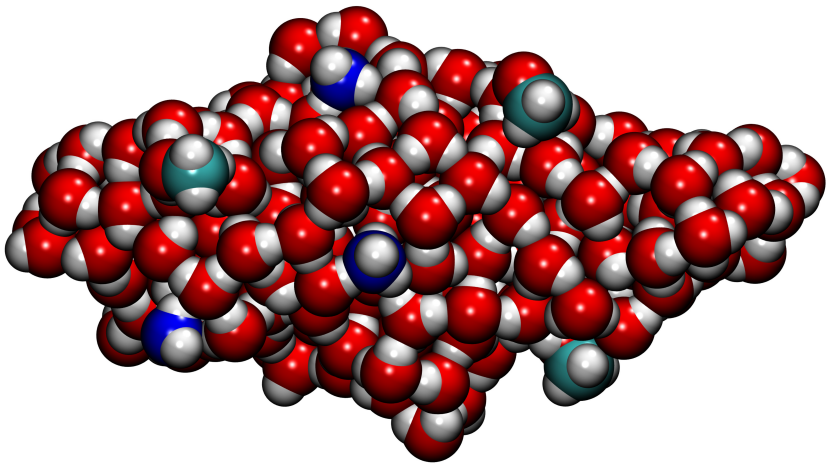

**Università degli Studi di Torino**
Doctoral School of the University of Torino
PhD Programme in Chemical and Materials Sciences XXXV Cycle

# Quantum chemistry computations of analogues of water amorphous ice covering interstellar grains, and binding energies of important interstellar molecules.



Supervisor:
Prof. Piero Ugliengo


**Università degli Studi di Torino**
Doctoral School of the University of Torino
PhD Programme in Chemical and Materials Sciences XXXV Cycle

# Quantum chemistry computations of analogues of water amorphous ice covering interstellar grains, and binding energies of important interstellar molecules.

Candidate: **Aurèle Germain**
Supervisor: Prof. **Piero Ugliengo**

Jury Members: Prof. **Stefan Bromley**
Universitat de Barcelona, Barcelona, Spain
Departament de Ciència de Materials i Química Física
& Institut de Química Teòrica i Computacional (IQTCUB)
Institució Catalana de Recerca i Estudis Avançats (ICREA),
Barcelona, Spain

Prof. **Francois Dulieu**
CY Cergy Paris University, Observatoire de Paris,
PSL University, Sorbonne Université,
CNRS, LERMA, Cergy, France

Prof. **Bartolomeo Civalleri**
Università degli Studi di Torino, Torino, Italy
Department of Chemistry

Head of the Doctoral School: Prof. Eleonora Bonifacio
PhD Programme Coordinator: Prof. Bartolome Civalleri

Torino, 2022

# Contents

# Preface

Astrochemistry is a multidisciplinary field mixing astronomical observation with chemical studies, both experimental and theoretical, in order to understand the chemical processes involved in the different astronomical objects present in the circumstellar medium (stars and their direct vicinity), and the interstellar medium (the space between the stars). In this context, my thesis is part of the "*Astro-Chemical Origins*" (ACO)[1] project, which is financed by the European Union's Horizon 2020 and Innovation Program, within the Innovative Training Networks (ITN) Marie Skłodowska-Curie Action (MSCA), grant agreement No 811312. ACO is a collaborative network of 13 European institutes and 8 partner organizations. As stated in the ACO website, the project has two main goals:

- to unveil the early history of the Solar System, using the chemical composition of today forming Solar-like planetary systems and comparing it with that of the Solar System primitive bodies;
- to train a new generation of researchers able to tackle this highly interdisciplinary problem, providing them with a wide-range of transferable skills, including the ability to communicate Science to a large audience.

[1] website: https://aco-itn.oapd.inaf.it/

These goals can only be achieved through an interdisciplinary collaboration among different communities. The project is divided into four scientific work packages (WP):

1. (Receivers) improve the detection capabilities of radio/millimeter receivers, in order to observe weak sources and low abundance molecules;

2. (Observations) build up a large and homogeneous database of astronomical spectral observations, in order to derive the chemical composition of forming Solar-like systems as a function of evolution and environment;

3. (Laboratory & Theory) understand the chemical processes in the exotic interstellar conditions via ad hoc quantum chemistry computations and laboratory experiments, in order to establish a reliable network of reactions occurring in the forming Solar-like systems;

4. (Model & Tools) develop sophisticated astrochemical models and tools, using modern techniques in order to predict the chemical composition of forming Solar-like planetary systems and compare it with observations, and finally build up a new model for the Proto Solar Nebula.

Each of the four objectives requires a specific expertise, possessed by different communities: instrumentation, astronomy, chemistry (experimental and theoretical) and computer science. The backbone of the project is constituted by 17 PhD students, who have carried out their thesis, each in different WPs.

Parallel to the scientific research, public engagement and outreach activities were carried out to increase the visibility and impact of astrochemistry among a broad non-scientific audience.

**Outreach activities** Within the ACO project, a part of this thesis was also focused on learning how to use tools dedicated to outreach activities and subsequently organize them with the other PhD students of the project. The main achievement was obtained with the development of a virtual reality (VR) project about astrochemistry and its presentation in science festival. This product is composed by several interactive and immersive activities using VR technology, which allows to introduce astrochemistry. The name of this project is "The journey of water to Earth" (See more in Appendix 5.6).

In the framework of outreach activity, Prof. Ugliengo, Lorenzo Tinacci, Marcus Keil (both PhD students of the ACO project), and I developed a

3D printed icy grain mantle model (Figure ). It allows to easily engage and explain to a broad audience the importance of the icy grain mantle in astrochemistry.

**Other activities** Several schools were also organized by the ACO project to increase our knowledge in fields not directly linked to ours, and to facilitate the exchanges between students of the ACO network. One international conference was organized in Turin on September 2021 (https://sites.google.com/inaf.it/aco-conference), and another is planned in Toulouse on June 2023 (https://aco-conference2023.sciencesconf.org).

As ACO students we also are assigned with "buddies", each ACO student is assigned to another ACO student. In my case my "buddy" is Stefano Ferrero, PhD student at the Autonomous University of Barcelona. As part of my thesis we each spent several months (6 for him and 3 for me) in our buddy's university.

All these activities were performed in addition to our obligations as PhD students, and not in replacement of it. In that regard, the amount of work needed to complete every task demanded by a project like ACO was sometimes overwhelming, even more so in the context of the COVID pandemic that hits only a few months after starting our PhDs, but ultimately very fulfilling. I have no doubt that the links created between every ACO students during the project activities will continue for years after the end of our respective thesis.

**Thesis context** In the ACO multidisciplinary environment, my thesis belongs to the Laboratory & Theory work package (WP3). Specifically, it focused on the modeling of interstellar icy grains mantle with a level of theory as close as possible to the best quantum mechanical approach, with an emphasis on developing a method of icy grain modeling that can be universally used by other astronomical theoretical groups in order to obtain more consistent results. The thesis was carried out at the Department of Chemistry of the Università degli studi di Torino (UniTo, Torino, Italy) under the supervision of Prof. Piero Ugliengo (theoretical and computational chemist) from November 2019 up to November 2022. Due to the COVID pandemic, a three months extension was authorized to all ACO PhD to get some extra time when finalizing the preparation of the thesis.

The thesis is organized as follows: after an introduction to several astro-

nomical concepts, regarding mainly the formation of chemical elements as well as dust grains and their icy mantles (Chapter 1), we will present briefly several computational methods and notions linked to this work (Chapter 2). This will be followed by a short review of the bibliographical work on grain models, with an emphasis on the ice mantle modeling, from both computation and experiments (Chapter 3). Finally, we will present the methods developed during this thesis to build up grain mantle (ACO-FROST python code, Chapter 4) to be used for studying adsorption of astrochemically relevant molecules at their surface, namely $NH_3$, $H_2O$, and $CH_3OH$ (Chapter 5).

# Chapter 1

# Introduction

# Abstract

In this chapter we will introduce the reader to several important aspects of the creation of matter in the universe, as well as the star formation and life cycle processes, directly linked to the production of heavy elements. Then we will present the medium of study of this thesis, the interstellar medium, known as the space between stars, and its different components. Finally we will describe the interstellar grains, submicrometric dust particles ubiquitous in the ISM, and their icy mantles formed in the coldest and densest parts of the ISM. We will introduce their formation, size, chemical complexity, and the processes happening on their surface. We hope that this chapter will contain the necessary information to understand fully the context and importance of the work presented in this thesis.


## 1.1 Introduction to astrochemistry

Astrochemistry is the study of abundance of molecules in the universe and their reactions. By combining observations of emitted astronomical particles with laboratory experiments and theoretical studies, astrochemists aim at understanding the chemical evolution of our universe, from its conception to present days. Through these studies, important questions regarding the formation and transformation of interstellar objects such as stars, meteorites, and planets can be answered. All these questions leading to the most important ones regarding the appearance of life on our planet and the formation of the necessary components *i.e.* water, complex organic molecules, amino acids, etc.

## 1.2 Short history of our universe

This section is mainly based and adapted from the books "Introduction to Cosmology" IInd edition by Barbara Ryden[1], and "Principles of Stellar Evolution and Nucleosynthesis" by Donald D. Clayton[2], as well as Ref.[3–5].

At the start of the 20th century, researchers observed that several galaxies visible from Earth were getting farther our own. By comparing observed absorption lines from molecular species in laboratories against the same emitted lines from nearby galaxies they found a shift in the measured wavelengths.

This shift, $z$, called "red-shift" when superior to 0, indicating a longer wavelength and hence an increase in distance from our galaxy, or "blue-shift" when inferior to 0, indicating a shorter wavelength and hence a decrease in distance from our galaxy, was found to be predominantly positive. Furthermore, the speed at which the galaxies were getting closer or farther from our own was much higher than the average speed of stars in our galaxy, eliminating the possibility of a simple observation of galaxies' respective gravitational influence. In 1927, Georges Lemaître, a Belgian cosmologist, after compiling the shift value of 42 galaxies and observing that 37 were positives, proposed, as an explanation to the over-representation of red-shifted galaxies and the speed at which they move away, that the universe could be expanding. By plotting the red-shift $z$ against the galaxies' distances $r$, and fitting them linearly, the astronomer Edwin Hubble derived the famous "Hubble's law" ($\nu = H_0 r$, where $\nu$ is called the recessional velocity, a product of $z$ and the speed of light $c$, and $H_0$ is the Hubble constant), showing that the farther a galaxy is from us, the faster it moves away. The Hubble constant, giving the speed of a galaxy 1 Mpc away (1 Mpc $= 3.09 \times 10^{22}$ m), was first computed to be approximately equal to 500 km $s^{-1}$ $Mpc^{-1}$, as of now, after observations from the *Hubble Space Telescope*, it is estimated to be equal to 68 $\pm$ 2 km $s^{-1}$ $Mpc^{-1}$.

The discovery of the universe's expansion in almost all directions brought forth two main models aimed at explaining this phenomenon. First, the "Big Bang" model, developed in the 20th century and widely accepted as of today, and second, the "Steady State" model, developed by Hermann Bondi, Thomas Gold, an Fred Hoyle in the 1940s, no more in use.

The "Big Bang" model describes the universe as having a start $t_0$ just after the big bang, and a defined quantity of matter that does not fluctuate. Thus, at a tiny fraction of seconds after the big bang, matter was condensed in a very hot and very dense state (T = $10^{32}$ K at $10^{-43}$ s). By expanding with time, its properties evolved and the universe became less dense and less hot, with galaxies getting farther away from each others. In this model, the Hubble constant $H_0$ evolves with time and is called the Hubble parameter for which $H_0$ is its current value in time.

In the "Steady State" model the *perfect cosmological principle* holds, which states that there is not a privilege moment in space and time, making the universe infinitely old and its properties constant. Thus, the Hubble constant remains constant with time and the length between galaxies tends toward 0 only when we look infinitely back in time. The density of the

universe stays also constant with its expansion, meaning that matters needs to be created, to maintain it, at an approximate rate of one hydrogen atom per $km^3$ per year.

In 1965, Arno Penzias and Robert Wilson discovered the "Cosmic microwave background" (CMB) that confirmed the superiority of the Big Bang model over the Steady State one. By using a microwave antenna they found an isotropic background of microwave radiations equivalent to black-body radiations. As of today, the CMB has been found to be well fitted to a black-body spectrum of approximately 2.7 K and involves about 411 photons per $cm^{-3}$. With the Big Bang model, unlike the Steady state model, the CMB is easily explained by the dense and hot state of the early universe containing highly interacting photons, scattered by the surrounding electrons, and possessing a black-body spectrum. By decreasing the density due to the universe's expansion, and other effects that will be discussed shortly, the black-body photons stopped scattering and flew freely across the universe.

### 1.2.1 The Big Bang model

The history of the universe, using the Big Bang model, can be divided in epochs from which three of them are highly important to understand the current state of the observable universe and the CMB.

Like said in the previous section, in its early state the universe was hot ($10^{27}$ K at $10^{-35}$ s, $10^9$ K at 100 s, 20000 K at $10^4$ yrs, 300 K at 0.3 Myrs) and dense. Matter and light were colliding frequently, and matter created from radiative recombination ( $p + e^- \rightarrow H + \gamma$) was almost directly victim of photonization ($H + \gamma \rightarrow p + e^-$) due to the highly energetic photons and their low mean free path. The universe's matter was ionized, and the free electrons of the universe made it opaque, equivalent to a black body. Photon and electrons were mainly interacting through scattering ($\gamma + e^- \rightarrow \gamma + e^-$) at a high enough rate to keep the temperature T of the universe constant via transfer of energy and momentum.

Following its expansion, the universe became less hot and less dense, decreasing the rate of photonization. The first epoch of interest, the *recombination* epoch, started when the universe went from ionized to neutral ($n_n = n_i$) due to the quicker creation of hydrogen and helium than their destruction through photonization. Through the decreasing of free electrons, the mean free path of scattering declined and became inferior to the Hubble parameter. Photons decoupled from matter and the universe became transparent,

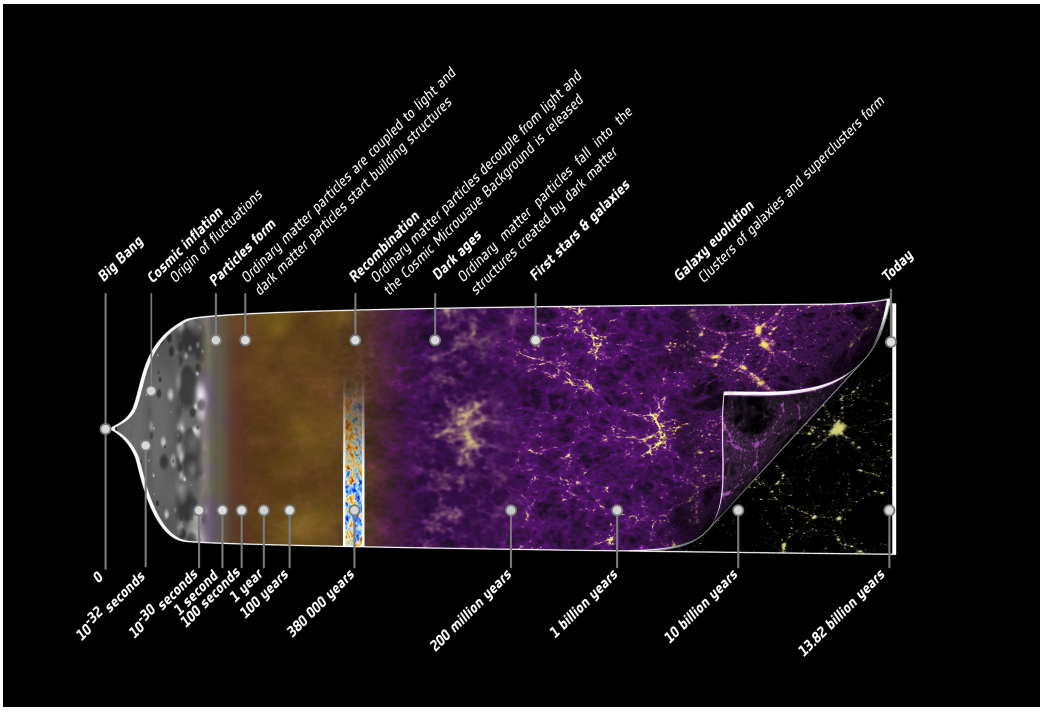



**Figure 1.1:** The 14-billion-year long history of the universe. © ESA – C. Carreau, taken from bit.ly/3UT4Y1h.

starting the second epoch of interest, the *photon decoupling* epoch. After some time, photons formed during the hot and dense phase of the universe stopped completely scattering with free electrons, starting the third epoch of interest, the *last scattering* epoch, and forming what is called now the CMB, a background radiation of non interacting photons dating back to the firsts hundred thousands of year of our 14 billion years old universe. The time scale between these different epochs, and other important events between the Big Bang and today, is shown on Figure 1.1.

### 1.2.2 Atomic history of the universe

In the first three minutes after the Big Bang, before the recombination epoch, light atoms were formed via fusion in what is called the epoch of Big Bang nucleosynthesis (BBN). These light atoms were mainly Hydrogen, Deuterium (D, also called heavy hydrogen, an stable isotope of the hydrogen atom with

one neutron and one proton), and Helium, but some Lithium (Li) was also present. As the temperature of the universe was too high ($10^{10}$ K), electrons could not be captured, making atoms ionized and short lived. With the expansion of the universe and the recombination epoch, the life expectancy of atoms became long enough to produce heavier ones, but the universe had cooled down too much to allow the fusion of atoms, and neutrons were either already decayed ($\tau_\nu \approx 15$ min) or bound into atoms of the BBN. Heavier elements than He and Li were not produced in the universe before the apparition of the first stars, around 100 Myr after the Big Bang.

#### 1.2.2.1 Primordial star formation

Stars are formed by the collapse of dense clouds of molecules, called dense molecular clouds, under their own gravitational force due to the overcoming of the thermal pressure inside the gas. In the present universe, molecular gases contain a vast amount of heavy elements allowing the cooling of the gas to temperatures as low as 10 K, but in the early universe, only light elements were present and thus gases could not cool below 100 K. To overcome the thermal pressure induced by the high temperature, an important mass of light elements needed to be present at the center of the gas. Hence, the chemical composition of the early universe encouraged the formation of stars heavier than the sun (of mass 1 $M_\odot = 2 \times 10^{30}$ kg). As larger stars evolve faster, the first formed stars were also short-lived (10 Byr for 1 $M_\odot$, and 6 Myr for 30 $M_\odot$). The critical mass needed to induce gravitational collapse at a pressure $P_{ext}$ is given by Ebert[6] and Bonnar[7]:

$$\begin{aligned} M_{BE} &= 1.18 M_\odot \frac{c_s^4}{G^{\frac{3}{2}}} P_{ext}^{-\frac{1}{2}} \\ c_s^2 &= \frac{dP}{d\rho} = \frac{\gamma k_B T}{\mu m_H} \end{aligned} \tag{1.1}$$

where T is the gas temperature, $\rho$ is the gas density, $m_H$ is the mass of a proton, G is the gravitational constant, $k_B$ is the boltzmann constant, $c_s$ is the speed of sound in the medium, and $\mu$ is the mean mass per particle in units of the proton mass. The collapse of the gas, seen as a free-fall of matter directed at its center, leads to its heating due to the release of potential energy. As the temperature increases, the pressure goes up and counteracts the gravitational collapse. Through chemical reactions such as three-body molecular hydrogen formation ($H + H + H \rightarrow H_2 + H$) the gas

cool down again and the gravitational collapse starts again. These steps of cooling followed by collapse and by subsequent cooling again continue until an equilibrium is found. When the gas is stabilized, a protostar is formed. The protostar continue to accrete matter contained in its envelope until the star is formed. The final mass of early star is still mostly unknown. As said before, a too heavy star would explode within several Myr (2 Myr for 100 $M_\odot$), and its accretion time would also be too important (approximately 5 Myr for 600 $M_\odot$) making its formation unlikely. A 30 $M_\odot$ is therefore a realistic assumption, and only a few Myr would be needed for the accretion of this amount of matter in the protostar.

