudier la faisabilité de telles méthodes en utilisant la base de données créée dans ce travail.

Dans sa définition la plus générale, le terme *machine learning* désigne l'ensemble des techniques ou algorithmes qui génèrent d'autres algorithmes, de manière à améliorer progressivement les performances sur une tâche. Il s'agit d'une méthode généralisée, ce qui signifie qu'elle ne nécessite pas un type de donnée spécifique. Les algorithmes de ML ne sont pas directement programmés pour réaliser une tâche spécifique mais plutôt pour apprendre à la réaliser. L'idée principale du ML est d'identifier des modèles dans les données. Il existe deux grandes familles de ML : supervisé et non supervisé. Ici, seul le ML supervisé a été utilisé. Le principe consiste à donner à l'algorithme un ensemble de "données d'apprentissage" contenant à la fois des valeurs d'entrée (également appelées caractéristiques ou descripteurs) et des valeurs de sortie (également appelées valeurs cibles). Les valeurs d'entrée, ou descripteurs, sont des quantités décrivant et différenciant les entrées de la base de données. Pour les structures cristallines telles que les zéolithes, plusieurs options peuvent être considérées comme les angles et les distances par exemple. Les valeurs de sortie correspondent aux propriétés que l'on cherche à prédire. L'algorithme va apprendre les règles générales reliant les valeurs d'entrée et de sortie. Les performances du modèle résultant sont testées à l'aide d'un ensemble de tests, qui correspond généralement à une petite partie des données d'apprentissage laissées de côté avant l'apprentissage. Ce processus s'appelle l'apprentissage du modèle. Une fois cet apprentissage effectué, le modèle nouvellement créé doit être capable de prédire

les valeurs de sortie d'un nouvel ensemble donné de caractéristiques d'entrée encore non vues. aL'objectif principal de ce travail a été de determiner s'il était possible d'utiliser la base de données de zéolithes créée afin de prédire les propriétés thermiques. Pour cela, j'ai donc utilisé les 120 structures optimisées précédemment comme jeu d'entrainement. J'ai défini plusieurs types de descripteurs possibles qui sont regroupés et présentés dans le table 4.4. L'intérêt est de pouvoir comparer les performances de différents types de descripteurs sur les zéolithes. J'ai utilisé un algorithme appelé *gradient boosting regression* (GBR)[241, 242] car c'est un algorithme efficace sur de petites bases de données comme la mienne[243]. Je me suis servi du module python sci-kit learn[244] pour effectuer les calculs de ML. Pour choisir et valider les hyperparamètres, j'ai utilisé la *cross-validation* (CV) et choisi comme mesure de la précision la racine de l'erreur quadratique moyenne (RMSE). A l'exception des descripteurs géométriques, toutes les courbes d'apprentissage réalisées montrent le même taux d'apprentissage, et atteignent la précision finale à environ 500 arbres. Cependant, dans le cas des descripteurs géométriques, on peut clairement voir une situation d'*over-fitting* car la RMSE sur le jeu de test commence à augmenter à environ 250 étapes de boosting alors qu'elle continue de diminuer dans le jeu d'entraînement. Par conséquent, j'ai décidé d'utiliser 250 arbres de décisions pour cet ensemble particulier de descripteurs. Pour les autres ensembles, ce phénomène ne se produit pas. C'est pourquoi j'ai choisi d'utiliser un nombre plus élevé d'arbres de décision (500). L'ensemble final des hyperparamètres utilisés pour cette étude est indiqué dans la table 4.5.

| Geometric descriptor | Unit | Volumetric descriptor | Unit |
|---|---|---|---|
| Si–O mean | Å | Accessible volume (AV) | $Å^3$ |
| Si–O harmonic mean | Å | Non accessible volume (NAV) | $Å^3$ |
| Si–O geometric mean | Å | AV + NAV | $Å^3$ |
| Si–O minimum value | Å | Accessible surface area (ASA) | $Å^2$ |
| Si–O maximum value | Å | Non accessible surface area (NASA) | $Å^2$ |
| Si–O variance | Å | ASA + NASA | $Å^2$ |
| Si–O–Si mean | ° | Density | $g.cm^{-3}$ |
| Si–O–Si harmonic mean | ° | Volume | Å |
| Si–O–Si geometric mean | ° | Largest free sphere | Å |
| Si–O–Si minimum value | ° | Largest included sphere | Å |
| Si–O–Si maximum value | ° | Number of channels contributing to ASA | - |
| Si–O–Si variance | ° | Number of channels contributing to AV | - |

| SOAP descriptors | Unit | Topological descriptors | Unit |
|---|---|---|---|
| Principal components of the SOAP analysis from the first to the 12th (PCAn, n=1, 2, ... ,12) | - | Number of neighbours in the successive spheres of coordination from the 2nd to the 13th (COORDn, n=2, 3, ..., 13) | - |

TABLE 4.4: Liste complète des descripteurs utilisés.

J'ai commencé mon étude en utilisant l'ensemble des descripteurs géométriques car les angles Si–O–Si et distances Si–O sont connues pour décrire efficacement les structures des zéolithes. Le résultat d'un GBR entraîné sur les descripteurs géométriques pour la prédiction des propriétés thermiques est affiché sur la fig. 4.27. On peut voir qu'il y a globalement une bonne prédiction de la dilatation thermique car la plupart des valeurs entre $-2$ et $-1$ $K^{-1}$ sont bien concentrées. Comme la majorité de l'ensemble des données se trouve dans cet intervalle, ces résultats étaient attendus. J'ai obtenu une valeur de RMSE de 4.24 $10^{-6}$ $K^{-1}$ correspondant à une erreur d'environ 20%. Compte tenu du petit ensemble de données que j'utilise, je pense qu'il s'agit d'une erreur raisonnable. Ce degré d'erreur est également observé dans d'autres modèles ML basés sur des descripteurs géométriques simples et prédisant des propriétés physiques macroscopiques[194]. Il est également important de noter que la déviation observée trouve son origine non seulement dans la taille réduite de l'ensemble de données, mais aussi dans l'incertitude des données d'entraînement elles-mêmes. En effet, les calculs DFT réalisés pour créer cette base de données sont accompagnés d'un certain degré d'incertitude. La QHA, utilisé pour calculer les propriétés thermiques, est un processus assez long et difficile. Dans le cadre d'une étude systématique, il est nécessaire de fixer un certain ensemble de paramètres pour toutes les structures (plages de température, modification du volume, etc.) Ces paramètres sont alors satisfaisants pour un grand nombre de zéolithes mais ne peuvent être adaptés dans certains cas extrêmes qui diffèrent beaucoup de la "structure moyenne". Des paramètres plus finement définis pour chaque zéolithe auraient pu permettre une meilleure précision des propriétés calculées. Néanmoins ce n'est pas réalisable dans une approche systématique. Le calcul des importances relatives des descripteurs a montré que les statistiques sur les angles sont plus importantes que les distances. Cela confirme l'intuition physique, car la dilatation thermique est dominée par des modes de vibration à basse fréquence, qui impliquent généralement des rotations des tétraèdres $SiO_4$ et des flexions d'angle Si–O–Si.

| **Parameter** | **Value** |
|---|---|
| Number of boosting stages | 250 [a] / 500 [b] |
| Learning rate | 0.01 |
| Minimum samples split | 2 |
| Maximum depth | 2 |
| Minimum samples leaf | 2 |
| Subsample | 0.4 |
| max features | square root of total features |
| loss function | least squares |

[a] Used for the geometric descriptors only to avoid over-fitting observed in the learning curves.

[b] Used for all other descriptors.

TABLE 4.5: Hyperparamètres choisis pour l'algorithme GBR.

J'ai ensuite décidé d'essayer de prédire l'expansion thermique avec les autres ensembles de descripteurs en utilisant exactement la même méthodologie. J'ai observé que, parmi les différents ensembles, les composantes principales des caractéristiques SOAP semblent être les plus performantes pour la prédiction de l'expansion thermique, avec une valeur de RMSE de 3.75 $10^{-6}$ $K^{-1}$. Cette valeur est encore plus faible que la RMSE obtenue précédemment avec les descripteurs géométriques. Cela est dû au fait que les caractéristiques SOAP contiennent les informations sur les angles

et les distances tout en incluant des effets plus complexes, ce qui donne une description complète de la géométrie locale. Ces descripteurs se sont également avérés être les plus efficaces pour la prédiction du module bulk. Ils ont donné une valeur de RMSE de 15,9 GPa qui est très proche de la valeur de 16,0 GPa obtenue avec les descripteurs géométriques. Cela montre une fois de plus que les angles Si–O–Si sont cruciaux pour la détermination des propriétés des zéolithes, car la différence de prédiction entre les descripteurs géométriques et les composantes principales des caractéristiques SOAP est vraiment faible (bien que les caractéristiques SOAP contiennent plus d'informations).

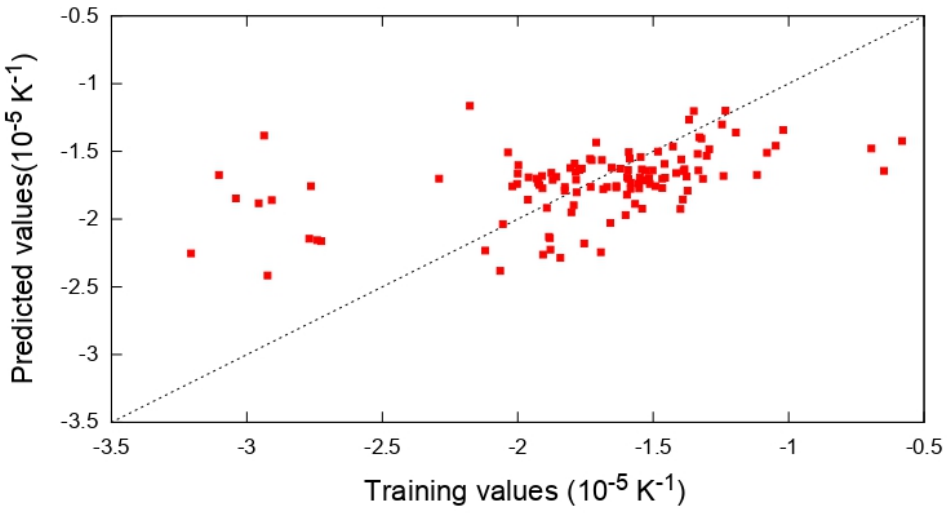


FIGURE 4.27: GBR cross-validation prediction of the thermal expansion $\alpha$, based on geometric descriptors.