The accretion of material by the protostar leads to more collapse and heating/cooling process. When equilibrium is reached and the temperature at the center of the star is high enough ($10^6$ K) fusion starts and the star nucleosynthesis can begin. The "fuel" present in the star's outer shell starts to "burn", meaning that the atoms present (H and He mainly) start to fuse and produce heavier elements: this is called the "Main sequence" of the star life-cycle. Hydrogen fuses into Helium, Helium fuses into Oxygen (O) and Carbon (C), and Carbon fuses into more varied elements. When the fuel is depleted (mainly H and He), the star dies.

#### 1.2.2.2 Formation of heavy elements

The formation of elements follows several chain reactions depending on the atom fueling the reactions. The creation of Helium through the burning of Hydrogen is done mainly via the PPI chain, also called proton-proton chain. It involves the creation of Deuterium through fusing of two hydrogen atoms and the subsequent fusing of D and H:

$$\begin{array}{lllllllll} \mathrm{H} & + & \mathrm{H} & \rightarrow & \mathrm{D} & + & \mathrm{e}^+ & + & \nu_e \\ \mathrm{D} & + & \mathrm{H} & \rightarrow & \mathrm{He}^3 & + & \gamma & & \\ \mathrm{He}^3 & + & \mathrm{He}^3 & \rightarrow & \mathrm{He}^4 & + & 2\mathrm{H} & & \end{array}$$

When $\mathrm{He}^4$ is formed in sufficient amount, and the temperature of the star reaches $10^8$ K, $\mathrm{C}^{12}$ can be produce via the $3\alpha$ reaction that fuses $\mathrm{He}^4$:

$$3\mathrm{He}^4 \rightarrow \mathrm{C}^{12} + \gamma$$

with an intermediate process forming an unstable Beryllium ($\mathrm{Be}^8$) by fusing two $\mathrm{He}^4$. Though $\mathrm{Be}^8$ is formed in an unstable condition, the rate of formation is such that an equilibrium forms where the rate of breakup equals the rate of formation and a small concentration of $\mathrm{Be}^8$ is built-up:

$$2\mathrm{He}^4 \quad \rightleftarrows \quad \mathrm{Be}^8$$

At temperatures of $10^8 - 10^9$ K, alpha captures involving $\mathrm{He}^4$ and $\mathrm{C}^{12}$ can occur, opening the doors to nucleosynthesis of more nuclei:

$$\begin{array}{lllllll} \mathrm{C}^{12} & + & \mathrm{He}^4 & \rightarrow & \mathrm{O}^{16} & + & \gamma \\ \mathrm{O}^{16} & + & \mathrm{He}^4 & \rightarrow & \mathrm{Ne}^{20} & + & \gamma \\ \mathrm{Ne}^{20} & + & \mathrm{He}^4 & \rightarrow & \mathrm{Mg}^{24} & + & \gamma \end{array}$$

and other nuclei such as $\mathrm{Si}^{28}$, $\mathrm{Ar}^{36}$,or $\mathrm{Fe}^{52}$. When the proportion of Carbon is high enough, it can start burning and produce elements directly through fusion:

$$\begin{array}{lllllll} \mathrm{C}^{12} & + & \mathrm{C}^{12} & \rightarrow & \mathrm{Mg}^{24} & + & \gamma \\ & & & \rightarrow & \mathrm{Mg}^{23} & + & \mathrm{n} \\ & & & \rightarrow & \mathrm{Na}^{23} & + & \mathrm{p} \\ & & & \rightarrow & \mathrm{Ne}^{20} & + & \mathrm{He}^4 \\ & & & \rightarrow & \mathrm{O}^{16} & + & 2\mathrm{He}^4 \end{array}$$

Similar reaction channels are available for $\mathrm{O}^{16}$ when its proportions allows it:

$$\begin{array}{lllllll} \mathrm{O}^{16} & + & \mathrm{O}^{16} & \rightarrow & \mathrm{S}^{32} & + & \gamma \\ & & & \rightarrow & \mathrm{P}^{31} & + & \mathrm{p} \\ & & & \rightarrow & \mathrm{S}^{31} & + & \mathrm{n} \\ & & & \rightarrow & \mathrm{Si}^{28} & + & \mathrm{He}^4 \\ & & & \rightarrow & \mathrm{Mg}^{24} & + & 2\mathrm{He}^4 \end{array}$$

As soon as $\mathrm{C}^{12}$ is formed through alpha capture, another cycle of nucleosynthesis starts, called carbon-nitrogen-oxygen (CNO):

$$\mathrm{C}^{12} \xrightarrow{(p,\gamma)} \mathrm{N}^{13} \xrightarrow{(e^+)} \mathrm{C}^{13} \xrightarrow{(p,\gamma)} \mathrm{N}^{14} \xrightarrow{(p,\gamma)} \mathrm{O}^{15} \xrightarrow{((e^+))} \mathrm{N}^{15} \xrightarrow{(p,\mathrm{He}^4)} \mathrm{C}^{12}$$

The CNO cycle is, at first, only a fraction of the energy produced by stars, but as their age increases, its fraction of produced energy increases.

Fusions processes continue during the burning stages of stars and, with the last stage (called the Silicon burning stage), produce heavier elements as long as the binding energy of the formed nucleus is superior to the binding energy of the constituents. Figure 1.2 shows the binding energy per nucleon of several elements, the Fe group (primarily Cr, Mn, Fe, Co, and Ni) being the peak and the heavier elements formed in stars through thermal fusion.

These elements are produced via the photodesintegration of lighter, previously formed, nuclei, producing alpha particles ($He^4$), neutrons and protons. The non-dissociated nuclei capture the alpha particles to produce new nuclei, and in the case of Si to produce heavier nuclei up to Fe and Ni:

$$
\begin{array}{ccccccc}
Si^{28} & + & He^4 & \rightarrow & S^{32} & + & \gamma \\
S^{32} & + & He^4 & \rightarrow & Ar^{36} & + & \gamma \\
Ar^{36} & + & He^4 & \rightarrow & Ca^{40} & + & \gamma \\
Ca^{40} & + & He^4 & \rightarrow & Ti^{44} & + & \gamma \\
Ti^{44} & + & He^4 & \rightarrow & Cr^{48} & + & \gamma \\
Cr^{48} & + & He^4 & \rightarrow & Fe^{52} & + & \gamma \\
Fe^{52} & + & He^4 & \rightarrow & Ni^{56} & + & \gamma
\end{array}
$$

Other processes are occurring but production of heavier elements than the Fe group cannot be done via fusion but are possible through fission occurring during the explosion of massive stars *i.e.* supernovae.

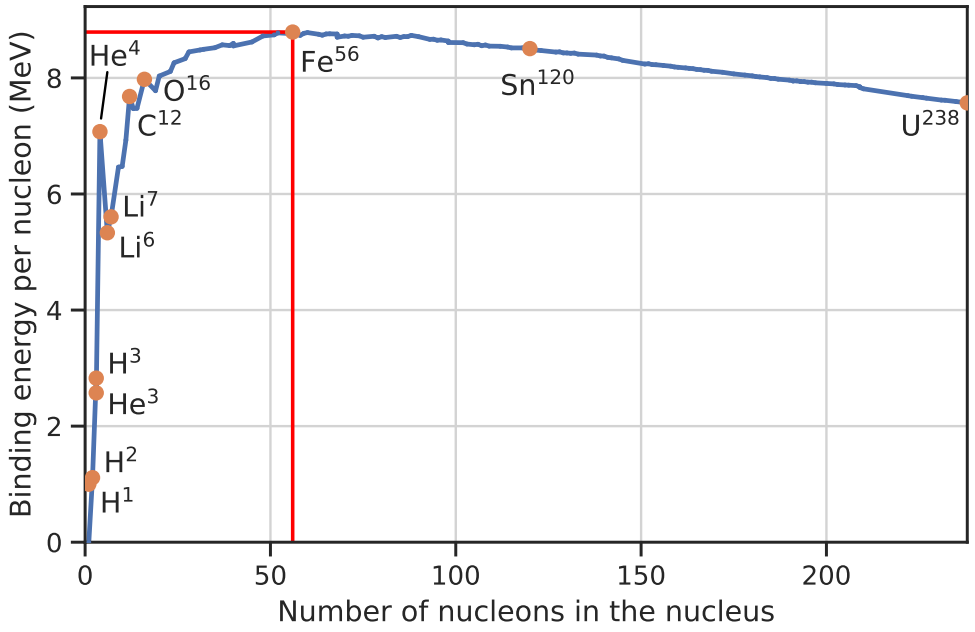



**Figure 1.2:** Binding energy (in MeV) per nucleon against the number of nucleons (mass number). Red lines indicate $Fe^{56}$. Data taken from the Atomic Mass Data Center: https://www.nndc.bnl.gov/nudat3/.

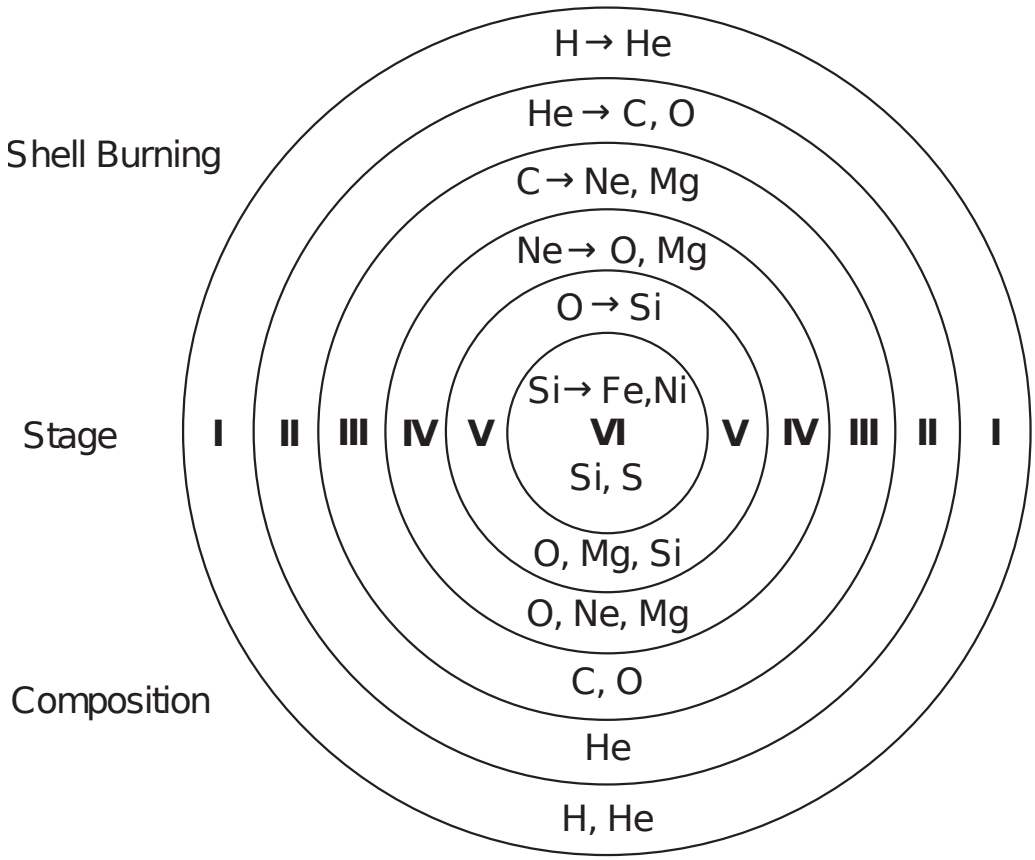



**Figure 1.3:** Schematic representation of the different burning stages and some elements produced during them. Taken from *Fusion The energy of the universe* Chapter 3[8].

The star goes from burning H, to He, to C, and to other elements forming successive burning stages (see Figure 1.3 for a graphical representation of the burning stages). The star burns slowly the fuel in its shells until hydrogen is depleted. The star stops generating enough energy and the gas pressure is insufficient to support the core. Through gravitational contraction, and subsequent temperature increase, the outer layers expand. The star becomes redder, and is therefore called a Red supergiant, and continues to burn He. When He is exhausted, the same process occurs again. As the star is massive, it is not able to reduce its mass to a sufficient low value and ends its life catastrophically through a supernova explosion. The explosion scatters

the produced elements in the gas phase of the surrounding environment, populating it with heavier atoms. The cycle can continue and another dense molecular cloud, now containing such elements as Carbon and Oxygen atoms through the scattering produced by the supernovae explosion, can undergo the same transformation.

#### 1.2.2.3 Present day star formation

The cooling of primordial gas was highly inefficient after 100 K due to the presence of only light elements produced during the early expansion of the universe after the Big Bang. With the death of first stars, the universe was populated with heavier elements such as oxygen and carbon, allowing a better cooling of molecular gas clouds in the star forming regions. Through fine-structure line cooling of ionized carbon or neutral atomic oxygen, the molecular clouds are able to reach typical temperatures of 10-20 K. Hyperfine structure transition like the $^2P_{3/2} - ^2P_{1/2}$ [C II] line (158 $\mu$m) or the $^3P_1 - ^3P_2$ [O I] line (63 $\mu$m) are believed to be the dominant cooling lines in molecular clouds. At these temperatures, the mass necessary to overcome the thermal pressure and form a protostar, shown in Eq. 1.1, is much lower than the one needed for primordial star formation. Stars formed in regions rich in elements with Z>2 are therefore favorably lighter than primordial ones, with masses closer to 1 $M_\odot$ than 30 or even 100 $M_\odot$.

The formed stars use gas containing heavier elements than H and He, but these light components are still the predominant ones present in the gas and the subsequent star formed. Therefore, the same processes are occurring in their layers, except if the formed star has a mass inferior to 0.08 $M_\odot$. In such cases, the formed star is called a "Brown Dwarf, a type of star that is not able to perform fusion of H atoms. After the main sequence of the low-mass stars, the same process as for massive stars occur, its outer layer becomes redder and its size increases to become a "Red Giant". Again, the star's life can end catastrophically in a supernova, if orbited by a white dwarf (a very dense and small star) and thus part of a binary system, or transform into a "Planetary Nebula", a region containing ionized elements formed during the star's life-cycle.

## 1.3 The Interstellar Medium

This section is mainly based and adapted from the books "Physics and chemistry of the Interstellar Medium" by Sun Kwok[9] and "Astrochemistry: from Astronomy to Astroiology" by Andrew M. Shaw[10], as well as Ref.[11,12].

The dust left around the formed stars becomes, over time, planets; planets revolving around a star form a solar system of roughly $10^{-2}$ light-years across; a group of solar systems, approximately $10^8 - 10^9$, form a galaxy of about $10^3 - 10^5$ ly in size; a group of galaxies form a cluster, then a super-cluster, and finally the cosmic-web. Space is, thus, not empty, but filled with stars and matter in-between these stars. Yet, the night-sky is dark and therefore seems to be void of anything outside of these objects. As the baryonic matter (the matter made of protons and neutrons, like atoms) is only 5% of the total matter of the universe (27% for dark matter, and 68% for dark energy) we know this to be untrue. It is also untrue if we concentrate only on the baryonic fraction of the universe. The Interstellar Medium (ISM) is defined as the space between stars and amount for 10% of the baryonic matter (7% for the stars). It is filled with gas, dense and cold, that obfuscate lights coming from nearby galaxies. Figure 1.4 shows a near-infrared image of a dark cloud, discovered in the 1930s and interpreted as a "hole" between stars, but now known to be a dark cloud of cold and dense particles adsorbing light.

The matter forming these dense, and sometimes diffuse, clouds are, like said in the previous section, incorporated through the catastrophic death of stars as well as solar winds ejecting atoms during their life-span. Stars change the chemical composition of their environment, therefore, the abundance of species in the ISM varies depending on its solar and chemical history, changing its composition between galaxies and even inside galaxies. Studying the chemical evolution of the ISM and its different parts is important for understanding stellar evolution and the chemical complexity of the universe and our solar system.

Table 1.1 shows the relative chemical abundance of several atoms in the present day sun, and therefore in its vicinity. As showed, hydrogen, in all its form, is the most abundant element, representing more than 90% of the baryionic matter, while He, the second most abundant one, represents less than 10%. After these two atoms, sole products, with Li, of the primordial matter formed during the big bang, the abundance drops drastically.

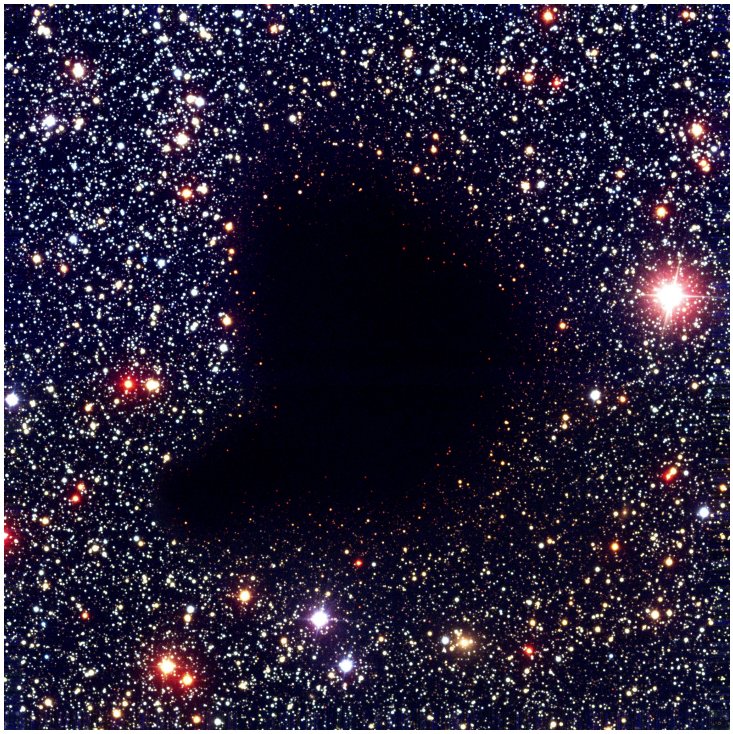


**Figure 1.4:** Near-infrared image of the dark cloud Barnard 68 discovered in 1919 by Edward Emerson Barnard. First cataloged in 1927 as a "dark nebulae" it is know known as a dark cloud of 16 K of radius 0.25 ly and a mass of about 2 $M_{\odot}$. Image taken from ESO https://www.eso.org/public/images/eso0102a/.

### 1.3.1 Interstellar molecules

The universe is not only filled with atomic elements such as hydrogen and helium, but also of molecules with various complexities. Interstellar molecules were first discovered in the 1930s with the detection of CH, $CH^+$, and CN[9,14–16]. Through the 1960s and 1970s and the development of radio receiver for both centimeter and millimeter wavelength, other interstellar molecules were detected: OH[17], CO[18], $H_2O$[19], $NH_3$[20], $H_2CO$[21], HCN[22], etc. As of today, 241 individual molecular species with 19 different elements were detected in the ISM and around stars[23], with a vast increase in the last decade. Even more spectral features detected by astronomers are still unidentified due to the lack of relevant theoretical and experimental studies.

**Table 1.1:** Relative abundance of the elements most present in the sun and its vicinity[13].

| Element | Abundance/H |
|---|---|
| Hydrogen | 1.0 |
| Helium | 0.0851 |
| Oxygen | $4.9 \times 10^{-4}$ |
| Carbon | $2.69 \times 10^{-4}$ |
| Nitrogen | $6.76 \times 10^{-5}$ |
| Magnesium | $3.98 \times 10^{-5}$ |
| Silicon | $3.24 \times 10^{-5}$ |
| Iron | $3.16 \times 10^{-5}$ |
| Sulfur | $1.32 \times 10^{-5}$ |

Molecular hydrogen ($H_2$) is the most abundant molecule, while formaldehyde ($H_2CO$) and hydrogen cyanide (HCN) are most abundant polyatomic species in the gas phase. The detection of interstellar molecules is an important part of understanding processes happening in the ISM. Molecules in clouds, like CO and $H_2O$, are effective for cooling interstellar gas via excitation through collision, mainly with $H_2$, and desexcitation via photon emission. Furthermore, presence of molecules in the ISM are of biological interest. HCN can turn into HCN polymers that, in turn, can react with water to form precursors of proteins and nucleic acids. When it comes to $NH_3$ and $H_2CO$, these two observed molecules are of prebiotic importance and could form amino acids.

### 1.3.2 The different components of the ISM

The ISM is composed mainly of gas that accounts for 99% of its mass, the remaining 1% is present in the form of dust particles. It is also composed of different phases that depend on the ionization status of H, their density and temperature. While the cold phases contain the most part of the mass of the ISM, the hot and ionized phases occupy most the space. Contrary to the always ionized and hot state of matter inside stars, the ISM contains a wide variety of state of matter and temperature conditions. Five phases of the ISM can be distinguished: the coronal gas, the diffuse ionized gas,

the intercloud H I, the diffuse clouds, the molecular/dark clouds, and the protostellar cores. H I designates regions of the ISM with hydrogen atoms predominantly neutral, while H II represents regions of the ISM with hydrogen atoms predominantly ionized.

The coronal gas is a very hot ($10^6$ K) gaseous object of the ISM that emits light. Its high temperature is believed to come either from mechanical heating by shock waves, or from bubbles of hot gas produce during supernova explosions. Although coronal gas are too hot or diffuse to form a star, they can expand via merging of another bubble of coronal gas, compressing the nearby molecular clouds and possibly starting the star formation process. As the hydrogen present in the coronal gas is completely ionized, it is classified as an H II region.

Diffuse ionized gas (DIG), also called warm ionized medium (WIM), is a widespread, low density, region of the ISM made mostly of ionized hydrogen (H II region) of high temperatures around $10^4$ K. Radiations produced by the youngest and most luminous stars (called O stars) of the galaxy, present in isolated regions of star formation and often surrounded by opaque clouds of neutral hydrogen, are believed to be the source of the extensive ionization of the DIG.[24]

The ISM contains interstellar gas clouds that account for half of its mass but very little of its volume. The remaining mass is scattered between those clouds in what is called "intercloud gas". As it is located between molecular clouds and ionized regions its density is very low (0.1 $cm^{-3}$). It is mainly constituted of neutral hydrogen atoms and is therefore an H I region.

Diffuse clouds are H I regions with temperatures around 100K and densities between $10 - 100$ $cm^{-3}$. Their diffuse nature makes them mostly semi-transparent and therefore few molecules can form. Nevertheless, several molecules like $H_2$ or CO have been detected, probably forming in the inner parts of the clouds, protected by the ionizing stars emitted photons.

Molecular clouds and dark clouds are neutral parts (H I regions) of the ISM with low temperatures around 50 K, suggesting no or very weak heating source, and high density. Dark clouds, like the Barnard 68 cloud shown in Fig. 1.4, are characterized by the absence of visible light due to its high density ($10^5$ $cm^{-3}$) of particles of interstellar dust, absorbing light. Molecular clouds and dark clouds are gravitionally unstable, their collapse leading to the formation of protostellar cores subsequently evolving to stars.

Protostellar cores, formed by the collapsing of giant molecular clouds as previously described, are the first stage of high-mass star formation. They

**Table 1.2:** The multiple phases of the ISM[9].

| ISM component | Designation | T (K) | Density[a] ($cm^{-3}$) | Fraction of ionization |
|---|---|---|---|---|
| Hot ionized medium (HIM) | Coronal gas | $10^6$ | 0.003 | 1 |
| Warm ionized medium (WIM) | Diffuse ionized gas (DIG) | $10^4$ | $> 10$ | 1 |
| Warm neutral medium (WNM) | Intercloud H I | $10^3 - 10^4$ | 0.1 | 0.1 |
| Atomic cold neutral medium | Diffuse clouds | 100 | $10 - 100$ | $\approx 10^{-3}$ |
| Molecular cold neutral medium | Molecular clouds, Dark clouds | $0 - 50$ | $10^3 - 10^5$ | $10^{-7} - 10^{-8}$ |
| Molecular hot cores | Protostellar cores | $100 - 300$ | $> 10^6$ | $\leq 10^{-8}$ |

[a] Density of $H_2$ in molecular clouds and cores, otherwise density of H.

are highly dense ($> 10^6$ $cm^{-3}$) with temperatures ranging from 100 to 300K. It is a neutral region and therefore classified as H I.

A summary of key properties of the different phases of the ISM are shown in Table 1.2.

#### 1.3.2.1 Ionization and heating processes

The ionization of atoms in the gas can happen in different ways. The most common method of ionization is from photons emitted from the sun. Energies $> 13.6$ eV can ionize hydrogen atoms while less energetic photons can ionize less abundant elements such as carbon atoms. Cosmic rays made of proton (87%) or He (12%) can also ionize the gas. High velocity shocks can produce a photon that will subsequently ionize an atom, or ionize it directly if the energy transferred during the shock is high enough.

The heating of the gas in the ISM has also different ways of happening. Photons of the CMD can heat atoms as their distribution is similar to black body radiations at 2.7 K. Photons emitted from stars can heat the gas through the "photo-electric" heating phenomena where a photon of energy $\approx 10$ eV can eject an electron from the dust via its absorption; the ejected

$e^-$ can then heat the gas. The "photo-electric" heating phenomena is the dominant form of heating in parts of the ISM. Electrons can also be ejected from H and $H_2$ via cosmic ray ionization. Collision between gas particles can heat and cool the gas through energy transfer. The dust can also cool the gas through black body emission, and emission from molecules' transition.

### 1.3.3 Interstellar grains

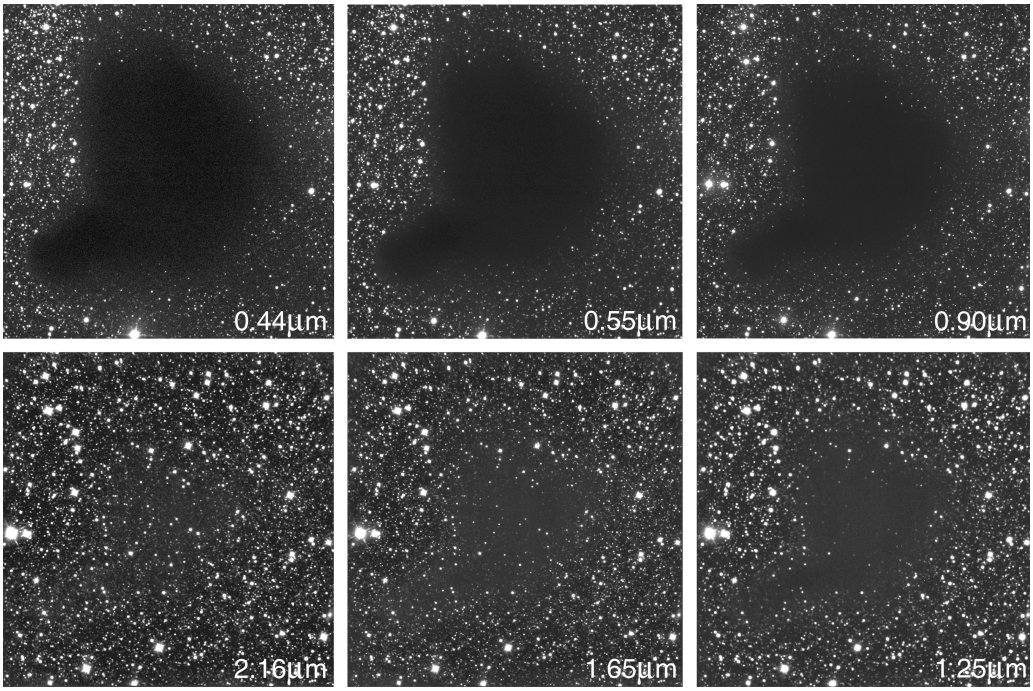



**Figure 1.5:** Image of the dark cloud Barnard 68 (also shown in Fig. 1.4) taken in six different wavelength. By increasing the wavelength the interstellar extinction is reduced, making the background visible. Image taken from ESO https://www.eso.org/public/images/eso9934b/.

Parts of the night sky are obscured by dark spots (see Fig. 1.4), preventing starlight from reaching us. Before the 20th century the explanation for these "holes" in the sky was unknown. Some astronomers leaned toward a lack of star distribution in parts of the universe, others, more numerous, were in support of interstellar objects obscuring the starlight from these spots and therefore causing interstellar extinction. In the 1930s, Robert Trumpler

showed that interstellar extinction is dependent on the wavelength. Blue light is more extinguished than red light, making the starlight more "reddened". By computing the amount of reddening linked to the starlight passing through the ISM, it was possible to obtain an estimate of the quantity of obscuring material.

In the middle of the 20th century the presence of interstellar dust as explanation for interstellar extinction was widely accepted, and the observed continuum emission was seen as dust grains heated by starlight and radiating thermally. In the 1960s infrared detectors were made widely available and observations showed that the visual extinction dependence on wavelength was continuing into the infrared, but that extinction was much less present in this region. By observing using infrared detectors, background light of dark clouds were not as obscured anymore, and stars were visible (see Figure 1.5).

#### 1.3.3.1 Chemical composition of ISM grains

Through IR observation and laboratory experiments the interstellar grains are believed to be made mostly of silicate and carbonaceous material. In the diffuse medium the silicate is assumed to be amorphous as in protoplanetary disks, a vast accretion of gas particles around forming stars, where it is mostly rich in magnesium and therefore present in the fosterite form ($Mg_2SiO_4$).

In hotter parts of the ISM ($> 100$ K) interstellar grains are bare, but in colder parts ($< 100$ K) a thin layer of ice starts to form around the core. The ice is mostly made of water, not coming to the dust already formed from the gas-phase, but synthesized through *in situ* recombination of oxygen and hydrogen atoms on the grain surface. Nitrogen and carbon atoms are also adsorbed on the grain and combined with hydrogen to form ammonia ($NH_3$) and methane ($CH_4$). In dark molecular clouds, where the temperature can reach a few dozens of K, the shielding of cosmic rays and energetic photons by the surrounding dust allows the accumulation of other molecules not formed at higher temperatures. A small part of the newly formed carbon monoxide contained in dark clouds, the second most abundant molecule present, starts to get adsorbed on the amorphous water surface (AWS) of grains. In the $\approx 10$ K zone of dense clouds, CO freeze-out at a level up to 90%[25–28], disappearing almost entirely from the gas phase of dark clouds. At this temperature, the gas-phase of dark clouds is mostly depleted of reactive species other than hydrogen and deuterium, making them the first source of reactive species adsorbed on the grain. Therefore, molecules like

CO are hydrogenated into formaldehyde ($CH_2O$) and methanol ($CH_3OH$) and deuterated. The process of formation of amorphous ice mantles on top of an interstellar dust grain is illustrated in Figure 1.6.

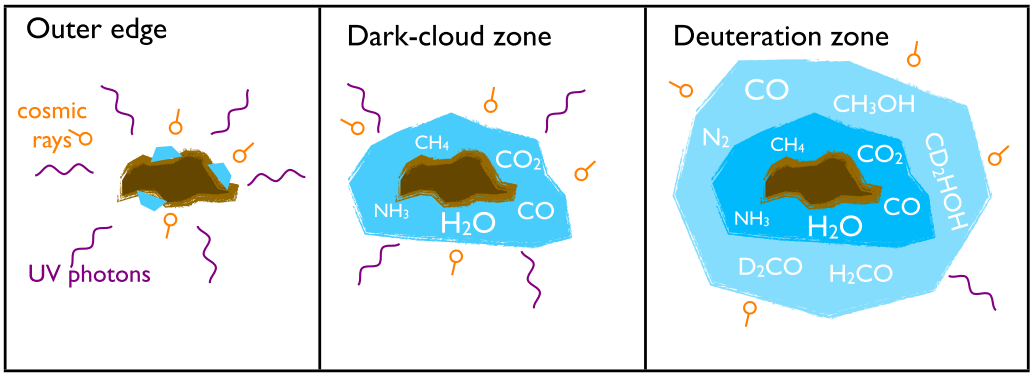



**Figure 1.6:** Evolution of the ice mantle from the outer-edge of a prestellar core ($100-300$ K) to the deuteration zone of a dark cloud ($\approx 10$ K). Adapted from Caselli et al.[12]

As of today, dozens of molecules have been detected in the mantles of interstellar icy grains by correlating observational data with laboratories' ones. After $H_2O$, the main component of ice mantles, $NH_3$, $CH_4$, and CO, that we already mentioned, astronomers have detected molecules such as $CO_2$, OCS, HCOOH, $CH_2O$, $OCN^-$, $C_2H_5OH$, $C_2H_6$[29–35] to name a few. A typical IR spectrum of ices in an interstellar object is shown in Figure 1.7. The different spectral features a labeled with the molecules they were identified with after laboratory studies. From these IR spectra, one can derive the "column density" of the observed atoms and compute their abundance in the observed objects. Table 1.3 shows the abundances of several molecules identified in interstellar ices, from different sources. The abundances are presented relatively to water, and for all sources we can see that the main constituent of icy mantles, after $H_2O$, are mostly CO, $CO_2$, $CH_3OH$, and $NH_3$.

#### 1.3.3.2 ISM grain formation

Interstellar grains are not directly produced in the ISM, as the temperature and density of matter needed for nucleation of dust grains are not elevated enough, but in the surrounding gas of stars called the Circumstellar medium (CSM). After ejection from the CSM, the different processes of the ISM

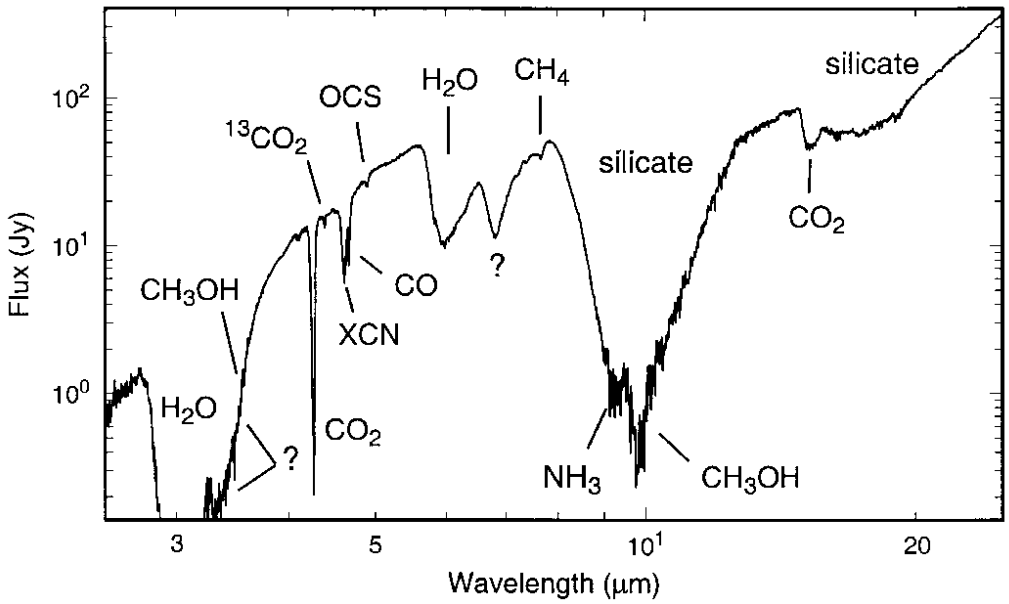



**Figure 1.7:** IR spectrum of the W33A protostar. The main absorption features of chemical components frozen into the ice mantles are indicated on the spectrum. By identifying these absorption lines, astronomers can identify the constituents of the interstellar ices. Reproduced from Gibb et al[31], by permission of the AAS.

evolve the produced interstellar dust and change its chemical composition and physical structure.

The composition of interstellar dust depends on the chemical abundance of the solar atmosphere from which the grain can be assembled. Carbon/oxygen rich star atmospheres would make carbon/silicates dusts. As the conditions for interstellar dust formation are quite delicate, only few types of stars are able to contribute to their formation, *i.e.* stars with cool stellar envelope like red giants and super giants.

Table 1.4 shows the main contributors to interstellar dust, as well as gas, in our galaxy. Asymptotic giant branch (AGB) stars are the main generator of interstellar dust. AGB are stars around 10 $M_\odot$ at the end of their life as a good portion and their hydrogen and helium has already been burned. Higher stages of burning are well underway and thus they contain lots of heavy elements needed for interstellar dust grains production (carbon, oxy-

**Table 1.3:** Abundances, relative to water, of several constituents of interstellar ices in different astronomical objects.

| Species (%) | W33A[a] | RAFGL 7009S[b] | NGC 7538 IRS9[c] | AFGL 961[d] |
|---|---|---|---|---|
| $H_2O$ | 100 | 100 | 100 | 100 |
| CO | 8 | 16 | 16 | 6 |
| $CO_2$ | 13 | 23 | 12 | 3 |
| $CH_3OH$ | 18 | $4-7$ | 65 | 87 |
| $CH_4$ | 2 | 4 | $< 1$ | $< 1.5$ |
| $NH_3$ | 15 | 17 | $< 10$ | $< 10$ |
| OCS | 0.2 | $\sim 0.2$ | $< 0.1$ | $< 0.1$ |
| $OCN^-$ | $2-8$ | 4 | 1 | – |

[a] Gibb et al. (2000)[31], and Demyk et al. (1998)[36].
[b] D'Hendecourt et al. (1996)[37], Dartois et al. (1999)[38], Brooke et al. (1999)[39], Ehrenfreund et al (1997)[40], Demyk et al. (1998)[36], and Gibb et al. (2004)[34].
[c] Tielens et al. (1987)[41], Whittet et al. (1996)[42], and Gibb et al. (2004)[34].
[d] Tielens et al. (1987)[41], D'Hendecourt et al. (1989)[43],and Palumbo et al. (1997)[44].

gen, magnesium, silicon, etc.). These products are ejected periodically in the stars' surroundings, and dust may nucleate and grow in this ejecta. After some time, the gas drift away from the stars, spreading the dust in the ISM. Compared to AGB, other massive stars produce only a very small proportion of dust grains. The amount and nature of the dust depends on stellar mass, metallicity (amount of element with Z> 2), and not least on the photospheric C/O ratio. For C/O < 1 (M-type AGB stars) the circumstellar chemistry is oxygen-dominated, and the type of dust that is forming is made of silicates, alumina ($Al_2O_3$), and other metal oxides[45]. The most evolved stars with an important rate of mass-loss show silicate features while the less evolved one with, with low mass-loss, show alumina features. In C-rich stars with C/O > 1, carbonaceous molecules prevail and condensate such as amorphous carbon and silicon carbide (SiC) constitute the dust grains[46]. About 90% of SiC grains are thought to come from C-rich low-mass ($\sim 1.5$ to $\sim 3.5$ $M_\odot$) AGB stars with metallicity close to the solar one[47]. Shocks from the star facilitate the formation of molecules in its vicinity. The chemistry in the dense gas surrounding the ABGs is controlled by the C/O ratio. In C-rich star, oxygen atoms are trapped in CO that dissociates following shocks, resulting in the

**Table 1.4:** Main contributors to gas and dust in the Milky Way. Estimated injection rates[a] are in $M_\odot$ $pc^{-2}$ $Myr^{-1}$.

| Source | Gas | Carbon dust | Silicate dust |
|---|---|---|---|
| AGB (C-rich) | 750 | 3 | |
| AGB (O-rich) | 750 | | 5 |
| Other massive stars | 150 | 0.1 | 0.03 |
| Novae | 6 | 0.3 | 0.03 |
| SN type Ia | | 0.3 | 2 |
| SN type II | 100 | 2 | 10 |

[a] Taken from The chemistry of Cosmic Dust Williams D.A. and Cecchi-Pestellini C.[11], from data obtained from Tielens et al. 2005[48], Massey et al. 2005[49], and Ferrarotti et al. 2006[50].

release of O atoms and the formation of O-bearing species like water. In O-rich stars, the opposite happens and C-bearing species form.