Il est également intéressant de comparer différents types de descripteurs lorsqu'ils sont combinés dans le même modèle. J'ai donc construit et entraîné un modèle avec le même nombre total de descripteurs et les mêmes hyperparamètres mais en utilisant cette fois les 3 meilleurs descripteurs de chaque type. La prédiction de l'expansion thermique a donné une précision légèrement meilleure que précédemment avec un RMSE de $3.64.10^{-6}$ $K^{-1}$. L'amélioration est néanmoins anecdotique et principalement localisée sur les faibles valeurs de dilatation thermique. L'importance relative des descripteurs a révélé que la deuxième composante principale des caractéristiques SOAP est le descripteur le plus important pour la prédiction de l'expansion thermique. C'était déjà le cas avec un modèle entrainé uniquement avec les SOAP. La dimensionnalité des caractéristiques SOAP devait être réduite afin d'être exploitable et les techniques de réduction de dimensionnalité entraînent généralement une certaine perte d'informations. Cependant, l'importance des SOAP prouve que cette perte n'a pas d'impact sur le potentiel descriptif de ces descripteurs. Les composantes principales des caractéristiques SOAP représentent alors des descripteurs très compétitifs, pesant même plus que les angles Si–O–Si dans l'entrainement d'un modèle.

Jusqu'à présent, j'ai prédit les propriétés des structures de zéolites qui ont été optimisées en DFT. Les prédictions ont prouvé qu'une utilisation de cette base de données pour du *machine learning* est possible avec une précision raisonnable. Cependant, bien que cela soit intéressant, j'ai considéré pour l'instant que les structures

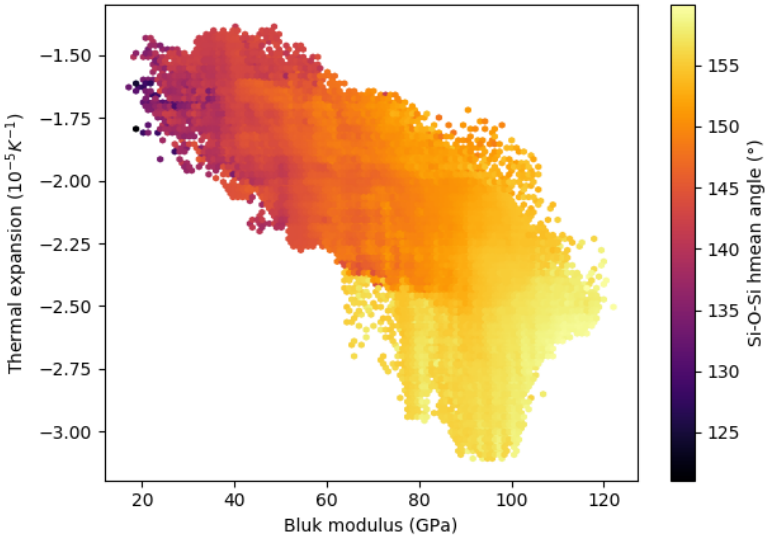


FIGURE 4.28: GBR prediction of the PCOD2 database on the thermal expansion and bulk modulus using the geometrical descriptors and our data set as training set.

pour lesquelles les propriétés devaient être prédites étaient optimisées en DFT, tout comme la base de données que j'ai créée. Ce besoin de précision équivalent à la DFT constitue un sérieux obstacle aux études potentielles de criblage. J'ai donc décidé de vérifier comment un modèle basé sur des structures non optimisées par la DFT se comporterait dans la prédiction des propriétés thermiques et mécaniques obtenues par la DFT. Ce modèle particulier donnera alors une indication de la sensibilité du modèle ML à la précision des géométries utilisées en entrée. En pratique, j'ai utilisé un nouvel ensemble de données de structures extraites de la base de données de l'IZA et contenant exactement les mêmes 120 topologies utilisées jusqu'à présent. J'ai utilisé la même combinaison de descripteurs utilisés précédemment ainsi que les même hyperparamètres. La RMSE obtenue pour ce nouveau modèle est de $3.63.10^{-6}$ $K^{-1}$ ce qui est très proche de la valeur de $3.64.10^{-6}$ $K^{-1}$ obtenue précédemment. Les importances relatives des descripteurs sont également très similaires entre ce modèle et le précédent. Cela est dû au fait que les structures de la base de données de l'IZA et les structures optimisées par la DFT sont proches les unes des autres, ce qui se traduit par des descripteurs proches également. Par conséquent, cela valide la prédiction des propriétés à partir de structures optimisées à un niveau inférieur à la DFT (par exemple, les structures optimisées par champ de force). Cette conclusion représente un gain de temps, car elle permet d'étudier des bases de données zéolitiques à très grande échelle, comme la base de données PCOD2 créée par Deem et al.[245, 246].

J'ai donc décidé d'essayer de prédire les propriétés de la base de données PCOD2 en utilisant ma base de données calculée par DFT. Après avoir comparé la distribution des descripteurs entre les deux bases de données, j'ai trouvé que les descripteurs géométriques représentaient le mieux la variété des structures de la base de données de structures hypothétiques. J'ai donc utilisé les descripteurs géométriques avec les mêmes hyperparamètres que ceux décrits précédemment avec ma base de données

de zéolite calculée par DFT comme jeu d'apprentissage. Les résultats de la prédiction ont été tracés sous forme de carte thermique, en utilisant un gradient de couleur comme troisième dimension. C'est ce que montre la fig. 4.28. Certaines tendances générales peuvent être observées à partir de cette figure, comme la corrélation statistique entre l'expansion thermique et le module bulk. Les structures ayant un module bulk plus élevé ont tendance à présenter une valeur de dilatation thermique plus grande (négative). Il apparait également que les deux quantités sont liées à la même caractéristique géométrique, à savoir l'angle Si–O–Si. Comme déjà expliqué précédemment, les angles se sont avérés être l'un des descripteurs les plus importants dans la prédiction des propriétés des zéolithes. Ici, il apparaît que les matériaux avec une valeur d'angle plus élevée présentent à la fois un module bulk plus élevé et une expansion thermique négative plus importante. Cette corrélation entre le comportement mécanique et thermique à travers une caractéristique géométrique relativement simple est un nouveau développement intéressant, qui devra être confirmé (à travers des calculs systématiques de structures représentatives de la base de données PCOD2 par exemple).

Dans cette thèse, j'ai étudié le comportement thermique d'un grand nombre de zeolithes purement $SiO_2$ par des calculs DFT. L'étude systématique a confirmé certaines hypothèses telles que la présence récurrente d'une expansion thermique négative dans les zéolithes et le fait que cette propriété est fortement liée à des modes de vibration de basses fréquences avec un paramètre de Grüneisen très négatif. Ce travail a également permis de créer une base de données exploitable de structures optimisées par DFT avec les propriétés calculées correspondantes. Cette base de données a ensuite été exploitée à l'aide du *machine learning*. J'ai démontré que ma base de données est appropriée pour la prédiction de l'expansion thermique. J'ai utilisé différents types de descripteurs pour comparer leurs performances et j'ai constaté que les méthodes de description de l'environnement comme le SOAP étaient les plus efficaces. En combinant les différents types de descripteurs, j'ai réussi à prédire le module bulk et l'expansion thermique de la base de données PCOD2 qui contient plus de 600 000 structures.

Cette thèse ouvre un certain nombre de questions et de développements possibles. Tout d'abord, la question de l'effet de la composition sur les propriétés des zéolithes. En effet, ici j'ai étudié l'effet de la topologie en gardant la composition de toutes les structures identique. Cependant, la plupart des zéolithes synthétisées ne le sont pas dans la composition $SiO_2$ et sont souvent composées d'autres atomes comme l'aluminium. La question que l'on peut se poser ici est la suivante : les résultats trouvés pour les zéolithes $SiO_2$ sont-ils transposables aux zéolithes isostructurales de compositions différentes ? Ou, en d'autres termes, les propriétés calculées dans ce travail sont-elles uniquement déterminées par la topologie du matériau, ou la composition joue-t-elle également un rôle ? La réponse viendrait de l'étude de matériaux de compositions différentes et de même topologie. Ceci pourrait être fait sur un petit nombre de structures, permettant de considérer un grand nombre de compositions et d'effectuer une analyse approfondie pour chaque cas. Une étude systématique pourrait également être envisagée, bien que l'inclusion du même nombre de structures par rapport à ce travail entraînerait un coût de calcul encore plus élevé. En effet, la même méthodologie utilisée dans cette thèse devrait être effectuée sur chaque composition différente de chaque structure. Cela limite le nombre de compositions qui peuvent être incluses, mais permettrait d'étudier les effets systématiques de la composition.

L'étude systématique des zéolithes $SiO_2$ a permis de créer une base de données de structures optimisées avec leurs propriétés correspondantes. Cette base de données s'ajoute aux ressources déjà existantes pour les zéolithes comme la base de données IZA ou la base de données PCOD2 de structures hypothétiques. C'est un outil qui peut compléter les autres bases de données ou être utilisé à des fins de comparaison. Mais elle peut aussi être exploitée seule pour diverses applications. L'une des principales limites aux études systématiques est le coût de calcul. La seule façon de calculer de plus en plus de structures en même temps est de sacrifier la précision, proportionnellement au nombre d'entrées à calculer. Les méthodes de champ de force (CF) sont bien indiquées pour cela. Les CF sont la plupart du temps développés et optimisés avec un grand nombre de calculs sur différents matériaux. Ils visent à reproduire la même précision sur laquelle leur développement est basé, à condition d'être utilisées sur la même famille de matériaux que celle utilisée pour leur optimisation. Certains CF sont orientés vers des applications plus générales et donc moins précises, tandis que d'autres sont beaucoup plus spécifiques et précis pour une gamme limitée de matériaux. Le développement de CFs nécessitant beaucoup de calculs de haute précision, la base de données calculée dans ce travail est adaptée à la création d'un CF pour les zéolithes. Cela pourrait permettre des études systématiques plus importantes sur cette famille de matériaux sans avoir à sacrifier la précision.