Ejected matter from Novae and supernovae (SN) explosions are an important place of dust formation in the CSM, and probably the only source of dust grains in the early universe. Novae and SN type Ia are types of explosion happening in systems containing binary stars, systems with two stellar objects that orbit each other, with one star being a white-dwarf (WD). While novea explosions happen periodically and do not destroy the WD, SN type 1a do destroy the WD. SN type II are supernovae explosions caused by the collapse of massive stars and are one of the main source of silicate dust in the Milky Way.

After grains are ejected from the CSM, and enter the ISM, they start to be processed by the surrounding environments, changing their size, shape, and composition. The accumulation of atoms and molecules in dark cloud to form amorphous ice mantles is one of these processes, but more exist.

Even though the concentration of dust grains in the ISM is very low, grains can grow through coagulation. This process can lead to, in extreme occasions, the formation of planetesimals a few kilometers in size, and then planet-size objects in star forming regions. The collision of dust grains can lead to their shattering and even vaporization, the release of the dust components into the gas phase, for the smallest ones if the impact energy is high enough. This process can lead to the disappearance of smaller dust grains.

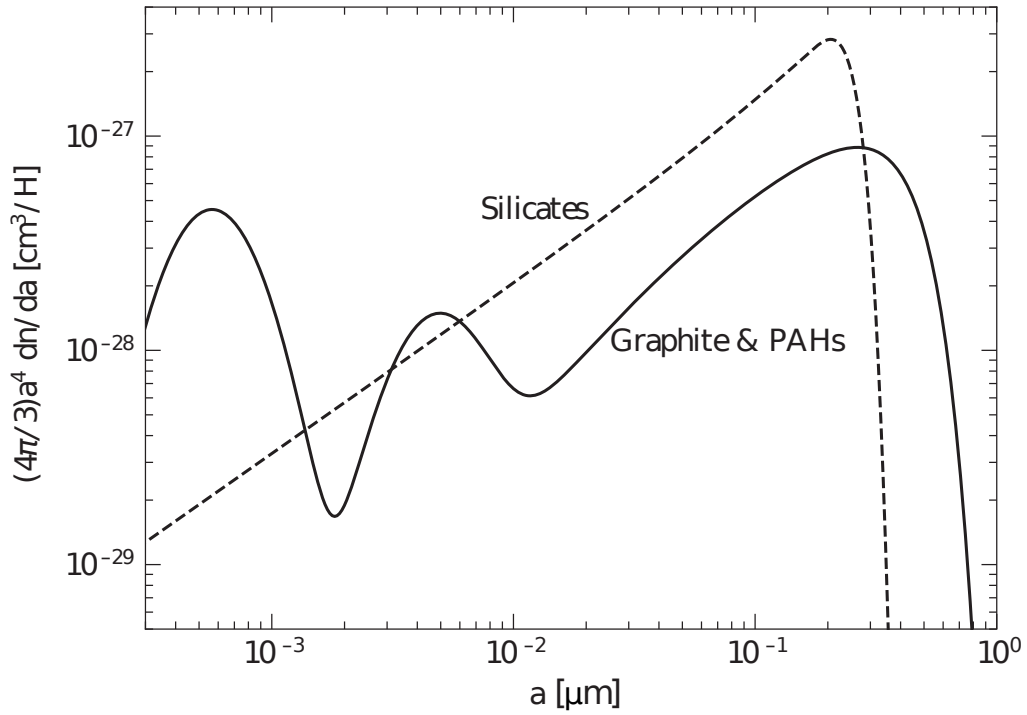



**Figure 1.8:** Distribution of dust grain sizes (expressed as the volume of dust) in function of their radius a in logarithmic scale. Taken from *Interstellar and Intergalactic Medium*[51] and using the Weingartner & Draine[52] (WD) model.

Changes in temperatures in different regions of the ISM can lead to thermal annealing and therefore crystallization of silicates. Reciprocally, cosmic rays like low energy ions ($He^+$), can amorphize crystalline silicates.

The present knowledge of dust grains properties is deduced from astronomical observations of absorption and diffusion of light on their surface and from models derived from these observations. Through several parameters governing the models, other properties that cannot be directly observed can be predicted. As models lead to predicted properties not directly observable and directly observable, its prediction can be validated by looking at the validity of the directly observable predicted properties (like interstellar extinction of light and its polarization).

One of these properties not directly observable is the size of interstellar dust grains. As we said before, interstellar grains are processed in the ISM and therefore are subjected to different treatments that alter their properties

after initial production. This imply that dust is present different shape and size in the ISM. Models of dust use usually a power law to describe the size distribution of grains:

$$\frac{dn_{gr}}{da} \propto a^{-\beta} \tag{1.2}$$

with $a$ being the radius and the grain volume $V = 4\pi a^3/3$. This power law is used in the classic MRN model[53] (for Mathis, Rumpl, and Nordsieck) published in 1977 and based on the observed extinction of starlight through diffuse clouds. The law was constrained by a minimum and maximum value of a of 50 Å and 0.25 $\mu$m respectively. Successive progress in astronomical observations lead to necessary revision of the model. For example emission in the $3 - 60$ $\mu$m range imply the presence of very small grain $< 50$ Å[54], and observation of features associated to polycyclic aromatic hydrocarbons[55] (PAHs) suggesting that graphite grains extends down to the molecular regime. Figure 1.8 shows the size distribution of interstellar dust grain using the Weingarter & Draine (WD) model[52] taking into account very small carbonaceous grains. We can see a peak for silicate grains around 0.2 $\mu$m, with a maximum at 0.4 $\mu$m, and a broad peak for Graphite/PAHs grains around 0.3 $\mu$m, with a maximum at 0.8 $\mu$m. Therefore, using the WD model, we can conclude that the typical bare dust grain size is in the order of $10^{-1}$ $\mu$m.

#### 1.3.3.3 Grain surface processes

Interstellar grains amount for only 1% of the total mass when compared to the gaseous elements of the ISM but the chemical processes happening on their surfaces are essential to explain the chemical abundance observed in the ISM. For example $H_2$, the most abundant molecule of the interstellar medium[56,57], is formed primarily by recombination of neutral hydrogen ($H + H \rightarrow H_2$) on the grain surface, a reaction path unavailable in the gas phase due to its instability that leads to direct dissociation. On the grain surface however, the reaction energy released due to the formation of molecular hydrogen can be dissipated into the grain mantle, leading to a stable molecule that can be ejected back to the gas-phase through reasons that are still partly unknown. Of course, gas-phase formation for $H_2$ is still possible, mainly through formation of an intermediate $H^-$ ion, but are mostly inefficient when compared to solid phase formation. Furthermore, as the icy mantle of grains capture molecules in the gases depending on their environment properties,

like temperature and density, they can be a rich reserve of molecular diversity.

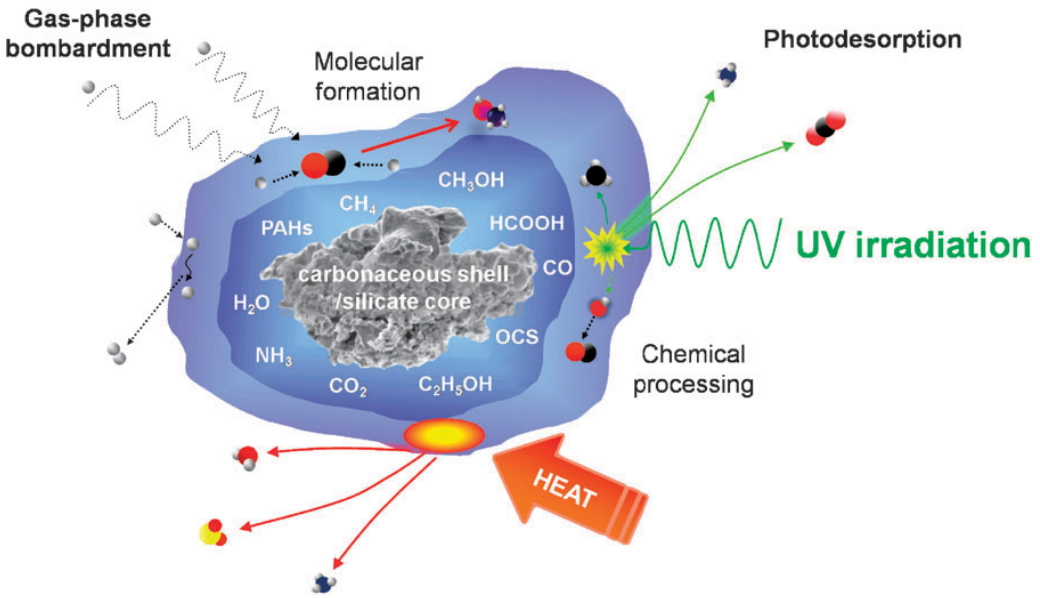



**Figure 1.9:** Main processes happening on the surface of interstellar ices, as well as main constituents of the ice mantles. Taken from Burke et al.[58]

In this part we will quickly describe the difference processes that can happen on the interstellar grain surface (also represented on Figure 1.9).

**Adsorption/Accretion** As said before, in molecular clouds, icy mantles form on the surface of dust grains. These mantles are formed via accretion (adsorption) of atoms and molecules contained in the gas-phase of molecular clouds, but also by coagulation of already formed grains. These processes participate in the dust enrichment.

**Diffusion** After being adsorption on the grain's surface atoms and molecules can diffuse on it to find a more stable binding site or to react with the other aggregated species. To diffuse a barrier needs to be overcome that depends on the binding strength of the species with its surroundings (*i.e.* the shape and composition of the binding site) and its temperature. In parts of the ISM where temperatures are too low for molecules to overcome the diffusion barrier themselves (like dark molecular clouds), the diffusion process is per-

formed mainly via quantum tunneling hopping[59], but also thermal hopping due to heating by way of stellar radiations.

**Molecular formation** The molecules present in molecular clouds can aggregate on the grain's surface directly, but also be formed, as showed previously for water molecule formation on the bare grain and mantles or for the formation of molecular hydrogen, after diffusion of the constituents. The catalytic properties of the icy grains have been extensively studied for molecules like $H_2$[60–63] and even $CH_3OH$[64,65]. And processes like absorption of photons have been linked to formation of molecules on ices[66]. Chemical reaction can also reconfigure the surface's amorphous shape, either by reformation of water molecules previously photodissociated[67], or by re-organization of the ice matrix due to release of energy after the reaction. Figure. 1.10 shows three different mechanisms of surface reactions: the Langmuir-Himshelwood mechanism, where two molecules are adsorbed on the surface, diffuse with low energy, and react to form a new product; Eley-Rideal mechanism, where one molecule is adsorbed on the surface, and another reacts directly with it from the gas-phase; the Hot atom mechanism, where one excited atom is adsorbed and subsequently diffuse with high energy until it reacts with another adsorbed element.

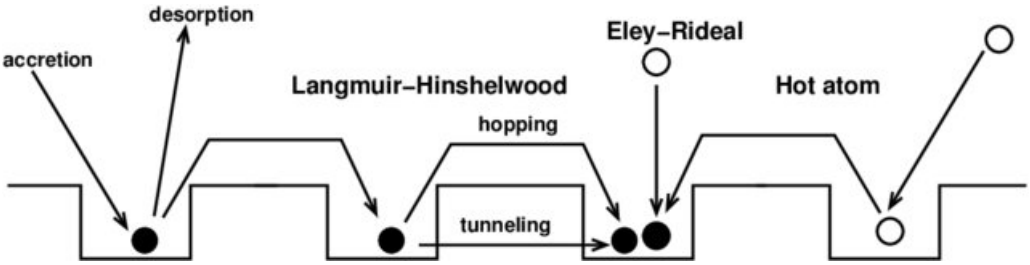



**Figure 1.10:** Several mechanisms for surface reactions: Langmuir-Hinshelwood, Eley-Rideal, and Hot atom. Filled circles are stable species, unfilled circles are excited species. Taken from https://hdl.handle.net/1887/16228

**Desorption** The formation of molecules on the grain surface helps to understand the abundance of several molecules of the ISM but only if these molecules desorb from the grains and re-populate the gas-phase of the interstellar medium. Several routes are available for molecules to desorb from

the surface of interstellar icy grains, starting with thermal desorption due to the gas-phase temperature. Indeed, dense molecular clouds can collapse and become star forming regions, increasing their temperatures and therefore provoking the sublimation of the ices, releasing the molecular content of the grains in the gas phase. Cosmic rays can also cause thermal desorption by passing through the grains. Molecules can also be photodesorbed by way of ultraviolet irradiation. The photodesorption of molecules is done either directly, or indirectly if the desorption is preceded by the photodestruction of the molecule and followed by the desorption of its constituents. In the case of water for example, it can dissociate into H + OH, leaving H to desorb from its excess of energy and OH to either react again on the grain's surface to form another water molecule or to also desorb[68]. Shocks that processes grains can lead to desorption of molecules by shattering and vaporization. Finally, chemical desorption is believed to happen via the partial dissipation of the reaction energy into the grain and the ejection of the formed molecule due to the remaining energy.

These different processes dictate in part the observed molecular abundances of the ISM. Furthermore, the binding energy of these observed elements, the energy necessary to remove an adsorbed atom or molecule, is of utmost importance as it rules the propensity of species to be adsorbed, diffused, or desorbed from the grain, making it a crucial part of numerical astrochemical models for obtaining accurate predictions of molecular abundances. In Chapter 5 of this thesis we will compute this quantity for three interstellar relevant molecules, name $NH_3$, $H_2$, and $CH_3OH$.

# References


[1] B. Ryden, *Introduction to Cosmology*. Cambridge University Press, nov 2016.

[2] D. Clayton, *Principles of Stellar Evolution and Nucleosynthesis*. Astronomy / Astrophysics, University of Chicago Press, 1983.

[3] J. A. Johnson, "Populating the periodic table: Nucleosynthesis of the elements," *Science*, vol. 363, pp. 474–478, feb 2019.

[4] T. Abel, G. L. Bryan, and M. L. Norman, "The Formation of the First Star in the Universe," *Science*, vol. 295, pp. 93–98, jan 2002.

[5] C. E. Rolfs and W. S. Rodney, *Cauldrons in the cosmos : nuclear astrophysics*. 1988.

[6] R. Ebert, "Über die Verdichtung von H I-Gebieten. Mit 5 Textabbildungen," *Zeitschrift für Astrophysik*, vol. 37, p. 217, Jan. 1955.

[7] W. B. Bonnar, "Boyle's Law and Gravitational Instability," *Monthly Notices of the Royal Astronomical Society*, vol. 116, pp. 351–359, jun 1956.

[8] G. McCracken and P. Stott, "Fusion in the Sun and Stars," in *Fusion*, pp. 15–29, Elsevier, jan 2013.

[9] S. Kwok, *Physics and Chemistry of The Interstellar Medium*. University Science Books, 2007.

[10] A. Shaw, *Astrochemistry: From Astronomy to Astrobiology*. Wiley, 2007.

[11] D. A. Williams and C. Cecchi-Pestellini, *The Chemistry of Cosmic Dust*. Cambridge: Royal Society of Chemistry, nov 2015.

[12] P. Caselli and C. Ceccarelli, "Our astrochemical heritage," *Astronomy and Astrophysics Review*, vol. 20, no. 1, 2012.

[13] M. Asplund, N. Grevesse, A. J. Sauval, and P. Scott, "The Chemical Composition of the Sun," *Annual Review of Astronomy and Astrophysics*, vol. 47, pp. 481–522, sep 2009.

[14] J. Dunham, T., "Interstellar Neutral Potassium and Neutral Calcium," *Publications of the Astronomical Society of the Pacific*, vol. 49, p. 26, feb 1937.

[15] P. Swings and L. Rosenfeld, "Considerations Regarding Interstellar Molecules," *The Astrophysical Journal*, vol. 86, p. 483, nov 1937.

[16] A. McKellar, "Evidence for the Molecular Origin of Some Hitherto Unidentified Interstellar Lines," *Publications of the Astronomical Society of the Pacific*, vol. 52, p. 187, jun 1940.

[17] S. Weinreb, A. H. Barrett, M. L. Meeks, and J. C. Henry, "Radio Observations of OH in the Interstellar Medium," *Nature*, vol. 200, pp. 829–831, nov 1963.

[18] Astronomy Correspondent, "Interstellar molecules: Carbon monoxide in space," *Nature*, vol. 226, no. 5244, pp. 409–410, 1970.

[19] A. C. Cheung, D. M. Rank, C. H. Townes, D. D. Thornton, and W. J. Welch, "Detection of Water in Interstellar Regions by its Microwave Radiation," *Nature*, vol. 221, pp. 626–628, feb 1969.

[20] A. C. Cheung, D. M. Rank, C. H. Townes, D. D. Thornton, and W. J. Welch, "Detection of $NH_3$ Molecules in the Interstellar Medium by Their Microwave Emission," *Physical Review Letters*, vol. 21, pp. 1701–1705, dec 1968.

[21] L. E. Snyder, D. Buhl, B. Zuckerman, and P. Palmer, "Microwave Detection of Interstellar Formaldehyde," *Physical Review Letters*, vol. 22, pp. 679–681, mar 1969.

[22] L. E. Snyder and D. Buhl, "Observations of Radio Emission from Interstellar Hydrogen Cyanide," *The Astrophysical Journal*, vol. 163, p. L47, jan 1971.

[23] B. A. McGuire, "2021 Census of Interstellar, Circumstellar, Extragalactic, Protoplanetary Disk, and Exoplanetary Molecules," *The Astrophysical Journal Supplement Series*, vol. 259, p. 30, apr 2022.

[24] L. M. Haffner, R.-J. Dettmar, J. E. Beckman, K. Wood, J. D. Slavin, C. Giammanco, G. J. Madsen, A. Zurita, and R. J. Reynolds, "The

warm ionized medium in spiral galaxies," *Reviews of Modern Physics*, vol. 81, pp. 969–997, jul 2009.

[25] K. Willacy, W. D. Langer, and T. Velusamy, "Dust Emission and Molecular Depletion in L1498," *The Astrophysical Journal*, vol. 507, pp. L171–L175, nov 1998.

[26] P. Caselli, C. M. Walmsley, M. Tafalla, L. Dore, and P. C. Myers, "CO Depletion in the Starless Cloud Core L1544," *The Astrophysical Journal*, vol. 523, pp. L165–L169, oct 1999.

[27] A. Bacmann, B. Lefloch, C. Ceccarelli, A. Castets, J. Steinacker, and L. Loinard, "The degree of CO depletion in pre-stellar cores," *Astronomy & Astrophysics*, vol. 389, pp. L6–L10, jul 2002.

[28] M. P. Redman, J. M. C. Rawlings, D. J. Nutter, D. Ward-Thompson, and D. A. Williams, "Molecular gas freeze-out in the pre-stellar core L1689B," *Monthly Notices of the Royal Astronomical Society*, vol. 337, pp. L17–L21, dec 2002.

[29] S. A. SANDFORD, "The inventory of interstellar materials available for the formation of the solar system," *Meteoritics & Planetary Science*, vol. 31, pp. 449–476, jul 1996.

[30] N. Boudin, W. A. Schutte, and J. M. Greenberg, "Constraints on the abundances of various molecules in interstellar ice: laboratory studies and astrophysical implications," *A&A*, vol. 331, pp. 749–759, 1998.

[31] E. L. Gibb, D. C. B. Whittet, W. A. Schutte, A. C. A. Boogert, J. E. Chiar, P. Ehrenfreund, P. A. Gerakines, J. V. Keane, A. G. G. M. Tielens, E. F. van Dishoeck, and O. Kerkhof, "An Inventory of Interstellar Ices toward the Embedded Protostar W33A," *The Astrophysical Journal*, vol. 536, pp. 347–356, jun 2000.

[32] P. Ehrenfreund and W. Schutte, "ISO observations of interstellar ices: Implications for the pristinity of comets," *Advances in Space Research*, vol. 25, pp. 2177–2188, jun 2000.

[33] P. Ehrenfreund and S. B. Charnley, "Organic Molecules in the Interstellar Medium, Comets, and Meteorites: A Voyage from Dark Clouds

to the Early Earth," *Annual Review of Astronomy and Astrophysics*, vol. 38, pp. 427–483, sep 2000.

[34] E. L. Gibb, D. C. B. Whittet, A. C. A. Boogert, and A. G. G. M. Tielens, "Interstellar Ice: The Infrared Space Observatory Legacy," *The Astrophysical Journal Supplement Series*, vol. 151, pp. 35–73, mar 2004.

[35] E. Dartois, "The Ice Survey Opportunity of ISO," *Space Science Reviews*, vol. 119, pp. 293–310, aug 2005.

[36] K. Demyk, E. Dartois, L. D'Hendecourt, M. Jourdain de Muizon, A. M. Heras, and M. Breitfellner, "Laboratory identification of the 4.62mu M solid state absorption band in the ISO-SWS spectrum of RAFGL 7009S," *A&A*, vol. 339, pp. 553–560, 1998.

[37] L. D'Hendecourt, M. Jourdain de Muizon, E. Dartois, M. Breitfellner, P. Ehrenfreund, J. Benit, F. Boulanger, J. L. Puget, and H. J. Habing, "ISO-SWS observations of solid state features towards RAFGL 7009S.," *A&A*, vol. 315, pp. L365–L368, 1996.

[38] E. Dartois, W. Schutte, T. R. Geballe, K. Demyk, P. Ehrenfreund, and L. D'Hendecourt, "Methanol: The second most abundant ice species towards the high-mass protostars RAFGL7009S and W 33A," *A&A*, vol. 342, pp. L32–L35, 1999.

[39] T. Y. Brooke, K. Sellgren, and T. R. Geballe, "New 3 Micron Spectra of Young Stellar Objects with $H_2O$ Ice Bands," *The Astrophysical Journal*, vol. 517, pp. 883–900, jun 1999.

[40] P. Ehrenfreund, L. D'Hendecourt, E. Dartois, M. Jourdain de Muizon, M. Breitfellner, J. L. Puget, and H. J. Habing, "ISO Observations of Interstellar Ices and Implications for Comets," *Icarus*, vol. 130, pp. 1–15, nov 1997.

[41] A. G. G. M. Tielens and L. J. Allamandola, "Evolution of Interstellar Dust," in *Physical Processes in Interstellar Clouds*, pp. 333–376, Dordrecht: Springer Netherlands, 1987.

[42] D. C. B. Whittet, W. A. Schutte, A. G. G. M. Tielens, A. C. A. Boogert, T. de Graauw, P. Ehrenfreund, P. A. Gerakines, F. P. Helmich,

T. Prusti, and E. F. van Dishoeck, "An ISO SWS view of interstellar ices: first results.," *A&A*, vol. 315, pp. L357–L360, 1996.

[43] L. B. D'Hendecourt and M. Jourdain de Muizon, "The discovery of interstellar carbon dioxide.," *A&A*, vol. 223, pp. L5–L8, 1989.

[44] M. E. Palumbo, T. R. Geballe, and A. G. G. M. Tielens, "Solid Carbonyl Sulfide (OCS) in Dense Molecular Clouds," *The Astrophysical Journal*, vol. 479, pp. 839–844, apr 1997.

[45] D. Gobrecht, I. Cherchneff, A. Sarangi, J. M. C. Plane, and S. T. Bromley, "Dust formation in the oxygen-rich AGB star IK Tauri," *Astronomy & Astrophysics*, vol. 585, p. A6, jan 2016.

[46] T. Henning, "Cosmic Silicates," *Annual Review of Astronomy and Astrophysics*, vol. 48, pp. 21–46, aug 2010.

[47] N. Liu, S. Cristallo, and D. Vescovi, "Slow Neutron-Capture Process: Low-Mass Asymptotic Giant Branch Stars and Presolar Silicon Carbide Grains," *Universe*, vol. 8, p. 362, jun 2022.

[48] A. G. G. M. Tielens, L. B. F. M. Waters, and T. J. Bernatowicz, "Origin and Evolution of Dust in Circumstellar and Interstellar Environments," in *Chondrites and the Protoplanetary Disk* (A. N. Krot, E. R. D. Scott, and B. Reipurth, eds.), vol. 341 of *Astronomical Society of the Pacific Conference Series*, p. 605, Dec. 2005.

[49] P. Massey, B. Plez, E. M. Levesque, K. A. G. Olsen, G. C. Clayton, and E. Josselin, "The Reddening of Red Supergiants: When Smoke Gets in Your Eyes," *The Astrophysical Journal*, vol. 634, pp. 1286–1292, dec 2005.

[50] A. S. Ferrarotti and H.-P. Gail, "Composition and quantities of dust produced by AGB-stars and returned to the interstellar medium," *Astronomy & Astrophysics*, vol. 447, pp. 553–576, feb 2006.

[51] B. Ryden and R. W. Pogge, *Interstellar and Intergalactic Medium*. Cambridge University Press, mar 2021.

[52] J. C. Weingartner and B. T. Draine, "Dust Grain–Size Distributions and Extinction in the Milky Way, Large Magellanic Cloud, and Small

Magellanic Cloud," *The Astrophysical Journal*, vol. 548, pp. 296–309, feb 2001.

[53] J. S. Mathis, W. Rumpl, and K. H. Nordsieck, "The size distribution of interstellar grains," *The Astrophysical Journal*, vol. 217, p. 425, oct 1977.

[54] B. T. Draine and N. Anderson, "Temperature fluctuations and infrared emission from interstellar grains," *The Astrophysical Journal*, vol. 292, p. 494, may 1985.

[55] A. Leger and J. L. Puget, "Identification of the Unidentified Infrared Emission Features of Interstellar Dust," *A&A*, vol. 137, pp. L5–L8, 1984.

[56] D. Hollenbach and E. E. Salpeter, "Surface Recombination of Hydrogen Molecules," *The Astrophysical Journal*, vol. 163, p. 155, jan 1971.

[57] G. Vidali, "$H_2$ Formation on Interstellar Grains," *Chemical Reviews*, vol. 113, pp. 8762–8782, dec 2013.

[58] D. J. Burke and W. A. Brown, "Ice in space: surface science investigations of the thermal desorption of model interstellar ices on dust grain analogue surfaces," *Physical Chemistry Chemical Physics*, vol. 12, p. 5947, jun 2010.

[59] W. D. Watson and E. E. Salpeter, "Molecule Formation on Interstellar Grains," *The Astrophysical Journal*, vol. 174, p. 321, jun 1972.

[60] G. Manicò, G. Ragunì, V. Pirronello, J. E. Roser, and G. Vidali, "Laboratory Measurements of Molecular Hydrogen Formation on Amorphous Water Ice," *The Astrophysical Journal*, vol. 548, pp. L253–L256, feb 2001.

[61] J. E. Roser, G. Manico, V. Pirronello, and G. Vidali, "Formation of Molecular Hydrogen on Amorphous Water Ice: Influence of Morphology and Ultraviolet Exposure," *The Astrophysical Journal*, vol. 581, pp. 276–284, dec 2002.

[62] L. Hornekær, A. Baurichter, V. V. Petrunin, D. Field, and A. C. Luntz, "Importance of Surface Morphology in Interstellar $H_2$ Formation," *Science*, vol. 302, pp. 1943–1946, dec 2003.

[63] S. Pantaleone, J. Enrique-Romero, C. Ceccarelli, S. Ferrero, N. Balucani, A. Rimola, and P. Ugliengo, "$H_2$ Formation on Interstellar Grains and the Fate of Reaction Energy," *The Astrophysical Journal*, vol. 917, p. 49, aug 2021.

[64] N. Watanabe, T. Shiraki, and A. Kouchi, "The Dependence of $H_2CO$ and $CH_3$ OH Formation on the Temperature and Thickness of $H_2O$-CO Ice during the Successive Hydrogenation of CO," *The Astrophysical Journal*, vol. 588, pp. L121–L124, may 2003.

[65] N. Watanabe, A. Nagaoka, T. Shiraki, and A. Kouchi, "Hydrogenation of CO on Pure Solid CO and CO-$H_2O$ Mixed Ice," *The Astrophysical Journal*, vol. 616, pp. 638–642, nov 2004.

[66] J. D. Thrower, D. J. Burke, M. P. Collings, A. Dawes, P. D. Holtom, F. Jamme, P. Kendall, W. A. Brown, I. P. Clark, H. J. Fraser, M. R. S. McCoustra, N. J. Mason, and A. W. Parker, "Desorption of Hot Molecules from Photon Irradiated Interstellar Ices," *The Astrophysical Journal*, vol. 673, pp. 1233–1239, feb 2008.

[67] P. Redondo, F. Pauzat, Y. Ellinger, and A. Markovits, "Reconstruction of water ice: the neglected process OH + OH $\rightarrow$ $H_2O$ + O," *Astronomy & Astrophysics*, vol. 638, p. A125, jun 2020.

[68] S. Andersson and E. F. van Dishoeck, "Photodesorption of water ice," *Astronomy & Astrophysics*, vol. 491, pp. 907–916, dec 2008.

# Chapter 2

# Computational methods

# Abstract

In this chapter we will introduce some important theoretical aspects of quantum chemistry and their application in computational chemistry methods. We will start by presented the basis of quantum chemistry by explaining the electronic hamiltonian and the Born-Oppenheimer approximation. Then, we will introduce computational quantum chemistry methods from Hartree-Fock to DFT and semi-empirical methods. For DFT we will present several functional widely in use in computational chemistry groups, from least to most accurate. A short part will be dedicated to non quantum methods called molecular mechanic methods. Next, we will introduce several calculation methods aimed at obtaining important properties of the studied structures, like stable geometries and harmonic frequencies. Finally we will present the GFNn-xTB set of methods, developed by the Grimme group at the University of Bonn, comprised of both semiempirical quantum mechanical methods and a Force-Field method, and extensively used in the work of this thesis.

This chapter is mainly based and adapted from the books "Introduction to Computational Chemistry" 3rd Edition by Frank Jensen[1], and "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory" by Attila Szabo and Neil S. Ostlund[2].


## 2.1 Introduction

Computational quantum chemistry is a field of physical chemistry that aims at predicting physical and chemical properties of molecules and materials by solving the corresponding Schrödinger equation which gives both the energy of the ground state and its associated wavefunction from which any physicochemical properties can be computed.

## 2.2 Quantum chemistry

The motion and interaction of electrons and nuclei can not be described via classical mechanics-based methods and needs the application of Quantum Mechanics for a correct and accurate description. Therefore, instead of Newton's law of motion we need to resort to the time dependent non relativistic

Schrödinger equation[3]:

$$\begin{aligned} i\hbar\frac{\partial\left|\Psi\right\rangle}{\partial t} &= \hat{H}\left|\Psi\right\rangle \\ &= \hat{T}\left|\Psi\right\rangle + \hat{V}\left|\Psi\right\rangle \\ &= -\frac{\hbar^2}{2m}\nabla_{\boldsymbol{r}}^2\left|\Psi\right\rangle + \hat{V}\left(\boldsymbol{r},t\right)\left|\Psi\right\rangle \end{aligned} \tag{2.1}$$

where $\left|\Psi\right\rangle = \left|\Psi\left(\boldsymbol{r},t\right)\right\rangle$ is the wave function describing the motion of the considered particle over all its degrees of freedom $\boldsymbol{r}$, $\hbar$ is the Planck constant divided by $2\pi$, $m$ is the particle's mass, and $\hat{H} = \hat{H}\left(\boldsymbol{r},t\right)$ is the Hamiltonian operator containing the kinetic ($\hat{T}$) and potential ($\hat{V}$) operators. If the potential part of the Hamiltonian operator $\hat{H}$ is time independent, the Hamiltonian becomes time independent and the wave function $\left|\Psi\right\rangle$ can be separated into two parts, one positional and one temporal taking the form of a phase factor:

$$\begin{aligned} \hat{V}\left(\boldsymbol{r},t\right) &= \hat{V}\left(\boldsymbol{r}\right) \\ \hat{H}\left(\boldsymbol{r},t\right) &= \hat{H}\left(\boldsymbol{r}\right) \\ \left|\Psi\left(\boldsymbol{r},t\right)\right\rangle &= \left|\Psi\left(\boldsymbol{r}\right)\right\rangle e^{-\frac{iEt}{\hbar}} \end{aligned} \tag{2.2}$$

This results in the time independent non relativistic Schrödinger equation:

$$\hat{H}\left(\boldsymbol{r}\right)\left|\Psi\left(\boldsymbol{r}\right)\right\rangle = E\left|\Psi\left(\boldsymbol{r}\right)\right\rangle \tag{2.3}$$

The goal of Quantum chemistry is to solve as accurately as possible equation 2.3 for the considered system.

The Hamiltonian operator for a chemical system can be written as[1]:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{en} + \hat{V}_{ee} + \hat{V}_{nn} \tag{2.4}$$

where the operators are, in order: the kinetic energy operator for nuclei; the kinetic energy operator for electrons; the electron-nucleus potential energy operator, or coulomb attraction operator, representing the attraction between nuclei and electrons; the electron-electron potential energy operator, representing the interaction between electrons; and the potential energy operator, or coulomb repulsion operator, representing the repulsion between nuclei. For a system of $N$ electrons and $M$ nuclei, we can rewrite Eq. 2.4 as[4]:

$$\begin{aligned}\hat{H} = &-\sum_{A}^{M}\frac{\hbar^2}{2M_A}\nabla^2_{\boldsymbol{R}_A} - \sum_{i}^{N}\frac{\hbar}{2m_e}\nabla^2_{\boldsymbol{r}_i} - \sum_{A}^{M}\sum_{i}^{N}\frac{Z_A e^2}{4\pi\epsilon_0\left|\boldsymbol{R}_A - \boldsymbol{r_i}\right|} \\ &-\sum_{i}^{N}\sum_{j>i}^{N}\frac{e^2}{4\pi\epsilon_0\left|\boldsymbol{r}_i - \boldsymbol{r}_j\right|} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B e^2}{4\pi\epsilon_0\left|\boldsymbol{R}_A - \boldsymbol{R}_B\right|}\end{aligned} \tag{2.5}$$

Inside the operators, $M_A$ is the mass of the $A_{th}$ nucleus; $\nabla^2$ is the Laplacian operator, which applied to the $i^{th}$ particles gives $\nabla^2_{\boldsymbol{r}_i} \equiv \nabla_{\boldsymbol{r}_i}\cdot\nabla_{\boldsymbol{r}_i} = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$ ; $Z_A$ is the atomic number of the $\mathrm{A}_{th}$ nucleus; $e$ is the charge of the electron; $\epsilon_0$ is the vacuum permittivity; $R_A$ and $r_i$ the position of the $\mathrm{A}_{th}$ nucleus A and the $\mathrm{i}_{th}$ electron, respectively.

To obtain a more handleable equation 2.5 we can use its dimensionless form where no $\hbar$, $e$, or $m_e$ appear using the atomic units[2]:

$$\begin{aligned}\hat{H} = &-\sum_{A}^{M}\frac{1}{2M_A}\nabla^2_{\boldsymbol{R}_A} - \sum_{i}^{N}\frac{1}{2}\nabla^2_{\boldsymbol{r}_i} - \sum_{A}^{M}\sum_{i}^{N}\frac{Z_A}{\boldsymbol{r}_{Ai}} \\ &+\sum_{i}^{N}\sum_{j>i}^{N}\frac{1}{\boldsymbol{r}_{ij}} + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R}_{AB}}\end{aligned} \tag{2.6}$$

where $\boldsymbol{r}_{Ai} = |\boldsymbol{R}_A - \boldsymbol{r}_i|$ is the absolute distance between nucleus A and electron i, $\boldsymbol{r}_{ij} = |\boldsymbol{r}_i - \boldsymbol{r}_j|$ is the absolute distance between electrons i and j, and $\boldsymbol{R}_{AB} = |\boldsymbol{R}_A - \boldsymbol{R}_B|$ is the absolute distance between nuclei A and B.