# Bibliography


[1] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, and A. Aspuru-Guzik. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*, 15(10):1120–1127, August 2016.

[2] G. Férey. Microporous solids: from organically templated inorganic skeletons to hybrid frameworks...ecumenism in chemistry. *Chemistry of Materials*, 13(10):3084–3098, September 2001.

[3] Iza commission on natural zeolite, 2005. Available from: http://www.iza-online.org/natural/Datasheets/Stilbite/stilbite.htm.

[4] T. Motsi, N. A. Rowson, and M. J. H. Simmons. Adsorption of heavy metals from acid mine drainage by natural zeolite. *International Journal of Mineral Processing*, 92(1-2):42–48, July 2009.

[5] J. Perić, M. Trgo, and N. Vukojević Medvidović. Removal of zinc, copper and lead by natural zeolite—a comparison of adsorption isotherms. *Water Research*, 38(7):1893–1899, April 2004.

[6] L. Ćurković, Š. Cerjan-Stefanović, and T. Filipan. Metal ion exchange by natural and modified zeolites. *Water Research*, 31(6):1379–1382, June 1997.

[7] A. Dyer. *An introduction to zeolite molecular sieves*. Elsevier Science - 3rd edition, 10 2007.

[8] J. dhainaut. *Vers l'identification d'inhibiteurs de croissance pour la synthèse de cristaux de zéolithes de taille nanométrique*. PhD thesis, université de Haute Alsace, 2012.

[9] T. Abdullahi, Z. Harun, and M. H. D. Othman. A review on sustainable synthesis of zeolite from kaolinite resources via hydrothermal process. *Advanced Powder Technology*, 28(8):1827–1840, August 2017.

[10] G. Zhu, S. Qiu, J. Yu, Y. Sakamoto, F. Xiao, R. Xu, and O. Terasaki. Synthesis and characterization of high-quality zeolite LTA and FAU single nanocrystals. *Chemistry of Materials*, 10(6):1483–1486, May 1998.

[11] D. Akolekar, A. Chaffee, and R. F. Howe. The transformation of kaolin to low-silica x zeolite. *Zeolites*, 19(5-6):359–365, November 1997.

[12] M. A. Camblor, A. Corma, and S. Valencia. Synthesis in fluoride media and characterisation of aluminosilicate zeolite beta. *Journal of Materials Chemistry*, 8(9):2137–2145, 1998.

[13] B. W. Mercer, Jr. The removal of cesium and strontium from condensate wastes with clinoptilolite. United States, 1960.

[14] A. Shanableh and A. Kharabsheh. Stabilization of cd, ni and pb in soil using natural zeolite. *Journal of Hazardous Materials*, 45(2-3):207–217, February 1996.

[15] G. Majano, L. Delmotte, V. Valtchev, and S. Mintova. Al-rich zeolite beta by seeding in the absence of organic template. *Chemistry of Materials*, 21(18):4184–4191, August 2009.

[16] Y. Kamimura, K. Iyoki, S. P. Elangovan, K. Itabashi, A. Shimojima, and T. Okubo. OSDA-free synthesis of MTW-type zeolite from sodium aluminosilicate gels with zeolite beta seeds. *Micrsporous and Mesoporous Materials*, 163:282–290, November 2012.

[17] S. Mintova, N. H. Olson, V. Valtchev, and T. Bein. Mechanism of zeolite a nanocrystal growth from colloids at room temperature. *Science*, 283(5404):958–960, February 1999.

[18] S. Mintova, N. H. Olson, and T. Bein. Electron microscopy reveals the nucleation mechanism of zeolite y from precursor colloids. *Angewandte Chemie - International Edition*, 38(21):3201–3204, November 1999.

[19] U. Olsbye, S. Svelle, M. Bjørgen, P. Beato, T. V. W. Janssens, F. Joensen, S. Bordiga, and K. P. Lillerud. Conversion of methanol to hydrocarbons: How zeolite cavity and pore size controls product selectivity. *Angewandte Chemie International Edition*, 51(24):5810–5831, April 2012.

[20] W. J. Roth, P. Nachtigall, R. E. Morris, P. S. Wheatley, V. R. Seymour, S. E. Ashbrook, P. Chlubná, L. Grajciar, M. Položij, A. Zukal, O. Shvets, and J. Čejka. A family of zeolites with controlled pore size prepared using a top-down method. *Nature Chemistry*, 5(7):628–633, June 2013.

[21] S. Liu. On the thermal stability of zeolite beta. *Journal of Catalysis*, 132(2):432–439, December 1991.

[22] G. Cruciani. Zeolites upon heating: Factors governing their thermal stability and structural changes. *Journal of Physics and Chemistry of Solids*, 67(9-10):1973–1994, September 2006.

[23] F. E. Trigueiro, D. F. J. Monteiro, F. M. Z. Zotin, and E. F. Sousa-Aguiar. Thermal stability of y zeolites containing different rare earth cations. *Journal of Alloys and Compounds*, 344(1-2):337–341, October 2002.

[24] R. Cai and Y. Yan. Corrosion-resistant zeolite coatings. *CORROSION*, 64(3):271–278, March 2008.

[25] M. E. Davis, C. Saldarriaga, C. Montes, J. Garces, and C. Crowdert. A molecular sieve with eighteen-membered rings. *Nature*, 331(6158):698–699, February 1988.

[26] C. T. Kresge, M. E. Leonowicz, W. J. Roth, J. C. Vartuli, and J. S. Beck. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature*, 359(6397):710–712, October 1992.

[27] O. M. Yaghi and H. Li. Hydrothermal synthesis of a metal-organic framework containing large rectangular channels. *Journal of American Chemical Society*, 117(41):10401–10402, October 1995.

[28] B. F. Hoskins and R. Robson. Infinite polymeric frameworks consisting of three dimensionally linked rod-like segments. *Journal of the American Chemical Society*, 111(15):5962–5964, July 1989.

[29] H. Li, M. Eddaoudi, M. O'Keeffe, and O. M. Yaghi. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature*, 402(6759):276–279, November 1999.

[30] 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. *Proceedings of the National Academy of Sciences*, 103(27):10186–10191, June 2006.

[31] M. Kandiah, M. Hellner Nilsen, S. Usseglio, S. Jakobsen, U. Olsbye, M. Tilset, C. Larabi, E. Alessandra Quadrelli, F. Bonino, and K. Petter Lillerud. Synthesis and stability of tagged UiO-66 zr-MOFs. *Chemistry of Materials*, 22(24):6632–6640, December 2010.

[32] M. Schlesinger, S. Schulze, M. Hietschold, and M. Mehring. Evaluation of synthetic methods for microporous metal–organic frameworks exemplified by the competitive formation of [cu2(btc)3(h2o)3] and [cu2(btc)(OH)(h2o)]. *Microporous and Mesoporous Materials*, 132(1-2):121–127, July 2010.

[33] J. Kim, H.-Y. Cho, and W.-S. Ahn. Synthesis and adsorption/catalytic properties of the metal organic framework CuBTC. *Microporous and Mesoporous Materials*, 16(2):106–119, April 2012.

[34] A. Pichon, A. Lazuen-Garay, and S. L. James. Solvent-free synthesis of a microporous metal–organic framework. *CrystEngComm*, 8(3):211, 2006.

[35] A. Pichon and S. L. James. An array-based study of reactivity under solvent-free mechanochemical conditions—insights and trends. *CrystEngComm*, 10(12):1839, 2008.

[36] P. J. Beldon, L. Fábián, R. S. Stein, A. Thirumurugan, A. K. Cheetham, and T. Friščić. Rapid room-temperature synthesis of zeolitic imidazolate frameworks by using mechanochemistry. *Angewandte Chemie International Edition*, 49(50):9640–9643, November 2010.

[37] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi. The chemistry and applications of metal-organic frameworks. *Science*, 341(6149), August 2013.

[38] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, and O. M. Yaghi. Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science*, 295(5554):469–472, January 2002.

[39] N. L. Rosi, J. Kim, M. Eddaoudi, B. Chen, M. O'Keeffe, and O. M. Yaghi. Rod packings and metal-organic frameworks constructed from rod-shaped secondary building units. *Journal of the American Chemical Society*, 127(5):1504–1518, January 2005.

[40] K. K. Tanabe and S. M. Cohen. Engineering a metal-organic framework catalyst by using postsynthetic modification. *Angewandte Chemie International Edition*, 48(40):7424–7427, September 2009.

[41] D. Farrusseng, S. Aguado, and C. Pinel. Metal-organic frameworks: Opportunities for catalysis. *Angewandte Chemie International Edition*, 48(41):7502–7513, September 2009.

[42] S. Couck, J. F. M. Denayer, G. V. Baron, T. Rémy, J. Gascon, and F. Kapteijn. An amine-functionalized MIL-53 metal-organic framework with large separation power for CO2 and CH4. *Journal of the American Chemical Society*, 131(18):6326–6327, April 2009.

[43] S. Couck, E. Gobechiya, C. E. A. Kirschhock, P. Serra-Crespo, J. Juan-Alcañiz, A. Martinez Joaristi, E. Stavitski, J. Gascon, F. Kapteijn, G. V. Baron, and J. F. M. Denayer. Adsorption and separation of light gases on an amino-functionalized metal-organic framework: An adsorption and in situ XRD study. *ChemSusChem*, 5(4):740–750, February 2012.

[44] T. Wu, L. Shen, M. Luebbers, C. Hu, Q. Chen, Z. Ni, and R. I. Masel. Enhancing the stability of metal–organic frameworks in humid air by incorporating water repellent functional groups. *Chemical Communications*, 46(33):6120, 2010.

[45] J. Yang, A. Grzech, F. M. Mulder, and T. J. Dingemans. Methyl modified MOF-5: a water stable hydrogen storage material. *Chemical Communications*, 47(18):5244, 2011.

[46] T.-H. Chen, I. Popov, O. Zenasni, O. Daugulis, and O. Š. Miljanić. Superhydrophobic perfluorinated metal–organic frameworks. *Chemical Communications*, 49(61):6846, 2013.

[47] H. Jasuja, Y. g. Huang, and K. S. Walton. Adjusting the stability of metal–organic frameworks under humid conditions by ligand functionalization. *Langmuir*, 28(49):16874–16880, November 2012.

[48] H. Jasuja, N. C. Burtch, Y. g. Huang, Y. Cai, and K. S. Walton. Kinetic water stability of an isostructural family of zinc-based pillared metal–organic frameworks. *Langmuir*, 29(2):633–642, January 2013.

[49] H. Deng, C. J. Doonan, H. Furukawa, R. B. Ferreira, J. Towne, C. B. Knobler, B. Wang, and O. M. Yaghi. Multiple functional groups of varying ratios in metal-organic frameworks. *Science*, 327(5967):846–850, February 2010.

[50] O. K. Farha, A. Özgür Yazaydın, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr, and J. T. Hupp. De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage capacities. *Nature Chemistry*, 2(11):944–948, September 2010.

[51] D. Yuan, D. Zhao, D. Sun, and H.-C. Zhou. An isoreticular series of metal-organic frameworks with dendritic hexacarboxylate ligands and exceptionally high gas-uptake capacity. *Angewandte Chemie International Edition*, 49(31):5357–5361, June 2010.

[52] Y. Yan, I. Telepeni, S. Yang, X. Lin, W. Kockelmann, A. Dailly, A. J. Blake, W. Lewis, G. S. Walker, D. R. Allan, S. A. Barnett, N. R. Champness, and M. Schröder. Metal-organic polyhedral frameworks: High h2 adsorption capacities and neutron powder diffraction studies. *Journal of the American Chemical Society*, 132(12):4092–4094, March 2010.

[53] B. Li, Z. Zhang, Y. Li, K. Yao, Y. Zhu, Z. Deng, F. Yang, X. Zhou, G. Li, H. Wu, N. Nijem, Y. J. Chabal, Z. Lai, Y. Han, Z. Shi, S. Feng, and J. Li. Enhanced binding affinity, remarkable selectivity, and high capacity of CO2 by dual functionalization of a rht-type metal-organic framework. *Angewandte Chemie International Edition*, 51(6):1412–1415, December 2011.

[54] J. Hafizovic Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, and K. Petter Lillerud. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *Journal of the American Chemical Society*, 130(42):13850–13851, September 2008.

[55] V. Colombo, S. Galli, H. Jin Choi, G. Ddeul Han, A. Maspero, G. Palmisano, N. Masciocchi, and J. R. Long. High thermal and chemical stability in pyrazolate-bridged metal–organic frameworks with exposed metal sites. *Chemical Science*, 2(7):1311, 2011.

[56] A. Kuc, A. Enyashin, and G. Seifert. Metal-organic frameworks: structural, energetic, electronic, and mechanical properties. *The Journal of Physical Chemistry B*, 111(28):8179–8186, June 2007.

[57] A. Schneemann, V. Bon, I. Schwedler, I. Senkovska, S. Kaskel, and R. A. Fischer. Flexible metal–organic frameworks. *Chem. Soc. Rev.*, 43(16):6062–6096, May 2014.

[58] N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, J. Kim, M. O'Keeffe, and O. M. Yaghi. Hydrogen storage in microporous metal-organic frameworks. *Science*, 300(5622):1127–1129, May 2003.

[59] M. P. Suh, H. J. Park, T. K. Prasad, and D.-W. Lim. Hydrogen storage in metal–organic frameworks. *Chemical Reviews*, 112(2):782–835, December 2011.

[60] B. Chen, X. Zhao, A. Putkham, K. Hong, E. B. Lobkovsky, E. J. Hurtado, A. J. Fletcher, and K. M. Thomas. Surface interactions and quantum kinetic molecular sieving for h2 and d2 adsorption on a mixed metal-organic framework material. *Journal of the American Chemical Society*, 130(20):6411–6423, April 2008.

[61] S. S. Han and W. A. Goddard. Lithium-doped metal-organic frameworks for reversible h2storage at ambient temperature. *Journal of the American Chemical Society*, 129(27):8422–8423, July 2007.

[62] K. L. Mulfort, O. K. Farha, C. L. Stern, A. A. Sarjeant, and J. T. Hupp. Post-synthesis alkoxide formation within metal-organic framework materials: A strategy for incorporating highly coordinatively unsaturated metal ions. *Journal of the American Chemical Society*, 131(11):3866–3868, February 2009.

[63] M. Dincă and J. R. Long. High-enthalpy hydrogen adsorption in cation-exchanged variants of the microporous metal-organic framework mn3[(mn4cl)3(BTT)8(CH3oh)10]2. *Journal of the American Chemical Society*, 129(36):11172–11176, August 2007.

[64] H. Furukawa, N. Ko, Y. B. Go, N. Aratani, S. B. Choi, E. Choi, A. Ö. Yazaydin, R. Q. Snurr, M. O'Keeffe, J. Kim, and O. M. Yaghi. Ultrahigh porosity in metal-organic frameworks. *Science*, 329(5990):424–428, July 2010.

[65] J. Arno, O. K. Farha, W. Morris, P. W. Siu, G. M Tom, M. H. Weston, and P. E. Fuller. ION-x dopant gas delivery system performance characterization at axcelis. In *2018 22nd International Conference on Ion Implantation Technology (IIT)*. IEEE, September 2018.

[66] P. L. Llewellyn, S. Bourrelly, C. Serre, A. Vimont, M. Daturi, L. Hamon, G. De Weireld, J.-S. Chang, D.-Y. Hong, Y. K. Hwang, S. H. Jhung, and G. Férey. High uptakes of CO2 and CH4 in mesoporous metal—organic frameworks MIL-100 and MIL-101. *Langmuir*, 24(14):7245–7250, March 2008.

[67] 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. *Chemical Reviews*, 112(2):724–781, December 2011.

[68] G. Férey, C. Serre, T. Devic, G. Maurin, H. Jobic, P. L. Llewellyn, G. De Weireld, A. Vimont, M. Daturi, and J.-S. Chang. Why hybrid porous solids capture greenhouse gases? *Chem. Soc. Rev.*, 40(2):550–562, 2011.

[69] M. Babaei, S. Salehi, M. Anbia, and M. Kazemipour. Improving CO2 adsorption capacity and CO2/CH4 selectivity with amine functionalization of MIL-100 and MIL-101. *Journal of Chemical & Engineering Data*, 63(5):1657–1662, April 2018.

[70] C. P. Cabello, G. Berlier, G. Magnacca, P. Rumori, and G. T. Palomino. Enhanced CO2 adsorption capacity of amine-functionalized MIL-100(cr) metal–organic frameworks. *CrystEngComm*, 17(2):430–437, 2015.

[71] L. Wang, F. Zhang, C. Wang, Y. Li, J. Yang, L. Li, and J. Li. Ethylenediamine-functionalized metal organic frameworks MIL-100(cr) for efficient CO2/n2o separation. *Separation and Purification Technology*, 235:116219, March 2020.

[72] M. Anbia and V. Hoseini. Enhancement of CO2 adsorption on nanoporous chromium terephthalate (MIL-101) by amine modification. *Journal of Natural Gas Chemistry*, 21(3):339–343, May 2012.

[73] J. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S. T. Nguyen, and J. T. Hupp. Metal–organic framework materials as catalysts. *Chemical Society Reviews*, 38(5):1450, 2009.

[74] J. Gascon, A. Corma, F. Kapteijn, and F. X. Llabrés i Xamena. Metal organic framework catalysis: Quo vadis? *ACS Catalysis*, 4(2):361–378, December 2013.

[75] C. Kato, M. Hasegawa, T. Sato, A. Yoshizawa, T. Inoue, and W. Mori. Microporous dinuclear copper(II) trans-1, 4-cyclohexanedicarboxylate: heterogeneous oxidation catalysis with hydrogen peroxide and x-ray powder structure of peroxo copper(II) intermediate. *Journal of Catalysis*, 230(1):226–236, February 2005.

[76] S. Horike, M. Dincă, K. Tamaki, and J. R. Long. Size-selective lewis acid catalysis in a microporous metal-organic framework with exposed mn2+ coordination sites. *Journal of the American Chemical Society*, 130(18):5854–5855, April 2008.

[77] A. Henschel, K. Gedrich, R. Kraehnert, and S. Kaskel. Catalytic properties of MIL-101. *Chemical Communications*, (35):4192, 2008.

[78] 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. *Chem. Commun.*, (27):2820–2822, 2007.

[79] E. Schrödinger. Quantisierung als Eigenwertproblem (Erste Mitteilung). (German) [Quantization as an eigenvalue problem (First report)]. *Annalen der Physik (1900)*, 384(4):361–376, 1926.

[80] E. Schrödinger. Quantisierung als Eigenwertproblem (Zweite Mitteilung). (German) [Quantization as an eigenvalue problem (Second report)]. *Annalen der Physik (1900)*, 384(6):489–527, 1926.

[81] E. Schrödinger. Quantisierung als Eigenwertproblem (Dritte Mitteilung: Störungstheorie, mit Anwendung auf den Starkeffekt der Balmerlinien). (German) [Quantization as an eigenvalue problem (Third report: perturbation theory, with application to the Stark effect of the Balmer lines)]. *Annalen der Physik (1900)*, 385(13):437–490, 1926.

[82] E. Schrödinger. Quantisierung als Eigenwertproblem (Vierte Mitteilung). (German) [Quantization as an eigenvalue problem (Fourth report)]. *Annalen der Physik (1900)*, 386(18):109–139, 1926.

[83] L. Butler. Chemical reaction dynamics beyond the born-oppenheimer approximation. *Annual Review of Physical Chemistry*, 49(1):125–171, October 1998.

[84] D. R. Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(1):89–110, January 1928.

[85] V. Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2):126–148, January 1930.

[86] J. C. Slater. A simplification of the hartree-fock method. *Physical Review*, 81(3):385–390, February 1951.

[87] E. Fermi. Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei*, 6(602-607):32, 1927.

[88] L. H. Thomas. The calculation of atomic fields. *Mathematical Proceedings of the Cambridge Philosophical Society*, 23(5):542–548, January 1927.

[89] R. I. G. Hughes. Theoretical practice: the bohm-pines quartet. *Perspectives on Science*, 14(4):457–524, December 2006.

[90] E. Teller. On the stability of molecules in the thomas-fermi theory. *Reviews of Modern Physics*, 34(4):627–631, October 1962.

[91] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[92] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, November 1965.

[93] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B*, 45(23):13244–13249, June 1992.