### 2.2.1 Born-Oppenheimer approximation

Analytical solutions to the Schrödinger equation are possible only for very small and simple systems, mainly the H atom. For more complex systems, numerical approximation are therefore the key to obtain results that are accurate but also affordable for the available computational resources. The most famous one is the Born-Oppenheimer approximation[5] where the wave functions of nuclei and electrons are treated separately. The justification for this approximation comes from the important mass velocity difference between nuclei and electrons, allowing us to consider the electrons as moving inside the field of fixed nuclei[6]. In this context we can consider the $\hat{T}_n$ operator

negligible and the $\hat{V}_{nn}$ constant. Hence, equation 2.6 can be approximated with the electronic Hamiltonian:

$$\hat{H}_{elec} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{2.7}$$

In this way we obtain a Schrödinger equation using only the electronic Hamiltonian:

$$\hat{H}_{elec} \left|\Psi_{elec}\right\rangle = E_{elec} \left|\Psi_{elec}\right\rangle \tag{2.8}$$

Its solution, *i.e.* the electronic wave function $\Psi_{elec}$ describing the motion of the electrons, is defined as:

$$\left|\Psi_{elec}\right\rangle = \left|\Psi_{elec}\left(\{\boldsymbol{r}_i\}\,;\{\boldsymbol{R}_A\}\right)\right\rangle \tag{2.9}$$

which depends explicitly on the electronic coordinates and only parametrically on the nuclei coordinates. The electronic energy, also parametrically dependent on the nuclei coordinates, is thus defined as:

$$E_{elec} = E_{elec}\left(\{\boldsymbol{R}_A\}\right) \tag{2.10}$$

Finally, in the Born-Oppenheimer approximation, the total energy $E_{tot}$ is equal to the electronic energy $E_{elec}$ plus the atomic repulsion, considered constant in the Born-Oppenheimer approximation:

$$E_{tot} = E_{elec} + \hat{V}_{nn} = E_{elec} + \sum_{A}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{\boldsymbol{R}_{AB}} \tag{2.11}$$

Reciprocally, the inverse logic can be applied to the nuclear part of the system to describe the motion of the nuclei, and to obtain a Schrödinger equation using the nuclear Hamiltonian $\hat{H}_{nucl}$ and its solution, the nuclear wave function$\left|\Psi\right\rangle_{nucl}$:

$$\begin{aligned} \hat{H}_{nucl} \left|\Psi_{nucl}\right\rangle &= E \left|\Psi_{nucl}\right\rangle \\ \left|\Psi_{nucl}\right\rangle &= \left|\Psi_{nucl}\left(\{\boldsymbol{R}_A\}\right)\right\rangle \end{aligned} \tag{2.12}$$

As electrons move much faster than nuclei we take the average of the electronic motion as an approximation in the nuclear Hamiltonian, implying that the electronic problem must be solved before the nuclear one

$$
\begin{aligned}
\hat{H}_{nucl} &= -\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 + \left\langle -\sum_{i=1}^{N}\frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \right\rangle \\
&\quad + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R}_{AB}} \\
&= -\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 + E_{elec}\left(\{\boldsymbol{R}_A\}\right) + \sum_{A}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\boldsymbol{R}_{AB}} \\
&= -\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 + E_{tot}\left(\{\boldsymbol{R}_A\}\right)
\end{aligned}
\tag{2.13}
$$

In Eq. 2.13, the first term describes the kinetics of nuclei and the second is the total energy $E_{tot}$ of Eq. 2.11 acting as a potential for the motion of the nuclei. Thus, in the Born-Oppenheimer approximation, the nuclei move on a Potential Energy Surface (PES) obtained after solving the electronic problem for the different nuclei positions. By solving Eq. 2.13 we obtain the energy levels for the nuclear motion, and more precisely for nuclear translation, vibration, rotation, the last two being essential in spectroscopy.

The approximation to the total wave function is thus:

$$
\Psi\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_A\}\right) = \Psi_{elec}\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_A\}\right)\Psi_{nucl}\left(\{\boldsymbol{R}_A\}\right) \tag{2.14}
$$

### 2.2.2 Slater determinant

To solve the Schrödinger equation we also need to represent the wave function $|\Psi(\boldsymbol{r}_i)\rangle$ of our system, whose mathematical form is unknown. For a single particle the wave function is represented by a spatial orbital $|\Phi_i(\boldsymbol{r})\rangle$ describing the spatial distribution of the particle (an electron) so that:

$$
P_{a\leq x\leq b} = \int_a^b \left|\Psi(\boldsymbol{r})\right|^2 d\boldsymbol{r} \tag{2.15}
$$

is the probability of finding the particle between the positions $a$ and $b$ of our system. Electrons are Fermions, so they posses a half-integer spin and must follow the Pauli exclusion principles saying that within a quantum system two electrons cannot share the same quantum state. To describe the spin

of the electrons we introduce two orthonormal spin functions $|\alpha\rangle$ and $|\beta\rangle$ representing respectively the spin up and spin down states:

$$\begin{aligned} \langle \alpha \,|\, \alpha \rangle &= \langle \beta \,|\, \beta \rangle = 1 \\ \langle \alpha \,|\, \beta \rangle &= \langle \beta \,|\, \alpha \rangle = 0 \end{aligned} \tag{2.16}$$

The wave function obtained for a single particle is a spin orbital $|\phi(\boldsymbol{x})\rangle$, where $\boldsymbol{x}$ indicates spatial and spin coordinates. It can take two forms: one with spin up $|\phi(\boldsymbol{r})\rangle \circledast |\alpha\rangle$, and one with spin down $|\phi(\boldsymbol{r})\rangle \circledast |\beta\rangle$. The spin orbitals are also orthonormal.

As said before, electrons follow the Pauli exclusion principle, also called the antisymetric principle, meaning that two particles in a system can not share the same quantum state. Therefore, for a system containing two particles, the wave function can not be a sum of the two possible states, represented as the product of two one particle wave functions, such as $\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \phi_i(\boldsymbol{x}_1)\,\phi_j(\boldsymbol{x}_2) + \phi_i(\boldsymbol{x}_2)\,\phi_j(\boldsymbol{x}_1)$, as the wave function would not be equal to zero if the two particles where in the same quantum states. Instead it is written such as $\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = -\Psi(\boldsymbol{x}_2, \boldsymbol{x}_1)$:

$$\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{\sqrt{2}} \left( \phi_i(\boldsymbol{x}_1)\,\phi_j(\boldsymbol{x}_2) - \phi_j(\boldsymbol{x}_1)\,\phi_i(\boldsymbol{x}_2) \right) \tag{2.17}$$

Eq. 2.17 can be rewritten as a determinant, called Slater determinant[7]:

$$\Psi(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \phi_i(\boldsymbol{x}_1) & \phi_j(\boldsymbol{x}_1) \\ \phi_i(\boldsymbol{x}_2) & \phi_j(\boldsymbol{x}_2) \end{vmatrix} \tag{2.18}$$

For a system of N electrons the Slater determinate becomes:

$$\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_i(\boldsymbol{x}_1) & \phi_j(\boldsymbol{x}_1) & \cdots & \phi_k(\boldsymbol{x}_1) \\ \phi_i(\boldsymbol{x}_2) & \phi_j(\boldsymbol{x}_2) & \cdots & \phi_k(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_i(\boldsymbol{x}_N) & \phi_j(\boldsymbol{x}_N) & \cdots & \phi_k(\boldsymbol{x}_N) \end{vmatrix} \tag{2.19}$$

The columns are labeled by spin orbitals while the row are labeled by electrons. Two identical columns would indicate two electrons in the same spin orbital, bringing the determinant to a value of 0. Swapping two columns inverts the sign of the determinant. Hence, a Slater determinant follows the

antisymmetry principle. To shorten the Slater determinant we can introduce a new notation only showing the diagonal elements and including the normalization factor:

$$\Psi\left(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\right)=\left|\phi_i\left(\boldsymbol{x}_1\right)\phi_j\left(\boldsymbol{x}_2\right)\cdots\phi_k\left(\boldsymbol{x}_N\right)\right\rangle \tag{2.20}$$

### 2.2.3 Variational method

As said before, the Schrödinger equation cannot be solved analytically for complex systems and needs to be solved numerically using approximations for the Hamiltonian and the wave function.

The time independent Schrödinger equation is an eigenvalue equation for which we search approximate solutions. However, given the Hamiltonian $\hat{H}$, we know that an infinite set of exact solutions to the Schrödinger equation exists with indices $\alpha=0,1,2,\ldots$:

$$\hat{H}\left|\Psi_\alpha\right\rangle=\epsilon_\alpha\left|\Psi_\alpha\right\rangle \tag{2.21}$$

where $\epsilon_0\leq\epsilon_1\leq\epsilon_2\leq\cdots\leq\epsilon_\alpha\leq\ldots$ are the set of real eigenvalues corresponding to the orthonormal eigenfunctions $\left|\Psi_\alpha\right\rangle$ meaning $\left\langle\Psi_\beta\right|\hat{H}\left|\Psi_\alpha\right\rangle=\epsilon_\alpha\delta_{\alpha\beta}$ ($\delta_{\alpha\beta}=1$ if $\alpha=\beta$ and 0 if $\alpha\neq\beta$). The variation method states that given an orthonormal trial wave function $\left|\Psi_{Trial}\right\rangle$ satisfying the necessary requirements, the energy associated with the trial function will always be greater to the true minimum energy of the system:

$$\left\langle\Psi_{Trial}\right|\hat{H}\left|\Psi_{Trial}\right\rangle\geq\epsilon_0 \tag{2.22}$$

The equality is true only if the trial wave function is equal to the exact ground state wave function ($\left|\Psi_{Trial}\right\rangle=\left|\Psi_0\right\rangle$). The goal is thus to minimize the energy associated to the trial wavefunction by finding the best trial function possible. The energy becomes a functional of the trial wavefunction

$$E\left[\Psi_{Trial}\right]=\left\langle\Psi_{Trial}\right|\hat{H}\left|\Psi_{Trial}\right\rangle \tag{2.23}$$

To minimize the energy of equation 2.23 the initial trial wavefunction is chosen and used to solve the Schrödinger equation. From the solution obtained is derived a new trial wavefunction with which we solve again the Schrödinger equation, thus obtaining newer and more accurate wavefunction. This iterative procedure is repeated until convergence of the wavefunction and is called Self Consistent Field (SCF) procedure.

## 2.3 Quantum Mechanical methods

To solve the Schrödinger equation various quantum mechanical (QM) methods are available. These QM methods can be classified into categories and sub-categories, depending on the number of parameters introduced in the Hamiltonian, from parameter free wave function-based methods (*ab-initio*), to more or less empirical Density Functionals, to semi-empirical methods. Ab-initio methods compute the properties of the studied system "from the beginning" or "from first principles" meaning that the inputs are only physical constants, while empirical methods relies on pre-obtained parameters. In the next section we presents some of these methods.

### 2.3.1 Hartree Fock method

The Hartree-Fock method[8–12] is used to solve the time independent non-relativistic Schrödinger equation for a system of N electrons. It takes into account the Born-Oppenheimer approximation, and its solution is a single Slater determinant $|\Psi_0\rangle$ representing the best possible approximation to the ground state of the system. Thus, we need to find a set of spin orbitals $\phi_a$ by minimizing the electronic energy via Eq. 2.23:

$$E_0 = \langle\Psi_0|\,\hat{H}\,|\Psi_0\rangle = \sum_i^N \langle\phi_i|\,h_i\,|\phi_i\rangle + \frac{1}{2}\sum_{ij}^N \left(\langle\phi_j|\,J_j - K_j\,|\phi_i\rangle\right) + V_{nn} \quad (2.24)$$

here $h$, $J$, and $K$ are respectively the mono-electronic Hamiltonian, the coulomb operator, and the exchange operator. They are defined as:

$$\begin{aligned} h_i &= -\frac{1}{2}\nabla^2 - \sum_a^M \frac{Z_a}{|\boldsymbol{R}_a - \boldsymbol{r}_i|} \\ J_i\,|\phi_j\rangle &= \langle\phi_i|\,\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}\,|\phi_i\phi_j\rangle \\ K_i\,|\phi_j\rangle &= \langle\phi_i|\,\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}\,|\phi_j\phi_i\rangle \end{aligned} \quad (2.25)$$

The exchange operator is a term with no classical interpretation that arises when two identical particles exchange both places and spin. The coulomb operator is a term coming from the interaction of electron $j$ over electron $i$.

To help find the solution for our system we introduce the Fock operator $F_i$:

$$F_i = h_i + \sum_{j}^{N} (J_j - K_j) \tag{2.26}$$

It represents the variation of energy for a single electron i associated to the spin orbital, also called molecular orbital (MO), $|\phi_i\rangle$, and describes with $h_i$ its kinetic energy and attraction to the nuclei, and through the $K$ and $J$ operators the interaction to the other electrons. In the basis set approximation, each MO, whose mathematical form is unknown, can be approximated as a linear combination of atomic orbitals (LCAO) $\chi$ also called basis function. The different ways available to represent these atomic orbitals will be discussed in section 2.3.3. For the HF solution we can use, for example, Slater type orbitals (STOs):

$$\phi_i = \sum_{\alpha}^{N_{basis}} c_{\alpha i}\chi_\alpha \tag{2.27}$$

where $N_{basis}$ is the number of basis function chosen in linear combination. Using the previous equation and the Fock operator of Eq. 2.26 we can write the Hartree-Fock equation:

$$F_i \sum_{\alpha}^{N_{basis}} c_{\alpha i}\chi_\alpha = \epsilon_i \sum_{\alpha}^{N_{basis}} c_{\alpha i}\chi_\alpha \tag{2.28}$$

Multiplying the left of Eq. 2.28 by $\phi_j$ basis function we can reformulate the HF problem as a general eigenvalue problem using the Roothaan-Hall equations[13,14]:

$$\begin{aligned} \mathbf{FC} &= \mathbf{SC}\epsilon \\ F_{\alpha\beta} &= \langle \chi_\alpha | \, \mathbf{F} \, | \chi_\beta \rangle \\ S_{\alpha\beta} &= \langle \chi_\alpha \, | \, \chi_\beta \rangle \\ C_{\alpha\beta} &= C_{\alpha i} \end{aligned} \tag{2.29}$$

The Fock matrix $\mathbf{F}$ contains the matrix elements $F_{\alpha\beta}$, the $\mathbf{S}$ matrix, also called the overlap matrix, describes the overlap between basis function via so-called overlap integrals.

The Fock matrix can only be known if the MOs coefficients $C_{\alpha i}$ are known, but to know the MO coefficients the Fock matrix needs to be diagonalized. As

the **F** matrix depends on its own solution, the resolution of Eq. 2.29 needs to be done iteratively. The process, called the self-consistent field (SCF) process (briefly seen in section 2.2.3), involves an initial guess of the **C** matrix, followed by the formation of the Fock matrix **F** and its diagonalization. The diagonalization results in a new set of coefficients for the **C** matrix. The process is reiterated until the coefficients obtained after diagonalization and the ones used for the formation of the Fock matrix converge to the same values, *i.e.* until they are equal.

### 2.3.2 Post Hartree Fock methods

The Hartree Fock method approximates the wave function as a single Slater determinant, and as such it does not take into account the whole electron correlation, *i.e.* the interaction between electrons in the studied system, but only on an average base for electrons with different spin (instantaneous correlation of motion is lost). Electrons with like spins are correctly treated via the exchange operator. Hence, the Hartree Fock energy is always superior to the exact energy of the system; the difference between the two quantities is called correlation energy, which is always negative as $E_{HF}$ is always superior to the exact energy $E_0$:

$$E_0 = E_{HF} + E_{corr} \tag{2.30}$$

The correlation energy is small compared to the exact energy, but it accounts for important effects such as bond breaking/making energy in proximity of transition state structures, magnetism and London dispersion interactions which are entirely missing by the HF method. Over the years several methods expanded the HF method to recover the electronic correlation and, in the following section, we present some of them.

#### 2.3.2.1 Configuration interaction

In the configuration interaction (CI) the wave function is represented as a linear combination of N-electrons Slater determinants. In theory, the exact solution to the Schrodinger equation can be computed using the full-CI, *i.e.* a linear combination of all Slater determinants needed to represent the ground state and every excited states of the system. However, it is affordable only for very simple systems as the number of Slater determinants increases drastically with the complexity of the structure, thus requiring too much computing time.

If we represent the Slater determinant as a set of spin orbitals$\{\chi_i\}$ that is itself separated into a set of occupied spin orbitals $\{\chi_a\}$, *i.e.* occupied by at least one electron, and a set of virtual spin orbitals $\{\chi_k\}$, *i.e.* without any electron, we can describe the ground state as:

$$|\phi_0\rangle = |\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\rangle \tag{2.31}$$

Following this convention, a determinant representing a single particle excitation, *i.e.* an electron in an occupied orbital is excited to a virtual orbital, as:

$$|\phi_a^r\rangle = |\chi_1\chi_2\cdots\chi_r\chi_b\cdots\chi_N\rangle \tag{2.32}$$

and doubly excited determinant, *i.e.* two electrons promoted to a virtual orbital, as:

$$|\phi_{ab}^{rs}\rangle = |\chi_1\chi_2\cdots\chi_r\chi_s\cdots\chi_N\rangle \tag{2.33}$$

For a complex system, the full CI wave function can therefore be written as:

$$\begin{aligned}|\Psi_0\rangle = c_0\,|\phi_0\rangle + \sum_{ar} c_a^r\,|\phi_a^r\rangle + \sum_{\substack{a<b\\r<s}} c_{ab}^{rs}\,|\phi_{ab}^{rs}\rangle \\ + \sum_{\substack{a<b<c\\r<s<t}} c_{abc}^{rst}\,|\phi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d\\r<s<t<u}} c_{abcd}^{rstu}\,|\phi_{abcd}^{rstu}\rangle + \cdots\end{aligned} \tag{2.34}$$

where $|\phi_{abc}^{rst}\rangle$ and $|\phi_{abcd}^{rstu}\rangle$ represent a triply and quadruply excited determinant respectively.

With K spin orbitals per Slater determinant the full CI wave function would contain $\binom{2K}{N} = \frac{(2K)!}{N!(2K-N)}$ determinants. For a system where $K = 10$ and $N = 4$, the number of electrons, the full CI wave function is made by 210 Slater determinants.

A convenient way to rewrite Eq. 2.34 is:

$$|\Psi_0\rangle = c_0\,|\phi_0\rangle + c_s\,|S\rangle + c_d\,|D\rangle + c_T\,|T\rangle + c_Q\,|Q\rangle + \cdots \tag{2.35}$$

where $|S\rangle$, $|D\rangle$, $|T\rangle$, $|Q\rangle$ represent single, double, triple, and quadruple excitation respectively.

The matrix obtained by solving the Schrödinger equation is

$$\begin{array}{c} \\ \langle\Psi_0| \\ \langle S| \\ \langle D| \\ \langle T| \\ \langle Q| \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} |\Psi_0\rangle & |S\rangle & |D\rangle & |T\rangle & |Q\rangle & \cdots \end{array} \\ \begin{bmatrix} \langle\Psi_0|\hat{H}|\Psi_0\rangle & \langle\Psi_0|\hat{H}|S\rangle & \langle\Psi_0|\hat{H}|D\rangle & \langle\Psi_0|\hat{H}|T\rangle & \langle\Psi_0|\hat{H}|Q\rangle & \cdots \\ & \langle S|\hat{H}|S\rangle & \langle S|\hat{H}|D\rangle & \langle S|\hat{H}|T\rangle & \langle S|\hat{H}|Q\rangle & \cdots \\ & & \langle D|\hat{H}|D\rangle & \langle D|\hat{H}|T\rangle & \langle D|\hat{H}|Q\rangle & \cdots \\ & & & \langle T|\hat{H}|T\rangle & \langle T|\hat{H}|Q\rangle & \cdots \\ & & & & \langle Q|\hat{H}|Q\rangle & \cdots \\ & & & & & \vdots \end{bmatrix} \end{array} \tag{2.36}$$

Some states have no coupling between them, *e.g.* ground state with single, triple, and quadruple excitations, or matrix elements which differ by more than 2 spin orbitals. As double excitations mix directly, contrary to single excitations, they have an important role in finding the correlation energy.