[94] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.

[95] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, December 1996.

[96] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006.

[97] 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. *The Journal of Chemical Physics*, 132(15):154104, April 2010.

[98] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55(22):2471–2474, November 1985.

[99] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007.

[100] M. Ducamp and F.-X. Coudert. Systematic study of the thermal properties of zeolitic frameworks. *The Journal of Physical Chemistry C*, 125(28):15647–15658, July 2021.

[101] W. Miller, C. W. Smith, D. S. Mackenzie, and K. E. Evans. Negative thermal expansion: a review. *Journal of Materials Science*, 44(20):5441–5451, July 2009.

[102] J. S. O. Evans, Z. Hu, J. D. Jorgensen, D. N. Argyriou, S. Short, and A. W. Sleight. Compressibility, phase transitions, and oxygen migration in zirconium tungstate. *Science*, 275(5296):61–65, January 1997.

[103] J. S. O. Evans. Negative thermal expansion materials. *Journal of the Chemical Society, Dalton Transactions*, (19):3317–3326, 1999.

[104] D. A. Woodcock, P. Lightfoot, L. A. Villaescusa, M.-J. Díaz-Cabañas, M. A. Camblor, and D. Engberg. Negative thermal expansion in the siliceous zeolites chabazite and ITQ-4: a neutron powder diffraction study. *Chemistry of Materials*, 11(9):2508–2514, August 1999.

[105] M. P. Attfield. Strong negative thermal expansion in siliceous faujasite. *Chemical Communications*, (5):601–602, 1998.

[106] M. P. Attfield and A. W. Sleight. Exceptional negative thermal expansion in AlPO4-17. *Chemistry of Materials*, 10(7):2013–2019, June 1998.

[107] A. L. Goodwin, M. Calleja, M. J. Conterio, M. T. Dove, J. S. O. Evans, D. A. Keen, L. Peters, and M. G. Tucker. Colossal positive and negative thermal expansion in the framework material ag 3 [co(CN) 6 ]. *Science*, 319(5864):794–797, February 2008.

[108] E. Grüneisen. Theorie des festen zustandes einatomiger elemente. *Annalen der Physik*, 344(12):257–306, 1912.

[109] E. Grüneisen. Zustand des festen körpers. In *Thermische Eigenschaften der Stoffe*, pages 1–59. Springer Berlin Heidelberg, 1926.

[110] H. Tanaka. Thermodynamic stability and negative thermal expansion of hexagonal and cubic ices. *The Journal of Chemical Physics*, 108(12):4887–4893, March 1998.

[111] H. D. Megaw and D. H. Templeton. Crystal structures: A working approach. *Physics Today*, 27(9):53–54, September 1974.

[112] M. T. Dove, M. Gambhir, K. D. Hammonds, V. Heine, and A. K. A. Pryde. Distortions of framework structures. *Phase Transitions*, 58(1-3):121–143, August 1996.

[113] M. T. Dove. Approaches to modelling the behaviour of ceramics and minerals. *Phase Transitions*, 61(1-4):1–17, May 1997.

[114] K. D. Hammonds, V. Heine, and M. T. Dove. Insights into zeolite behaviour from the rigid unit mode model. *Phase Transitions*, 61(1-4):155–172, May 1997.

[115] P. R. L. Welche, V. Heine, and M. T. Dove. Negative thermal expansion in beta-quartz. *Physics and Chemistry of Minerals*, 26(1):63–77, November 1998.

[116] Y. Yamamura, S. Ikeuchi, and K. Saito. Characteristic phonon spectrum of negative thermal expansion materials with framework structure through calorimetric study of sc2m3o12 (m = w and mo). *Chemistry of Materials*, 21(13):3008–3016, June 2009.

[117] A. K. A. Pryde, K. D. Hammonds, M. T. Dove, V. Heine, J. D. Gale, and M. C. Warren. Origin of the negative thermal expansion in zrw2o8 and zrv2o7. *Journal of Physics: Condensed Matter*, 8(50):10973–10982, December 1996.

[118] D. A. Woodcock, P. Lightfoot, and C. Ritter. Negative thermal expansion in y2(WO4)3. *Journal of Solid State Chemistry*, 149(1):92–98, January 2000.

[119] P. M. Forster, A. Yokochi, and A. W. Sleight. Enhanced negative thermal expansion in lu2w3o12. *Journal of Solid State Chemistry*, 140(1):157–158, October 1998.

[120] G. A. Rossetti, J. P. Cline, and A. Navrotsky. Phase transition energetics and thermodynamic properties of ferroelectric PbTiO3. *Journal of Materials Research*, 13(11):3197–3206, November 1998.

[121] I. Bull, P. Lightfoot, L. A. Villaescusa, L. M. Bull, R. K. B. Gover, J. S. O. Evans, and R. E. Morris. An x-ray diffraction and MAS NMR study of the thermal expansion properties of calcined siliceous ferrierite. *Journal of the American Chemical Society*, 125(14):4342–4349, March 2003.

[122] A. K. Tyagi, S. N. Achary, and M. D. Mathews. Phase transition and negative thermal expansion in a2(MoO4)3 system (a=fe3$^+$, cr3$^+$ and al3$^+$). *Journal of Alloys and Compounds*, 339(1-2):207–210, June 2002.

[123] Negative or Zero Thermal Expansion Materials (Special Issue Name). *Journal of the Chinese Ceramic Society*, 37:651-759, 2009.

[124] J. S. O. Evans, T. A. Mary, T. Vogt, M. A. Subramanian, and A. W. Sleight. Negative thermal expansion in ZrW2o8 and HfW2o8. *Chemistry of Materials*, 8(12):2809–2823, January 1996.

[125] T. A. Mary, J. S. O. Evans, T. Vogt, and A. W. Sleight. Negative thermal expansion from 0.3 to 1050 kelvin in ZrW 2 o 8. *Science*, 272(5258):90–92, April 1996.

[126] J. Graham, A. D. Wadsley, J. H. Weymouth, and L. S. Williams. A new ternary oxide, ZrW2o8. *Journal of the American Ceramic Society*, 42(11):570–570, November 1959.

[127] A. K. A. Pryde, K. D. Hammonds, M. T. Dove, V. Heine, J. D. Gale, and M. C. Warren. Rigid unit modes and the negative thermal expansion in ZrW2o8. *Phase Transitions*, 61(1-4):141–153, May 1997.

[128] V. Korthuis, N. Khosrovani, A. W. Sleight, N. Roberts, R. Dupree, and W. W. Jr. Warren. Negative thermal expansion and phase transitions in the ZrV2-xPxO7 series. *Chemistry of Materials*, 7(2):412–417, February 1995.

[129] N. Khosrovani, A. W. Sleight, and T. Vogt. Structure of ZrV2o7from -263 to 470°c. *Journal of Solid State Chemistry*, 132(2):355–360, September 1997.

[130] T. A. Mary and A. W. Sleight. Bulk thermal expansion for tungstate and molybdates of the type a2m3o12. *Journal of Materials Research*, 14(3):912–915, March 1999.

[131] E. T. Ritz and N. A. Benedek. Strain game revisited for complex oxide thin films: Substrate-film thermal expansion mismatch in PbTiO3. *Physical Review Materials*, 4(8), August 2020.

[132] V. Srikant, E. J. Tarsa, D. R. Clarke, and J. S. Speck. Crystallographic orientation of epitaxial BaTiO3films: The role of thermal-expansion mismatch with the substrate. *Journal of Applied Physics*, 77(4):1517–1522, February 1995.

[133] J. S. O. Evans, T. A. Mary, and A. W. Sleight. Negative thermal expansion in a large molybdate and tungstate family. *Journal of Solid State Chemistry*, 133(2):580–583, November 1997.

[134] J. E. Readman, S. E. Lister, L. Peters, J. Wright, and J. S. O. Evans. Direct synthesis of cubic ZrMo2o8 followed by ultrafast in situ powder diffraction. *Journal of the American Chemical Society*, 131(48):17560–17562, November 2009.

[135] L. C. Kozy, M. N. Tahir, C. Lind, and W. Tremel. Particle size and morphology control of the negative thermal expansion material cubic zirconium tungstate. *Journal of Materials Chemistry*, 19(18):2760, 2009.

[136] Z. Hu, J. D. Jorgensen, S. Teslic, S. Short, D. N. Argyriou, J. S. O. Evans, and A. W. Sleight. Pressure-induced phase transformation in ZrW2o8 — compressibility and thermal expansion of the orthorhombic phase. *Physica B: Condensed Matter*, 241-243:370–372, December 1997.

[137] C. A. Perottoni and J. A. H. da Jornada. Pressure-induced amorphization and negative thermal expansion in ZrW 2 o 8. *Science*, 280(5365):886–889, May 1998.

[138] H. Holzer and D. C. Dunand. Phase transformation and thermal expansion of cu/ZrW2o8 metal matrix composites. *Journal of Materials Research*, 14(3):780–789, March 1999.

[139] S. Yilmaz and D. C. Dunand. Finite-element analysis of thermal expansion and thermal mismatch stresses in a cu–60vol%ZrW2o8 composite. *Composites Science and Technology*, 64(12):1895–1898, September 2004.

[140] X. Yan, X. Cheng, G. Xu, C. Wang, S. Sun, and R. Riedel. Preparation and thermal properties of zirconium tungstate/copper composites. *Materialwissenschaft und Werkstofftechnik*, 39(9):649–653, September 2008.

[141] P. Lommens, C. De Meyer, E. Bruneel, K. De Buysser, I. Van Driessche, and S. Hoste. Synthesis and thermal expansion of ZrO2/ZrW2o8 composites. *Journal of the European Ceramic Society*, 25(16):3605–3610, November 2005.

[142] L. Sun, A. Sneller, and P. Kwon. ZrW2o8-containing composites with near-zero coefficient of thermal expansion fabricated by various methods: Comparison and optimization. *Composites Science and Technology*, 68(15-16):3425–3430, December 2008.

[143] X. Yang, J. Xu, H. Li, X. Cheng, and X. Yan. In situ synthesis of ZrO2/ZrW2o8composites with near-zero thermal expansion. *Journal of the American Ceramic Society*, 90(6):1953–1955, June 2007.

[144] J. Tani, H. Kimura, K. Hirota, and H. Kido. Thermal expansion and mechanical properties of phenolic resin/ZrW2o8composites. *Journal of Applied Polymer Science*, 106(5):3343–3347, August 2007.

[145] X. Chu, R. Huang, H. Yang, Z. Wu, J. Lu, Y. Zhou, and L. Li. The cryogenic thermal expansion and mechanical properties of plasma modified ZrW2o8 reinforced epoxy. *Materials Science and Engineering: A*, 528(9):3367–3374, April 2011.

[146] L. M. Sullivan and C. M. Lukehart. Zirconium tungstate (ZrW2o8)/polyimide nanocomposites exhibiting reduced coefficient of thermal expansion. *Chemistry of Materials*, 17(8):2136–2141, March 2005.

[147] C. Lind, M. R. Coleman, L. C. Kozy, and G. R. Sharma. Zirconium tungstate/polymer nanocomposites: Challenges and opportunities. *physica status solidi (b)*, 248(1):123–129, August 2010.

[148] K. Scheel. Versuche ueber die ausdehnung fester koerper, insbesondere von quarz in richtung der hauptachse, platin, palladium und quarzglas bei der temperatur der fluessigen Luft. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 9:3–23, 1907.

[149] K. Scheel. Ueber die ausdehnung des quarzglases. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 9:719–721, 1907.

[150] P. Tschaufeser and S. C. Parker. Thermal expansion behavior of zeolites and AlPO4s. *The Journal of Physical Chemistry*, 99(26):10609–10615, June 1995.

[151] S. H. Park, R.-W. Große Kunstleve, H. Graetsch, and H. Gies. The thermal expansion of the zeolites MFI, AFI, DOH, DDR, and MTN in their calcined and as synthesized forms. In *Studies in Surface Science and Catalysis*, pages 1989–1994. Elsevier, 1997.

[152] K. Yamahara, K. Okazaki, and K. Kawamura. Molecular dynamics studies on thermal behavior of an MFI-type zeolite. *Catalysis Today*, 23(4):397–402, April 1995.

[153] D. S. Bhange and V. Ramaswamy. Negative thermal expansion in silicalite-1 and zirconium silicalite-1 having MFI structure. *Materials Research Bulletin*, 41(7):1392–1402, July 2006.

[154] D. S. Bhange and V. Ramaswamy. Thermal stability of the mobil five type metallosilicate molecular sieves—an in situ high temperature x-ray diffraction study. *Materials Research Bulletin*, 42(5):851–860, May 2007.

[155] D. S. Bhange and V. Ramaswamy. Enhanced negative thermal expansion in MFI molecular sieves by varying framework composition. *Microporous and Mesoporous Materials*, 130(1-3):322–326, May 2010.

[156] Structural study of pure silica and gecontaining zeolite ITQ-24. In *Tenth European Powder Diffraction Conference*, pages 393–398. Oldenbourg Wissenschaftsverlag, December 2007.

[157] T. Carey, A. Corma, F. Rey, C. C. Tang, J. A. Hriljac, and P. A. Anderson. The effect of extra framework species on the intrinsic negative thermal expansion property of zeolites with the LTA topology. *Chemical Communications*, 48(47):5829, 2012.

[158] X. Wang, J. C. Hanson, J. Szanyi, and J. A. Rodriguez. Interaction of h2o and NO2 with BaY faujasite: complex contraction/expansion behavior of the zeolite unit cell. *The Journal of Physical Chemistry B*, 108(43):16613–16616, September 2004.

[159] P. M. Jardim, B. A. Marinkovic, A. Saavedra, L. Y. Lau, C. Baehtz, and F. Rizzo. A comparison between thermal expansion properties of hydrated and dehydrated orthorhombic HZSM-5 zeolite. *Microporous and Mesoporous Materials*, 76(1-3):23–28, December 2004.

[160] R. Lakes and K. W. Wojciechowski. Negative compressibility, negative poisson's ratio, and stability. *physica status solidi (b)*, 245(3):545–551, February 2008.

[161] F.-X. Coudert. Systematic investigation of the mechanical properties of pure silica zeolites: stiffness, anisotropy, and negative linear compressibility. *Phys. Chem. Chem. Phys.*, 15:16012, 2013.

[162] J. Lin, X. F. Shu, and J. X. Dong. The experimental determination of mechanical properties of zeolite ferrierite crystal. *Materials Letters*, 59(12):1595–1597, 2005.

[163] Z. Li, M. C. Johnson, M. Sun, E. T. Ryan, D. J. Earl, W. Maichen, J. I. Martin, S. Li, C. M. Lew, J. Wang, M. W. Deem, M. E. Davis, and Y. Yan. Mechanical and dielectric properties of pure-silica-zeolite low-k materials. *Angewandte Chemie*, 118(38):6477–6480, 2006.

[164] R. Astala, S. M. Auerbach, and P. A. Monson. Density functional theory study of silica zeolite structures: stabilities and mechanical properties of SOD, LTA, CHA, MOR, and MFI. *The Journal of Physical Chemistry B*, 108(26):9208–9215, May 2004.

[165] J. D. Evans and F.-X. Coudert. Predicting the mechanical properties of zeolite frameworks by machine learning. *Chem. Mater.*, 29:7833–7839, 2017.

[166] A. Erba. On combining temperature and pressure effects on structural properties of crystals with standardab initiotechniques. *J. Chem. Phys.*, 141(12):124115, September 2014.

[167] A. Erba, M. Shahrokhi, R. Moradian, and R. Dovesi. On how differently the quasi-harmonic approximation works for two isostructural crystals: Thermal properties of periclase and lime. *J. Chem. Phys.*, 142(4):044114, January 2015.

[168] A. Erba, J. Maul, M. De La Pierre, and R. Dovesi. Structural and elastic anisotropy of crystals at high pressures and temperatures from quantum mechanical methods: The case of $Mg_2SiO_4$ forsterite. *J. Chem. Phys.*, 142(20):204502, May 2015.

[169] A. Erba, J. Maul, R. Demichelis, and R. Dovesi. Assessing thermochemical properties of materials through ab initio quantum-mechanical methods: the case of $\alpha$-$Al_2O_3$. *Phys. Chem. Chem. Phys.*, 17(17):11670–11677, 2015.

[170] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. Llunell, M. Causà, Y. Noël, L. Maschio, A. Erba, M. Rerat, and S. Casassa. CRYSTAL17 User's Manual. University of Torino: Torino, 2017.

[171] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman. Quantum-mechanical condensed matter simulations with CRYSTAL. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 8(4):e1360, March 2018.

[172] R. Nada, J. B. Nicholas, M. I. McCarthy, and A. C. Hess. Basis sets for ab initio periodic hartree–fock studies of zeolite/adsorbate interactions: He, ne, and ar in silica sodalite. *Int. J. Quantum Chem.*, 60(4):809–820, November 1996.

[173] J. Muscat. *The phase stability, surface structure and defect chemistry of titanium dioxide from first principles techniques*. PhD thesis, University of Manchester, 1999.

[174] 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. *Phys. Rev. Lett.*, 100(13):136406, April 2008.

[175] H. J. Monkhorst and J. D. Pack. Special points for brillouin-zone integrations. *Phys. Rev. B*, 13(12):5188–5192, June 1976.

[176] T. Carey, C. C. Tang, J. A. Hriljac, and P. A. Anderson. Chemical control of thermal expansion in cation-exchanged zeolite a. *Chem. Mater.*, 26(4):1561–1566, February 2014.

[177] M. Baise, P. M. Maffettone, F. Trousselet, N. P. Funnell, F.-X. Coudert, and A. L. Goodwin. Negative hydration expansion in ZrW2o8 : Microscopic mechanism, spaghetti dynamics, and negative thermal expansion. *Physical Review Letters*, 120(26), June 2018.

[178] L. Leardini, S. Quartieri, G. Vezzalini, and R. Arletti. Thermal behaviour of siliceous faujasite: Further structural interpretation of negative thermal expansion. *Micropor. Mesopor. Mater.*, 202:226–233, January 2015.

[179] D. A. Woodcock, P. Lightfoot, P. A. Wright, L. A. Villaescusa, and M. A. Camblor. Strong negative thermal expansion in the siliceous zeolites itq-1, itq-3 and ssz-23. *J. Mater. Chem.*, 9:349–351, 1999.

[180] L. A. Villaescusa, P. Lightfoot, S. J. Teat, and R. E. Morris. Variable-temperature microcrystal x-ray diffraction studies of negative thermal expansion in the

pure silica zeolite IFR. *Journal of the American Chemical Society*, 123(23):5453–5459, May 2001.

[181] M. G. Tucker, A. L. Goodwin, M. T. Dove, D. A. Keen, S. A. Wells, and J. S. O. Evans. Negative thermal expansion inZrW2o8: Mechanisms, rigid unit modes, and neutron total scattering. *Physical Review Letters*, 95(25), December 2005.

[182] J. Z. Tao and A. W. Sleight. The role of rigid unit modes in negative thermal expansion. *Journal of Solid State Chemistry*, 173(2):442–448, July 2003.

[183] J. A. Kosinski, J. G. Gualtieri, and A. Ballato. Thermal expansion of alpha quartz. In *Proceedings of the 45th Annual Symposium on Frequency Control 1991*. IEEE, 1991.

[184] G. D. Barrera, J. A. O. Bruno, T. H. K. Barron, and N. L. Allan. Negative thermal expansion. *J. Phys.: Condens. Matter*, 17:R217–R252, 2005.

[185] Z. Liu, Q. Gao, J. Chen, J. Deng, K. Lin, and X. Xing. Negative thermal expansion in molecular materials. *Chem. Commun.*, 54:5164–5176, 2018.

[186] R. Astala, Scott M. Auerbach, and P. A. Monson. Normal mode approach for predicting the mechanical properties of solids from first principles: Application to compressibility and thermal expansion of zeolites. *Phys. Rev. B*, 71(1), January 2005.

[187] A. Bieniok and K. D. Hammonds. Rigid unit modes and the phase transition and structural distortions of zeolite rho. *Micropor. Mesopor. Mater.*, 25(1-3):193–200, December 1998.

[188] J. D. Gale. Gulp: A computer program for the symmetry-adapted simulation of solids. *Journal of the Chemical Society, Faraday Transactions*, 93(4):629–637, 1997.