#### 2.3.2.2 Coupled cluster

Another well known method is the coupled cluster method[15–18] in which the wave function is written as:

$$|\Psi\rangle = e^{\boldsymbol{T}}|\phi_0\rangle \tag{2.37}$$

where $\boldsymbol{T}$ is the cluster operator[19] composed by all the excitation operators, which in turn are composed by all the excitation coefficients ($\boldsymbol{T_1}$ for single excitation coefficients, $\boldsymbol{T_2}$ for double excitation coefficients, etc...)

$$\boldsymbol{T} = \boldsymbol{T_1} + \boldsymbol{T_2} + \boldsymbol{T_3} + \boldsymbol{T_4} + \cdots \tag{2.38}$$

The Schrödinger equation in this formalism is written as:

$$\hat{H}|\Psi_0\rangle = \hat{H}e^{T}|\phi_0\rangle = Ee^{T}|\phi^0\rangle \tag{2.39}$$

Therefore, for the CCSD method, which stands for coupled cluster single, doubles, the cluster operator is written as $\boldsymbol{T} = \boldsymbol{T_1} + \boldsymbol{T_2}$ and the different elements of the CCSD matrix are:

$$\begin{aligned} \langle\Psi_0|e^{-(\boldsymbol{T_1}+\boldsymbol{T_2})}\hat{H}e^{(\boldsymbol{T_1}+\boldsymbol{T_2})}|\Psi_0\rangle &= E \\ \langle S|e^{-(\boldsymbol{T_1}+\boldsymbol{T_2})}\hat{H}e^{(\boldsymbol{T_1}+\boldsymbol{T_2})}|\Psi_0\rangle &= 0 \\ \langle D|e^{-(\boldsymbol{T_1}+\boldsymbol{T_2})}\hat{H}e^{(\boldsymbol{T_1}+\boldsymbol{T_2})}|\Psi_0\rangle &= 0 \end{aligned} \tag{2.40}$$

In the method name, the highest excitation order may be written in parenthesis (such as CCSD(T), CCSDT(Q)), which means that those excitations are treated with the perturbation theory.

### 2.3.3 Basis set

Molecular orbitals (MOs) from Slater determinant are represented via Basis set (BS) functions. These BS functions are linear combination of chosen function or atomic orbitals, as the form of molecular orbitals is unknown:

$$f_i = \sum_{\alpha}^{M} c_{\alpha i} \chi_\alpha \tag{2.41}$$

Different ways exist to describe basis function, the two most well known among localized basis sets being Slater-type orbitals (STOs):

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N Y_{l,m}(\theta,\varphi)\, r^{n-1} e^{-\zeta r} \tag{2.42}$$

and Gaussian-type orbitals (GTOs):

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = N Y_{l,m}(\theta,\varphi)\, r^{2n-2-l} e^{-\zeta r^2} \tag{2.43}$$

$$\chi_{\zeta,l_x,l_y,l_z}(x,y,z) = N x^{l_x} y^{l_y} z^{l_z} e^{-\zeta r^2} \tag{2.44}$$

which are the two ways for describing GTOs, *i.e.* with polar (2.43) and Cartesian (2.44) coordinates. In both STOs and GTOs $Y_{l,m}$ are spherical harmonic functions, $\zeta$ the exponent of the basis function, and $N$ a normalization factor. For cartesian GTOs, the parameters $l_x$, $l_y$, and $l_z$ determine the type of orbital, while for STOs and polar GTOs it is determined by $n$, $l$, and $m$.

In term of performance, STOs have better accuracy than GTOs, as GTOs have zero slope at the nucleus while STOs possess a cusp, thus GTOs have difficulties in representing the proper behavior of an electron near the nucleus. Furthermore, GTOs fall too rapidly far from the nucleus and represent loosely the tail of the wave-function. Hence, for obtaining a basis set of GTOs as accurate as a STOs basis set, we need more GTOs functions than STOs (roughly three times more). However, even though their accuracy is inferior to STOs, GTOs are more computationally cheap to treat than STOs, which

are mainly used for mono-atomic or bi-atomic systems when in need of precise results. GTOs are thus more widely spread in computational chemistry.

Choosing the number of functions to be used in the basis set is also important. As a minimum, one STO basis function is a good approximation for representing an atomic orbital, for GTOs one function is not enough, hence, an AO will be represented at minimum by a linear combination, also called contracted function, of GTOs. Each individual GT of the contracted function is called a primitive function. So, for example, the minimum basis set needed for an Hydrogen atom will be comprised of one 1s STO, or contracted function. For an atom of the second row, the minimum basis set will be comprised of a 1s and 2s function, and one set of 2p functions ($2p_x$, $2p_y$, and $2p_z$). Adding additional functions per AO can be used to better represent electrons distribution in systems where the distribution is different in different direction. Similarly, electron density near a nucleus can be polarized via the interaction of another nucleus. This polarization and the resulting asymmetry can be included in the basis set by adding what is called polarization functions. For a 1s AO of an Hydrogen atom, for example, a set of 2p functions can be added to the atom to introduce a polarization to the s orbital. Likewise, d-orbitals can be used to polarize p-orbitals, f-orbitals for d-orbitals, etc. In systems like anions or excited stated, where electrons are weakly bonded, or where the interesting properties depend on weakly bonded electrons, basis function with small exponent $\zeta$, called diffuse functions, can be use to augment the basis set. These functions are used to better represent the outer part, or tail, of the AO.

A third type of basis functions available in computational codes are plane waves. They are mostly used to model extended systems, generally periodic ones. They are modeled around the solution to the Schrödinger equation for a free electron and, because of that, they are widely used for metallic systems. This solution can be written as a sum of complex exponentials or cosine and sine functions:

$$\begin{aligned} \phi(x) &= Ae^{ikx} + Be^{-ikx} \\ \phi(r) &= Acos(kx) + Bsin(kx) \end{aligned} \tag{2.45}$$

For infinite systems, the different between consecutive energy levels becomes negligible and molecular orbitals fuse into so-called *bands* where electrons can be described by a basis set of *plane waves* that can be written as a complex function

$$\chi_k(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tag{2.46}$$

where $\boldsymbol{k}$ is the wave vector related to the electron kinetic energy via $E = \frac{1}{2}k^2$, and plays the same role as $\zeta$ in a GTO.

#### 2.3.3.1 Basis set superposition error

When using whatever STO or GTO, atoms at short distances use basis functions of their neighbors to complete their own basis set, in order to give to the electron the best variational degrees of freedom, so finding the best solution possible for the energy. However, this basis function overlap is not present at long distances thus creating an inconsistency, *i.e.* interaction at short distances will be maximized and at long distances will be minimized. This discrepancy is called the basis set superposition errors (BSSE)[20] and can be corrected either by increasing the number of basis functions, expensive in computational power, or by corrective methods such as the counterpoise correction (CP)[21,22], a cheaper solution than increasing the basis set.

The BSSE between two monomers A and B separated by a distance R can be estimated using the CP correction[22]:

$$BSSE = E_A^{AB}(R) + E_B^{AB}(R) - E_A - E_B \tag{2.47}$$

where $E_A$ is the energy of the isolated monomer A using the A basis set, $E_B$ is the energy of the isolated monomer B using the B basis set, $E_A^{AB}$ is the monomer energy of A using the full AB dimer basis set {AB} at the studied AB geometry, and $E_B^{AB}$ is the monomer energy of B using the full AB dimer basis set {AB} at the studied AB geometry. More precisely, $E_A^{AB}$ and $E_B^{AB}$ are computed at the dimer geometry by removing the nuclei B and A, respectively, but not their basis functions. The functions of a basis set without the corresponding nuclei (and electrons) are referred as *ghost functions*.

### 2.3.4 Density Functional Theory

Density functional Theory (DFT) is the most popular form of quantum chemistry method. It stems from an article by Pierre Hohenberg and Walter Kohn published in 1964[23].

Starting from the electron density represented by the following equation:

$$\rho(\boldsymbol{r}) = N \int d\boldsymbol{r}_2 \cdots \int d\boldsymbol{r}_N \Psi^*(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)\, \Psi(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{2.48}$$

and the Born Oppenheimer approximation (vide supra), we can imagine two external potentials $V_{ext}$ and $V_{ext'}$ resulting in two different Hamiltonians $\hat{H}$ and $\hat{H}'$ with corresponding wave functions $\Psi$ and $\Psi'$ but giving the same electron density $\rho_{\boldsymbol{r}}$. By employing the variation principle from Eq. 2.22 and using $\Psi$ as a trial wave function for $\hat{H}'$ and $\Psi'$ as a trial wave function for $\hat{H}$, we obtain:

$$\begin{aligned}
&\langle\Psi|\,\hat{H}'\,|\Psi\rangle > E_0' \\
&\langle\Psi|\,\hat{H}\,|\Psi\rangle + \langle\Psi|\,\hat{H}' - \hat{H}\,|\Psi\rangle > E_0' \\
&E_0 + \langle\Psi|\,V_{ext}' - V_{ext}\,|\Psi\rangle > E_0' \\
&E_0 + \int \rho\,(\boldsymbol{r})\,(V_{ext}' - V_{ext})\,d\boldsymbol{r} > E_0'
\end{aligned} \tag{2.49}$$

Similarly for $\hat{H}$ we have:

$$E_0' + \int \rho\,(\boldsymbol{r})\,(V_{ext} - V_{ext}')\,d\boldsymbol{r} > E_0 \tag{2.50}$$

Summing both equation we get the inconsistent inequality:

$$E_0 + E_0' > E_0 + E_0' \tag{2.51}$$

From this discrepancy we can see that the potential is directly linked to the electron density, meaning that the energy is a unique functional of $\rho$, $E\,[\rho]$, and can be computed exactly provided that the explicit form of the universal functional is known. By using the electron density at each point of the system instead of the position of its electrons, the system go from a wave function characterizing N electrons with 4N variables (three spatial positions plus one spin for every electron), to one quantity, only dependent on three spatial coordinates and independent of the total number of electrons.

The energy can be divided into three parts, the kinetic energy $T\,[\rho]$, the interaction between electrons $E_{ee}\,[\rho]$, and the attraction between electrons and nuclei $E_{ne}\,[\rho]$. Like the Hartree Fock method, the electron-electron interaction term can be divided into two parts: the Coulomb part $J\,[\rho]$, and the exchange part $K\,[\rho]$. The nucleus-electron attraction and the coulomb interactions are exactly known but it is not the case for the electron-electron interaction and the kinetic energy of the electrons.

A second article from Kohn and Sham[24], in which orbitals are introduced, separates the kinetic energy into a non interaction part, where electrons

are assumed to not interact between each other, and an interaction part treated as a correction to the first one and that can be considered as a kinetic correlation energy.

According to this theory the DFT energy can be written as:

$$E_{DFT} = T[\rho] + E_{en}[\rho] + J[\rho] + E_{XC}[\rho] \tag{2.52}$$

where $E_{XC}$ is the exchange correlation energy containing every terms describing the interaction between electrons not accounted for by the other terms. In the Khon Sham theory, $E_{XC}$ is the only unknown term and thus the aim of the DFT theory is to find how to exactly calculate this functional or, until then, to find a way to compute as accurately as possible these interactions depending on the system studied and the computational power available.

#### 2.3.4.1 Some DFT exchange-correlation functionals

The choice of a specific DFT functional is mainly a choice on how to treat the exchange-correlation term, which is what distinguishes from one DFT functional to another. Although an exact term could in principle exist, it has not been found yet, and the choice of method mainly depends on the system studied, whether it is a molecular or periodic system, an open- or closed shell, as well as the physico-chemical properties one is interested in.

DFT methods are numerous, either ab-initio if derived from first principles, or hybrid if using pre-computed parameters. It is therefore common to use already well establish functionals available in the most well known quantum chemistry programs, and to test the accuracy of these functionals for the system studied against more accurate methods, like CCSD(T), for example.

In the following we will present a brief description of some, mostly, widespread DFT methods, sorted from least accurate to most accurate, depending on the approximation used for the exchange-correlation term of the DFT energy.

PW[25], for Perdew and Wang, is a local density approximation (LDA) functionnal. LDA functional are based on the uniform electron-gas model and depend only on the electron density value at each point in space. They are seldom used for chemical system and are more common in solid state physics for studying metals. The exchange-correlation term is decomposed as:

$$E_{XC} = E_X + E_C \tag{2.53}$$

For all LDA functionals, the exchange term takes the form of the energy of a uniform electron gas:

$$E_X^{LDA}[\rho] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} \int \rho(r)^{\frac{4}{3}}\, dr \tag{2.54}$$

In the case of PW, the LDA approximation is extended for cases where the density of electron is different depending on the spin by using the local spin density approximation (LSDA). Hence, the density of Eq. 2.54 is replaced with the sum of two individual densities, $\rho_\alpha$ and $\rho_\beta$, raised to the power of $\frac{4}{3}$:

$$E_X^{LSDA}[\rho_\alpha, \rho_\beta] = -\frac{3}{4}\left(\frac{6}{\pi}\right)^{\frac{1}{3}} \int \left(\rho_\alpha(r)^{\frac{4}{3}} + \rho_\beta(r)^{\frac{4}{3}}\right) dr \tag{2.55}$$

In the PW functional the correlation energy is defined as:

$$E_C^{PW} = -2A(1+\alpha_1 r_s)\ln\left(1 + \frac{1}{2A\left(\beta_1 r_s^{\frac{1}{2}} + \beta_2 r_s + \beta_3 r_s^{\frac{3}{2}} + \beta_4 r_s^2\right)}\right) \tag{2.56}$$

where A, $\alpha_1$, $\beta_1$, $\beta_2$, $\beta_3$, and $\beta_4$ are fitting parameters, and $r_s = \left(\frac{3}{4\pi}\rho^{-1}\right)^{\frac{1}{3}}$ is the local Seitz radius. LSDA approximation provide reasonable results for metals, but for molecular systems the exchange energy is underestimated by a factor of approximately 10%, leading to errors larger than the exchange energy. Thus, LSDA functionals are rarely used by computational chemists.

PBE[26], for Perdew-Burke-Ernzerhof, is a generalized gradient approximation (GGA) functional, meaning that at each point it uses both the density and its gradient:

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int d^3r\ f\ \ (\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta) \tag{2.57}$$

contrary to the LSDA exchange and correlation terms show in Eq. 2.55 and Eq. 2.56 that only uses the density. In Eq. 2.57, $f(\rho_\alpha, \rho_\beta, \nabla\rho_\alpha, \nabla\rho_\beta)$ contains exchange and correlation elements. These elements vary depending on the functional chosen.

In PBE, the exchange and correlation energies are separated as two different terms, like for PW. The exchange term takes the form:

$$\begin{aligned} E_X^{GGA}[\rho_\alpha, \rho_\beta] &= \int d^3r\ \ (\rho_\alpha + \rho_\beta)\, \epsilon_X^{unif}(\rho_\alpha, \rho_\beta)\, F_X(s) \\ \zeta &= \frac{\rho_\alpha - \rho_\beta}{\rho_\alpha + \rho_\beta} \end{aligned} \tag{2.58}$$

where $\zeta$ the relative spin polarization, $s = |\Delta\rho| / 2 \left(3\pi^2\right)^{1/3} \rho^{4/3}$ is a dimensionless density gradient, $F_X$ an enhancement factor, and $\epsilon_X^{unif}$ is the exchange energy per particle of a uniform electron gas i.e the LDSA exchange energy per particle. Effectively, the PBE exchange energy is equal to the LSDA coupled to an enhancement factor handling deviations from the uniform electron gas. The enhancement factor $F_X$ is defined in the PBE functional as:

$$F_X(s) = 1 + \kappa - \frac{\kappa}{1 + \beta s^2} \tag{2.59}$$

where $\kappa$ and $\beta$ are fitted parameters. In LSDA and LDA functionals, $F_X = 1$.

The correlation energy of PBE is defined as:

$$E_C^{GGA}[\rho_\alpha, \rho_\beta] = \int d^3r \ (\rho_\alpha + \rho_\beta) \left[\epsilon_X^{unif}(r_s, \zeta) + H(r_s, \zeta, t)\right] \tag{2.60}$$

where $t = |\Delta\rho| / 2\phi k_s \rho$ is a dimensionless density gradient like s, and H a gradient contribution working like $F_X$ as an enhancement factor. In t, $k_s$ is the Thomas-Fermi screening wave number, and $\phi$ is a spin-scaling factor. Effectively, the correlation energy of PBE is the LDSA correlation energy with an enhancement factor added to it. In LDSA or LDA functionals, $H = 0$.

TPSS[27], for Tao-Perdew-Staoverov-Scuseria, functional is a Meta-GGA method, an extension of the GGA methods, that can be considered as an improvement over the PBE functional. The meta-GGA functionals depend on the second derivative of the density, also called Laplacian of the density $\nabla^2\rho$, but can also depend on the orbital kinetic energy density:

$$\tau(r) = \frac{1}{2} \sum_i^{occup} |\phi_i(r)|^2 \tag{2.61}$$

a term analog to the Laplacian of the density. For TPSS, the double derivative of the density is included in the exchange term via the enhancement $F_X$, that can be rewritten as:

$$F_X(s) = 1 + \kappa - \frac{\kappa}{1 + \frac{x}{\kappa}} \tag{2.62}$$

where $x$ is chosen as to be dependent on the Laplacian of the electron density $\nabla^2\rho$ or, equivalently, on the orbital kinetic energy $\tau$.

The correlation energy is a minor change to another meta-GGA functional call PKZB[28], for Perdew–Kurth–Zupan–Blaha. Functional that improved other the PBE correlation energy by mainly changing the enhancement factor H by including $\nabla^2\rho$ or, equivalently, $\tau$.

B3LYP[29–32] is a hybrid DFT functional derived from a non-hybrid GGA functional called B-LYP. As it is called an "hybrid" method, it uses in parts the exact exchange term of the Hartree-Fock theory seen in eq 2.25 with a combination of other exchange-correlation term that are either ab-initio or using pre-computed parameters. The B3LYP functional is one a the most popular functional in quantum chemistry and its name stands for Becke, 3 parameters, Lee, Yang, Parr.

Its exchange-correlation term is written as[32]:

$$E_{xc}^{B3LYP} = aE_x^{HF}\,(1-a)E_x^{LSDA} + b\Delta E_x^{B88} + cE_c^{LYP} + (1-c)\,E_c^{LSDA} \tag{2.63}$$

where $a$, $b$, and $c$ are the three parameters of value 0.20, 0.72, and 0.81 respectively, taken from Becke original fitting of the B3PW91 functional[29]. $E_c^{LSDA}$ is the exchange-correlation energy computed with the Local Density Approximation (LDA) that depends only on the electron density of each point in space, and modified to take into account the spin value of the electrons as seen previously with the PW functional; $E_x^{B88}$ is the Becke 88 exchange functional[29] with gradient correction; $E_c^{LYP}$ is the Lee, Yang, and Parr correlation functional[30].

B3LYP is the functional of choice for a number of computational chemistry teams as it has been benchmark for a multitude of system over the decades it has been in use. Like other DFT methods, it fails to describe accurately long-range electron correlations responsible for dispersion forces. Different types of methods exist to correct the DFT energy for dispersion interaction. For the B3LYP functional, it is mostly use in its B3LYP-D3 declination which incorporate the DFT-D3 correction[33,34] that adds an $R^{-6}$ energy term for each atom pair.

### 2.3.5 Semi-empirical quantum mechanical methods

DFT methods are overall accurate, but the computational cost increases exponentially with the size of the system (with the $2^{nd} - 3^{rd}$ power). Semi-empirical quantum mechanical (SQM) methods contain, as their name suggests, empirical data computed using more accurate methods, and are based

on the Hartree-Fock formalism. This means that the bi-electronic integrals appearing in the Roothann-Hall equations are reduced by drastic approximations, bringing the computational cost to increase almost linearly with the system size. However, the price to be paid is in terms of the accuracy, and, as semi-empirical methods are parametrized on specific systems and to solve targeted chemical problems, a minor general applicability with respect to DFT.

Molecular orbitals in *ab-initio* methods are represented as linear combination of atomic orbitals like seen in Eq. 2.41. In turn, the atomic orbitals are themselves expanded for all electrons in the basis sets, meaning that both core electrons, *i.e.* those not participating in chemical bonding, and valence electrons, *i.e.* those participating in chemical bonding, are represented in the basis set. Semi-empirical methods generally expand the basis set only to represent the valence electrons of the system, treating the core electrons as being part of the nuclei field.