[189] M. J. Sanders, M. Leslie, and C. R. A. Catlow. Interatomic potentials for $SiO_2$. *J. Chem. Soc. Chem. Commun.*, pages 1271–1273, 1984.

[190] P. Vieillard. A predictive model for the entropies and heat capacities of zeolites. *Eur. J. Mineral.*, 22(6):823–836, December 2010.

[191] Y. G. Bushuev and G. Sastre. Feasibility of pure silica zeolites. *J. Phys. Chem. C*, 114(45):19157–19168, October 2010.

[192] 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. *Microporous and Mesoporous Materials*, 149(1):134–141, February 2012.

[193] R. L. Martin, B. Smit, and M. Haranczyk. Addressing challenges of identifying geometrically diverse sets of crystalline porous materials. *Journal of Chemical Information and Modeling*, 52(2):308–318, December 2011.

[194] S. Chibani and F.-X. Coudert. Systematic exploration of the mechanical properties of 13 621 inorganic compounds. *Chem. Sci.*, 10:8589–8599, 2019.

[195] R. Gaillac, S. Chibani, and F.-X. Coudert. Speeding up discovery of auxetic zeolite frameworks by machine learning. *Chem. Mater.*, 32:2653–2663, 2020.

[196] M. Siddorn, F.-X. Coudert, K. E. Evans, and A. Marmier. A systematic typology for negative poisson's ratio materials and the prediction of complete auxeticity in pure silica zeolite jst. *Phys. Chem. Chem. Phys.*, 17:17927–17933, 2015.

[197] J. Dagdelen, J. Montoya, M. de Jong, and K Persson. Computational prediction of new auxetic materials. *Nature Commun.*, 8:323, 2017.

[198] M. de Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. van der Zwaag, J. J. Plata, C. Toher, S. Curtarolo, G. Ceder, K. A. Persson, and M. Asta. Charting the complete elastic properties of inorganic crystalline compounds. *Scientific Data*, 2(1), March 2015.

[199] B. Coasne, J. Haines, C. Levelut, O. Cambon, M. Santoro, F. Gorelli, and G. Garbarino. Enhanced mechanical strength of zeolites by adsorption of guest molecules. *Phys. Chem. Chem. Phys.*, 13(45):20096, 2011.

[200] 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. *J. Phys. Chem. Lett.*, 4(11):1861–1865, May 2013.

[201] K. W. Chapman and P. J. Chupas. Pressure enhancement of negative thermal expansion behavior and induced framework softening in zinc cyanide. *J. Am. Chem. Soc.*, 129(33):10090–10091, August 2007.

[202] H. Fang, A. E. Phillips, M. T. Dove, M. G. Tucker, and A. L. Goodwin. Temperature-dependent pressure-induced softening in $Zn(CN)_2$. *Phys. Rev. B*, 88(14):144103, October 2013.

[203] H. Fang and M. T Dove. A phenomenological expression to describe the temperature dependence of pressure-induced softening in negative thermal expansion materials. *J. Phys. Condens. Matter*, 26(11):115402, March 2014.

[204] H. Fang and M. T. Dove. Pressure-induced softening as a common feature of framework structures with negative thermal expansion. *Phys. Rev. B*, 87(21):214109, June 2013.

[205] H. Fang, M. T. Dove, and A. E. Phillips. Common origin of negative thermal expansion and other exotic properties in ceramic and hybrid materials. *Phys. Rev. B*, 89(21):214103, June 2014.

[206] M. Ducamp and F.-X. Coudert. Prediction of thermal properties of zeolites through machine learning. *The Journal of Physical Chemistry C*, 126(3):1651–1660, January 2022.

[207] F. Rosenblatt. The Perceptron, A Perceiving and Recognizing Automaton, Project Para Report No. 85-460-1, *Cornell Aeronautical Laboratory*, 1957.

[208] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[209] F. Rosenblatt. Perceptron simulation experiments. *Proceedings of the IRE*, 48(3):301–309, March 1960.

[210] B. Widrow. Self-adaptive discrete systems, theory self adapf. In *Contr. Syst. Proc. IFAC Symp. 1st*, 1961.

[211] M. M. Bongard. Simulation of the recognition process on a digital computing machine. *Biophysics*, 4(2), 1961.

[212] E. M. Braverman. The experiments with training a machine to recognize patterns. *Automat. Remote Contr*, 23(3), 1962.

[213] M. Minsky and S. Papert. Perceptrons: an introduction to computational chemistry. 1969.

[214] S. I. Toos and A. bagherian and I. A. akhlaghi. Benchmark Problems or Test Functions. November 2011.

[215] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82, April 1997.

[216] M. Fernandez and A. S. Barnard. Identification of nanoparticle prototypes and archetypes. *ACS Nano*, 9(12):11980–11992, November 2015.

[217] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang. An introduction to machine learning. *Clinical Pharmacology & Therapeutics*, 107(4):871–885, March 2020.

[218] F. H. Allen. *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*. International Union of Crystallography, 1987.

[219] Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019. *Journal of Chemical & Engineering Data*, 64(12):5985–5998, November 2019.

[220] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, and O. M. Yaghi. The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. *Accounts of Chemical Research*, 41(12):1782–1789, October 2008.

[221] H. E. Pence and A. Williams. ChemSpider: An online chemical information resource. *Journal of Chemical Education*, 87(11):1123–1124, August 2010.

[222] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108(5), January 2012.

[223] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Physical Review B*, 89(20), May 2014.

[224] L. Ward, R. Liu, A. Krishna, V. I. Hegde, A. Agrawal, A. Choudhary, and C. Wolverton. Including crystal structure attributes in machine learning models of formation energies via voronoi tessellations. *Physical Review B*, 96(2), July 2017.

[225] R. Christensen, H. A. Hansen, and T. Vegge. Identifying systematic DFT errors in catalytic reactions. *Catalysis Science & Technology*, 5(11):4946–4949, 2015.

[226] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke. Finding density functionals with machine learning. *Physical Review Letters*, 108(25), June 2012.

[227] J. Hollingsworth, T. E. Baker, and K. Burke. Can exact conditions improve machine-learned density functionals? *The Journal of Chemical Physics*, 148(24):241743, June 2018.

[228] N. Mardirossian and M. Head-Gordon. $\omega$b97m-v: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of Chemical Physics*, 144(21):214110, June 2016.

[229] O. Schütt and J. VandeVondele. Machine learning adaptive basis sets for efficient large scale density functional theory simulation. *Journal of Chemical Theory and Computation*, 14(8):4168–4175, June 2018.

[230] X. Wang, M. Simard, and J. D. Wuest. Molecular tectonics. three-dimensional organic networks with zeolitic properties. *Journal of the American Chemical Society*, 116(26):12119–12120, December 1994.

[231] D. S. Wragg, R. E. Morris, and A. W. Burton. Pure silica zeolite-type frameworks: A structural analysis. *Chemistry of Materials*, 20(4):1561–1570, January 2008.

[232] L. Zoubritzky. Identification and Classification of Crystal Topologies. *Internship report*, École normale supérieure — PSL University, 2021.

[233] A. P. Bartók, R. Kondor, and G. Csányi. On representing chemical environments. *Phys. Rev. B*, 87(18):184115, May 2013.

[234] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18:13754–13769, 2016.

[235] E. A. Engel, A. Anelli, M. Ceriotti, C. J. Pickard, and R. J. Needs. Mapping uncharted territory in ice from zeolite networks to ice structures. *Nature Commun.*, 9:135701, 2018.

[236] T. C. Nicholas, A. L. Goodwin, and V. L. Deringer. Understanding the geometric diversity of inorganic and hybrid frameworks through structural coarse-graining. *Chem. Sci.*, 11(46):12580–12587, 2020.

[237] B. A. Helfrecht, R. Semino, G. Pireddu, S. M. Auerbach, and M. Ceriotti. A new kind of atlas of zeolite building blocks. *J. Chem. Phys.*, 151(15):154112, October 2019.

[238] J. Byggmästar, A. Hamedani, K. Nordlund, and F. Djurabekova. Machine-learning interatomic potential for radiation damage and defects in tungsten. *Phys. Rev. B*, 100(14):144105, October 2019.

[239] L. Himanen, M. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke, and A. S. Foster. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.*, 247:106949, 2020.

[240] Z. Chaker, M. Salanne, J.-M.Delaye, and T. Charpentier. Nmr shifts in aluminosilicate glasses *via* machine learning. *Phys. Chem. Chem. Phys.*, 21:21709–21725, 2019.

[241] J. Friedman. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 29(5):1189–1232, October 2001.

[242] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning - Data mining, inference and prediction*. Springer, New York, 2 edition, 2009.

[243] R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning - ICML '06*. ACM Press, 2006.

[244] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011.

[245] M. W. Deem, R. Pophale, P. A. Cheeseman, and D. J. Earl. Computational discovery of new zeolite-like materials. *J. Phys. Chem. C*, 113(51):21353–21360, October 2009.

[246] R. Pophale, P. A. Cheeseman, and M. W. Deem. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.*, 13(27):12407, 2011.

[247] O. T. Qazvini, R. Babarao, Z.-L. Shi, Y.-B. Zhang, and S. G. Telfer. A robust ethane-trapping metal–organic framework with a high capacity for ethylene purification. *Journal of the American Chemical Society*, 141(12):5014–5020, March 2019.

[248] O. T. Qazvini, V.-J. Scott, L. Bondorf, M. Ducamp, M. Hirscher, F.-X. Coudert, and S. G. Telfer. Flexibility of a metal–organic framework enhances gas separation and enables quantum sieving. *Chemistry of Materials*, 33(22):8886–8894, November 2021.

[249] A. U. Ortiz, A. Boutin, and F.-X. Coudert. Prediction of flexibility of metal–organic frameworks CAU-13 and NOTT-300 by first principles molecular simulations. *Chemical Communications*, 50(44):5867, 2014.

[250] H. Wu, T. Yildirim, and W. Zhou. Exceptional mechanical stability of highly porous zirconium metal–organic framework UiO-66 and its important implications. *The Journal of Physical Chemistry Letters*, 4(6):925–930, March 2013.

[251] T. D. Bennett, A. K. Cheetham, A. H. Fuchs, and F.-X. Coudert. Interplay between defects, disorder and flexibility in metal-organic frameworks. *Nature Chemistry*, 9(1):11–16, December 2016.

[252] S. M. J. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier, and V. Van Speybroeck. Thermodynamic insight in the high-pressure behavior of UiO-66: Effect of linker defects and linker expansion. *Chemistry of Materials*, 28(16):5721–5732, August 2016.

[253] L. R. Redfern, L. Robison, M. C. Wasson, S. Goswami, J. Lyu, T. Islamoglu, K. W. Chapman, and O. K. Farha. Porosity dependence of compression and lattice rigidity in metal–organic framework series. *Journal of the American Chemical Society*, 141(10):4365–4371, February 2019.

[254] J. Winarta, B. Shan, S. M. Mcintyre, L. Ye, C. Wang, J. Liu, and B. Mu. A decade of UiO-66 research: A historic review of dynamic structure, synthesis mechanisms, and characterization techniques of an archetypal metal–organic framework. *Crystal Growth & Design*, 20(2):1347–1362, December 2019.

[255] T. D. Bennett, J.-C. Tan, S. A. Moggach, R. Galvelis, C. Mellot-Draznieks, B. A. Reisner, A. Thirumurugan, D. R. Allan, and A. K. Cheetham. Mechanical properties of dense zeolitic imidazolate frameworks (ZIFs): A high-pressure x-ray diffraction, nanoindentation and computational study of the zinc framework zn(im)2, and its lithium–boron analogue, LiB(im)4. *Chemistry - A European Journal*, 16(35):10684–10690, August 2010.

[256] P. G. Yot, K. Yang, V. Guillerm, F. Ragon, V. Dmitriev, P. Parisiades, E. Elkaïm, T. Devic, P. Horcajada, C. Serre, N. Stock, J. P. S. Mowat, P. A. Wright, G. Férey, and G. Maurin. Impact of the metal centre and functionalization on the mechanical behaviour of MIL-53 metal–organic frameworks. *European Journal of Inorganic Chemistry*, 2016(27):4424–4429, July 2016.

[257] L. R. Redfern, M. Ducamp, M. C. Wasson, L. Robison, F. A. Son, F.-X. Coudert, and O. K. Farha. Isolating the role of the node-linker bond in the compression of UiO-66 metal–organic frameworks. *Chemistry of Materials*, 32(13):5864–5871, June 2020.

[258] M. Zhou, K. Wang, Z. Men, C. Sun, Z. Li, B. Liu, G. Zou, and B. Zou. Pressure-induced isostructural phase transition of a metal–organic framework co2(4, 4ı-bpy)3(NO3)4·xH2o. *CrystEngComm*, 16(20):4084–4087, 2014.

[259] R. Dovesi, R. Orlando, A. Erba, C. M. Zicovich-Wilson, B. Civalleri, S. Casassa, L. Maschio, M. Ferrabone, M. De La Pierre, P. D'Arco, Y. Noël, M. Causà, M. Rérat, and B. Kirtman. CRYSTAL14 : A program for the ab initio investigation of crystalline solids. *International Journal of Quantum Chemistry*, 114(19):1287–1317, March 2014.

[260] R. Dovesi and V. R. Saunders and C. Roetti and R. Orlando and C. M. Zicovich-Wilson and F. Pascale and B. Civalleri and K. Doll and N. M. Harrison and I. J. Bush and P. D'Arco and M. Llunell and M. Causà and Y. Noël. CRYSTAL14 User's Manual. University of Torino: Torino, 2014.

[261] C. L. Hobday, R. J. Marshall, C. F. Murphie, J. Sotelo, T. Richards, D. R. Allan, T. Düren, F.-X. Coudert, R. S. Forgan, C. A. Morrison, S. A. Moggach, and T. D. Bennett. A computational and experimental approach linking disorder, high-pressure behavior, and mechanical properties in UiO frameworks. *Angewandte Chemie International Edition*, 55(7):2401–2405, January 2016.

[262] M. Taddei, D. Tiana, N. Casati, J. A. van Bokhoven, B. Smit, and M. Ranocchiari. Mixed-linker UiO-66: structure–property relationships revealed by a combination of high-resolution powder x-ray diffraction and density functional theory calculations. *Physical Chemistry Chemical Physics*, 19(2):1551–1559, 2017.

[263] R. Gaillac, P. Pullumbi, K. A. Beyer, K. W. Chapman, D. A. Keen, T. D. Bennett, and F.-X. Coudert. Liquid metal–organic frameworks. *Nature Materials*, 16(11):1149–1154, October 2017.

[264] R. Gaillac, P. Pullumbi, and F.-X. Coudert. Melting of zeolitic imidazolate frameworks with different topologies: Insight from first-principles molecular dynamics. *The Journal of Physical Chemistry C*, 122(12):6730–6736, March 2018.

[265] R. N. Widmer, G. I. Lampronti, S. Anzellini, R. Gaillac, S. Farsang, C. Zhou, A. M. Belenguer, C. W. Wilson, H. Palmer, A. K. Kleppe, M. T. Wharmby, X. Yu, S. M. Cohen, S. G. Telfer, S. A. T. Redfern, F.-X. Coudert, S. G. MacLeod, and

T. D. Bennett. Pressure promoted low-temperature melting of metal–organic frameworks. *Nature Materials*, 18(4):370–376, March 2019.

[266] D. B. Mitzi. Introduction: Perovskites. *Chemical Reviews*, 119(5):3033–3035, March 2019.

[267] W. Li, Z. Wang, F. Deschler, S. Gao, R. H. Friend, and A. K. Cheetham. Chemically diverse and multifunctional hybrid organic–inorganic perovskites. *Nature Reviews Materials*, 2(3), February 2017.

[268] W. Li, A. Stroppa, S. Gao, and Z.-M. Wang. *Hybrid organic-inorganic perovskites*. Wiley, 2020.

[269] J. Burschka, N. Pellet, S.-J. Moon, R. Humphry-Baker, P. Gao, M. K. Nazeeruddin, and M. Grätzel. Sequential deposition as a route to high-performance perovskite-sensitized solar cells. *Nature*, 499(7458):316–319, July 2013.

[270] M. M. Lee, J. Teuscher, T. Miyasaka, T. N. Murakami, and H. J. Snaith. Efficient hybrid solar cells based on meso-superstructured organometal halide perovskites. *Science*, 338(6107):643–647, November 2012.

[271] Y. Wang, X. Lü, W. Yang, T. Wen, L. Yang, X. Ren, L. Wang, Z. Lin, and Y. Zhao. Pressure-induced phase transformation, reversible amorphization, and anomalous visible light response in organolead bromide perovskite. *Journal of the American Chemical Society*, 137(34):11144–11149, August 2015.

[272] T. Ou, J. Yan, C. Xiao, W. Shen, C. Liu, X. Liu, Y. Han, Y. Ma, and C. Gao. Visible light response, electrical transport, and amorphization in compressed organolead iodine perovskites. *Nanoscale*, 8(22):11426–11431, 2016.

[273] D. B. Mitzi. Synthesis, crystal structure, and optical and thermal properties of (c4h9nh3)2mi4 (m = ge, sn, pb). *Chemistry of Materials*, 8(3):791–800, January 1996.

[274] B. Kumar Shaw, A. R. Hughes, M. Ducamp, S. Moss, A. Debnath, A. F. Sapnik, M. F. Thorne, L. N. McHugh, A. Pugliese, D. S. Keeble, P. Chater, J. M. Bermudez-Garcia, X. Moya, S. K. Saha, D. A. Keen, F.-X. Coudert, F. Blanc, and T. D. Bennett. Melting of hybrid organic–inorganic perovskites. *Nature Chemistry*, 13(8):778–785, May 2021.

[275] J. M. Bermúdez-García, S. Yáñez-Vilar, A. García-Fernández, M. Sánchez-Andújar, S. Castro-García, J. Mira, J. A. Moreira, T. A. Centeno, and M. A. Señarís-Rodríguez. A simple in situ synthesis of magnetic m@CNTs by thermolysis of the hybrid perovskite [TPrA][m(dca)3]. *New Journal of Chemistry*, 41(8):3124–3133, 2017.

[276] 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. *Computer Physics Communications*, 167(2):103–128, April 2005.

[277] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, October 1996.

[278] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Physical Review B*, 54(3):1703–1710, July 1996.

[279] C. Chakravarty, P. G. Debenedetti, and F. H. Stillinger. Lindemann measures for the solid-liquid phase transition. *The Journal of Chemical Physics*, 126(20):204508, May 2007.

## RÉSUMÉ

Durant ma thèse, j'ai réalisé des simulations pour étudier les comportements "anormaux" présents dans certains matériaux. Je me suis particulièrement intéressé à l'expansion thermique négative, un comportement caractérisé par la contraction du matériau lors de son chauffage. J'ai choisi d'étudier les zéolithes, une famille de matériaux naturels existant également sous forme synthétique et largement utilisés en industrie. Les zéolithes sont connues pour avoir une expansion thermique négative, au point que ce comportement est considéré comme naturel. J'ai ainsi mené une étude systématique sur plus de 100 structures pour approfondir les connaissances sur cette propriété et sur sa présence dans les zéolithes. J'ai ensuite utilisé la base de données ainsi créée pour étudier son utilisation potentielle à travers le machine learning. Enfin, j'ai étudié plusieurs metal-organic frameworks lors de collaborations pour étudier divers phénomènes tels que la fusion ou l'ouverture de structures.

## MOTS CLÉS

Matériaux poreux, Zéolithes, Simulation moléculaire, Dynamique moléculaire

## ABSTRACT

During my PhD, I used molecular simulations techniques to investigate the "abnormal" behaviors present in different classes of material. Among the possible properties, I particularly focused on the negative thermal expansion, a behavior characterized by the contraction of the material upon heating. For this, I chose to study zeolites, a family of natural material existing also in synthetic form and widely used industrially. Zeolites are known for showing negative thermal expansion, to the point where this behavior is considered natural in this family. I thus conducted a systematic study of more than 100 structures to deepen the understanding of such property and its occurence in zeolites. I later used the created database to study its potential use in machine learning and the applicability to large databases of theoretical zeolites. I also work several metal-organic frameworks in collaboration with experiments to study various phenomena such as melting or opening of frameworks.

## KEYWORDS

Porous materials, zeolites, Molecular simulation, Molecular dynamics