In Eq. 2.41, the coefficients $C_{\alpha i}$ are computed by solving in a self-consistent way the Roothaan-Hall equations, that can be written into a generalized eigenvalue problem:

$$\mathbf{FC} = \mathbf{SC}\epsilon \tag{2.64}$$

where $\boldsymbol{F}$ is the Fock matrix, $\boldsymbol{C}$ is the molecular orbitals coefficients matrix, $\boldsymbol{S}$ is the overlap matrix, and $\boldsymbol{\epsilon}$ is a diagonal matrix containing the energies of the orbitals. The Fock matrix $\boldsymbol{F}$ can be separated into two parts: the one electron part $h$, and two electrons parts $J$ and $K$, all written in Eq. 2.25, which can be rewritten into the semi-empirical form where $\boldsymbol{G}$ represents the two electrons part:

$$\begin{aligned} F_{\mu\nu} &= h_{\mu\nu} + G_{\mu\nu} \\ h_{\mu\nu} &= \langle\mu| - \frac{1}{2}\nabla^2 |\nu\rangle - \sum_a^M Z_a \langle\mu| \frac{1}{|\boldsymbol{R}_a - \boldsymbol{i}|} |\nu\rangle \\ G_{\mu\nu} &= \sum_{\lambda\sigma}^{M_{basis}} D_{\lambda\sigma} \left( \langle\mu\nu \,|\, \lambda\sigma\rangle - \frac{1}{2} \langle\mu\lambda \,|\, \nu\sigma\rangle \right) \end{aligned} \tag{2.65}$$

Approximating these matrix elements and the integrals involved in the resolution of the Roothan-Hall equations is the main aim of SM methods[35]. These approximation involving the calculation of integrals are numerous, the main one being the zero-differential overlap (ZDO) approximation.[36] It involves two electrons integrals of the form $\langle\mu\nu \,|\, \lambda\sigma\rangle$ and calculates only the

two-electrons integrals of the type $\langle \mu\mu \,|\, \nu\nu \rangle$ reducing the two electrons part of the Fock operator from $N^4$ to $N^2$ in the order of number of integrals.

With the ZDO approximation and the truncation of the basis to only treat valence electrons, these methods represents a further approximation with respect to the HF theory, which turn lacks of accuracy as we already highlighted before. To recover part of the lost accuracy, without increasing the computing time, semi-empirical methods replace parts of the missing integrals by parameters obtained empirically, either experimentally or via ab-initio computation methods like CCSD.

#### 2.3.5.1 DFTB

Density Functional Tight Binding (DFTB) is a generalization of the DFT method where electrons are supposed to be tightly bound to the atom they belong to, and to have limited interactions with surrounding atoms of the system. As a class of SQM methods, DFTB uses a minimal AO basis set, treating exclusively valence electrons, and using integral approximations to determined AO Hamiltonian matrix elements beforehand.

The DFTB model is based on a reference density $\rho_0$ representing a superposition of densities from neutral atoms[35]

$$\rho_0 = \sum_A \rho_0^A \tag{2.66}$$

From this reference density, the DFT total energy is rewritten using a density deviation $\delta\rho$ from the $\rho_0$ reference density, i.e. it is written as a Taylor expansion of the DFT energy around the reference density[37,38]:

$$E[\rho] = E^0[\rho_0] + E^1[\rho_0, (\delta\rho)] + E^2\left[\rho_0, (\delta\rho)^2\right] + E^3\left[\rho_0, (\delta\rho)^3\right] + \cdots \tag{2.67}$$

Depending on the order at which Eq. 2.67 is truncated, the DFTB model will be classified as DFTB1, DFTB2 or DFTB3. After several approximations to the integrals, the approximate DFTB3 total energy can be written as:

$$\begin{aligned} E[\rho] &= E^0[\rho_0] + E^1[\rho_0, (\delta\rho)] + E^2\left[\rho_0, (\delta\rho)^2\right] + E^3\left[\rho_0, (\delta\rho)^3\right] \\ &= \frac{1}{2}\sum_{AB} V_{AB}^{rep} + \sum_{iAB}\sum_{\mu\in A}\sum_{\nu\in B} n_i C_{\mu i} C_{\nu i} H^0_{\mu\nu} \\ &\quad + \frac{1}{2}\sum_{AB} \Delta q_A \Delta q_B \gamma^h_{AB} + \frac{1}{3}\sum_{AB} (\Delta q_A)^2 \Delta q_B \Gamma_{AB} \end{aligned} \tag{2.68}$$

The first term, $E^0[\rho_0]$, depends only on the reference density, and is a repulsive potential that takes the form of a short range function. This potential does not affect strongly non-bonding properties but it is critical for the covalent bonding properties such as bond lengths and therefore the associated vibrational frequencies. The second term, $E^1[\rho_0, (\delta\rho)]$, involves the first order $\delta\rho$ fluctuation term and it is essential to describe covalent bonds in tight binding methods. The third term, $E^2\left[\rho_0, (\delta\rho)^2\right]$, is a second order self-consistent charge (SCC) term describing the Coulomb interaction between atomic partial charges. The fourth term, $E^3\left[\rho_0, (\delta\rho)^3\right]$, is a third order SCC term to help describing a more diffuse electron density for non neutral atoms (in particular di anions).

The DFTB formalism is well documented and widely used for SQM methods. The efficacity of the SQM methods using this formalism depends on the parametrization used and the definition of certain formula and functions such as $V_{AB}^{rep}$, the core-core repulsion, or $\gamma_{AB}$, an analytical function used in one of the charge-charge interaction terms, that interpolates between values depending whether the two charges are separated by a long distance or a very small distance.

Among the most modern incarnation of DFTB methods it is worth mentioning the GFN-xTB (for Geometry Frequency Non-covalent extended Tight Binding), a semi-empirical quantum mechanical method developed by the Grimme's group at the Bonn University using the DFTB3 formalism. It will be presented in details in a specific sections as it was the method of choice for many of the calculations carried out during the present work.

## 2.4 Non QM methods

### 2.4.1 Force Fields

Force Fields (FF) methods, are based on molecular mechanics (MM) methods, and treat atoms with force fields. They are not based on quantum mechanical laws, thus electrons are not considered at all, only atoms are. Like QM methods in the BO approximation, the movements of these atoms is treated classically.

According to the above-mentioned features of MM-based methods, geometry information like covalent bonds must be given explicitly instead of being obtained by solving the Schrödinger equation. The electronic energy

becomes a parametric equation of the atomic coordinates and the parameters are fitted to experimental or higher level computational data.

MM-based methods describe molecules by a "ball and spring" model where atoms are defined by their radius and electronic charge, and bonds by their length and rigidity[39–43]. In the Force-Field scheme, molecules are seen as composed by structural units that share approximately the same value between different molecules (in a similar way as functional groups are considered in organic chemistry). As an example the bond length between an atom A and B present in two different molecules share the same features.

The force-field energy can be written as[1]:

$$\begin{aligned} E = & \quad E_{bonded} \quad + \quad E_{non-bonded} \\ = & \overbrace{E_{stretch} + E_{bend} + E_{tors} + E_{cross}} + \overbrace{E_{vdw} + E_{elec}} \end{aligned} \tag{2.69}$$

The above equation is made by two parts, the first one describing interactions between bonded atoms ($E_{bonded}$), and the second part describing interactions between non-bonded atoms ($E_{non-bonded}$). In the bonded part, the first three terms describe the energy needed to stretch, bend, and rotate a covalent bond. The fourth term is the coupling between these first three energies. In the non-bonded part, the first term is the van der Waals energy ($E_{vdw}$) describing the repulsion and attraction between atoms that are not bonded by covalent bonds. The second term is an electrostatic energy ($E_{elec}$) describing attraction and repulsion between atoms and molecules due to charge distribution in molecules.

These energies forming the bonded and non bonded part of the FF energy are not universal and vary from one FF to another. For example some force fields methods do not use the coupling term $E_{cross}$, or use more than one, others add correction terms for hydrogen or halogen bonds. Force Fields methods can also vary in what information they use for fitting the parameters, and how they implements the energy terms. For example the van der Walls energy can be written as a Leonard-Jones (LJ) potential[44]:

$$E_{LJ}(R) = \epsilon \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^{6} \right] \tag{2.70}$$

where $\epsilon$ is the dispersion energy, the energy needed to separate the atoms, $R_0$ is the minimum energy distance, the distance at which $E_{LJ}$ is equal to 0, and $R$ is the distance between the two atoms considered. The first member

of Eq. 2.70 represents the repulsive contribution to the energy, the second one the attractive contribution. This is a popular way of describing the vdW interaction, but it can also be described as a "Hill" type potential[45]:

$$E_{Hill}(R) = Ae^{-BR} - \frac{C}{R^6} \tag{2.71}$$

where A, B, and C are constant. Here the repulsive part is described using an exponential function.

A widely known Force-Field method used to describe water quickly, and at good accuracy, is the Transferable Intermolecular Potential Functions[46] (TIPS). Since its first publication, the TIPS method has been expanded into a multitude of derivative methods. Among them, it can be found a 3-sites water model based on LJ potential and Coulomb interactions. Two interaction sites are placed on the hydrogen atoms and with a charge $q_H$, the last interaction site is placed on the oxygen atom with a given charge of $q_O = -2q_H$. It was re-parametrized with the TIP3P method, and modified into a 4-sites model with the TIPS2[47] method and TIP4P[48], and a 5-sites model with TIP5P[49]. For the 4-sites TIPS methods, the oxygen charge is moved off from the oxygen atom and placed toward the hydrogen atoms on the bisector of the HOH angle. The 5-sites model modifies again the position of the oxygen charge and changes it into two lone pair oriented tetrahedrally around the oxygen. As of today, the TIP3P method is the most popular, as it is cheaper in computational power than the 4-sites and 5-sites models. The Coulomb and Lennard-Jones interactions of the TIP models are applied between all intermolecular pairs of sites:

$$\epsilon_{mn} = \sum^{on}_{m} \sum^{on}_{n} \left( \frac{q_i q_j e^2}{r_{ij}} + \frac{A_i A_j}{r_{ij}^{12}} - \frac{C_i C_j}{r_{ij}^{6}} \right) \tag{2.72}$$

$\epsilon_{mn}$ is the interaction energy between monomers $m$ and $n$; $q$, $A$, and $C$ are three parameters empirically determined for each site to obtain reasonable energetic and structural results.

Force-Field methods possess advantages, like the relative speed of calculations, and the computational power needed for these calculations (as of today it is possible to perform computation on very large systems with a simple personal laptop in very little time). But also disadvantages, the main one being the over specialization of the parameters, making the method suitable for only a very limited number of molecule and atomic types. For example, the TIPS methods cannot treat other atoms than H and O, and in the

specific form of $H_2O$. This disadvantage is not true for every FF methods. UFF (add citation) and the more recent GFN-FF, a FF method belonging to the GFN-xTB set of methods and which will be talked about more in the following sections, are parametrized for atomic types up to Z=86, and has been extensively tested for various molecules.

## 2.5 Computational algorithms

Up to now what was presented was limited to single point calculations, i.e. computing the energy and wave function of a system depending on its parameters at a specific geometry. Single point calculations are not the only types of calculation available to theoretical chemists, and computing the energy of a system and its wave function is a first step leading to the determination of several molecular properties, *e.g.* global and local minima of the PES or vibrational frequencies, depending on the order of the derivative of the energy with respect to the needed variable, and other properties such as terms added to the Hamiltonian, the type of wave function, etc.

### 2.5.1 Geometry

A single point calculation performed to obtain the energy of our system is done on a fixed structure. By finding stationary points, both local minima and first order saddle points, on the PES, it is possible, via theoretical chemistry computation, to find structures in transition states or stable states, i.e. in state where the inter-atomic forces for every atoms are close to 0. On the PES, local minima are stable states, and first order saddle points are transition states (TS). Perturbing slightly a system geometry out of its stable state will induce a change to reduce the perturbation and bring the system back to the stable state. Perturbing slightly a system geometry out of its transition state will bring the system either to another TS or to a stable state.

The energy $E$ of a system, obtained through the methods explained previously, is a parametric function of the atoms coordinates. Thus, it is possible to expand the energy $E$ into a Taylor series of $\boldsymbol{R}$ where $\boldsymbol{R} =$

$(X_1, X_2, X_3, \ldots, X_{3N})$ are the atomic coordinates:

$$E(\boldsymbol{R}) = E(\boldsymbol{R}_0) + \boldsymbol{g}(\boldsymbol{R} - \boldsymbol{R}_0) + \frac{1}{2}\boldsymbol{H}(\boldsymbol{R} - \boldsymbol{R}_0)^2 + \cdots$$
$$\boldsymbol{g} = \frac{\partial E}{\partial \boldsymbol{R}};\ \boldsymbol{H} = \frac{\partial^2 E}{\partial \boldsymbol{R}^2} \tag{2.73}$$

where $\boldsymbol{g}$ is the energy gradient and $\boldsymbol{H}$ is the second derivative of the energy, in short the Hessian matrix. The gradient $\boldsymbol{g}$ is a vector pointing in the direction where the energy function increases the most. The Hessian $\boldsymbol{H}$, is a matrix that, through its diagonalization, and in tandem with the gradient, indicates the step direction to take in order to minimize the energy. Through a series of iterative steps, the geometry optimization will find a geometry $\boldsymbol{R}_N$ for which the energy $E$ is at a minimum, meaning that $\boldsymbol{g} = 0$. As diagonalizing the Hessian matrix is computationally demanding, much more than computing the gradient, its full calculation is mostly reserved to frequency computation (presented in the next subsection). Thus, diverse methods for finding stationary points have been developed, which either do not need the Hessian matrix, or can use an approximation. The Steepest Descent (SD), or gradient descent, method is one of them. It aims to find a local minima to the function through a sequence of iterative steps in the opposite direction of the gradient. When the function starts to increase again, a new gradient is computed and the process continues. This technique is very robust and do not need the Hessian matrix. A local minima will always be found if the process is done accurately, but the minima will be approached at an increasingly slower speed, slowing down the convergence. Another method is the Newton-Raphson method, were the energy is expanded into a Taylor series like Eq. 2.73 but truncated after the second order derivative. It creates a sequence of guesses $\{R_k\}$ starting from an initial one $R_0$, and through a second-order Taylor expansion of E, converges towards the local minima $R_N$ of E.

$$E(\boldsymbol{R_k}) = E(\boldsymbol{R}_k) + \boldsymbol{g}(\boldsymbol{R} - \boldsymbol{R}_k) + \frac{1}{2}\boldsymbol{H}(\boldsymbol{R} - \boldsymbol{R}_k)^2 \tag{2.74}$$

The next iteration minimizes Eq. 2.74 in $(\boldsymbol{R} - \boldsymbol{R}_k)$, which is equivalent to setting its derivative with respect to $(\boldsymbol{R} - \boldsymbol{R}_k)$ to 0, giving:

$$\boldsymbol{g} + \boldsymbol{H}(\boldsymbol{R} - \boldsymbol{R}_k) = 0 \tag{2.75}$$

which leads to:

$$(\boldsymbol{R} - \boldsymbol{R}_k) = -\boldsymbol{H}^{-1}\boldsymbol{g} \tag{2.76}$$

the step of iteration performed by the Newton-Raphson method.

Contrary to the gradient descent method, the Newton-Raphson method needs the Hessian to compute the iterative steps, but, as said previously, computing the Hessian matrix is a very costly ordeal. One way to do it approximately is through the updating scheme used in Hessian update methods. In these methods, the Hessian is approximated (as a unit matrix for example) and updated after each steps of the minimization of the energy using the gradients of the previous and current coordinates, giving good approximation to the Hessian matrix in the direction of the steps after just a few iterations.

$$\boldsymbol{H}_k = \boldsymbol{H}_{k-1} + \Delta\boldsymbol{H} \tag{2.77}$$

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is an updating scheme used commonly to compute $\Delta\boldsymbol{H}$:

$$\Delta\boldsymbol{H}_{BFGS} = \frac{\Delta\boldsymbol{g}\Delta\boldsymbol{g}^t}{\Delta\boldsymbol{g}^t\Delta\boldsymbol{R}} - \frac{\boldsymbol{H}\Delta\boldsymbol{R}\Delta\boldsymbol{R}^t\boldsymbol{H}}{\Delta\boldsymbol{R}^t\boldsymbol{H}\Delta\boldsymbol{R}} \tag{2.78}$$

where $\Delta\boldsymbol{g} = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$, and $\Delta\boldsymbol{R} = \boldsymbol{R}_k - \boldsymbol{R}_{k-1}$.

### 2.5.2 Harmonic Frequency Calculation

The second derivative of the energy as a function of the geometry is needed to calculate the Hessian matrix (Eq. 2.73). Diagonalization of $\boldsymbol{H}$ gives as Eigenvectors the normal modes of vibrations which lead to vibrational frequencies of our system. Through the vibrational frequencies and their corresponding intensities one obtains the Infrared (IR) spectrum, an important data to compare theoretical and experimental results. In addition, the sign of the Hessian matrix eigen values can indicate if the stationary point is a TS (one negative eigen value) or a stable geometry (only positive eigen values). The diagonalization of $\boldsymbol{H}$ leads to a unitary transformation from the cartesian coordinates $\boldsymbol{R}$ (x,y, and z for each atom, or $3N_{atom}$) to a vibrational normal coordinates $\boldsymbol{q}$ ($3N_{atom} - 5$ or 6 for linear/not linear molecules) with eigenvalues $\epsilon_i$. The vibrational frequencies $\nu_i$ are linked to the eigenvalues $\epsilon_i$ of the Hessian matrix by $\nu_i = \sqrt{\epsilon_i}/2\pi$.

The vibrational Hamiltonian of the system is written as[1,50]:

$$\boldsymbol{H}_{vib} = -\sum_{i=1}^{3N_{atom}} \left( \frac{1}{2}\frac{\partial}{\partial q_i^2} + \frac{1}{2}\epsilon_i q_i^2 \right) \tag{2.79}$$

The solution of Eq. 2.79 gives the energy levels $E_v$ associated to each frequency $\nu_i$ as:[1,50]

$$E_v(\nu_i) = \left(v + \frac{1}{2}\right)\hbar\nu_i \tag{2.80}$$

in which $v$ is the vibration quantum number. As a result, the minimum energy of the $i$-th harmonic oscillator is not zero but $\frac{1}{2}\hbar\nu_i$. Therefore, by summing up the energies associated to all normal modes, we obtain the zero-point energy (ZPE)[1,50]:

$$E_{ZPE} = \sum_{i}^{N_{modes}} \frac{\hbar\nu_i}{2} \tag{2.81}$$

By adding the ZPE energy to the electronic energy we obtain the zero-point corrected energy (ZPVE) $E_0$, i.e. the energy of the system at 0K[50].:

$$E_0 = E_{elec} + E_{ZPE} \tag{2.82}$$

In the ISM, and in particular in dense molecular clouds, the temperature is as low as 10 K and, therefore, the $E_{ZPE}$ is the only important correction to the $E_{elec}$.

### 2.5.3 Molecular dynamics

Molecular Dynamics (MD) is the study of the time evolution of a given system by solving the specific equation of nuclear motion (either quantum or classical) over time. Basically, due to the very complex nature of a true quantum mechanical treatment, and considering the almost classical nature of the nuclei motion, the latter is studied by solving the Newton's second law of motion[1,51–53] $\boldsymbol{F} = m\boldsymbol{a}$ which can be written as:

$$-\frac{dV}{d\boldsymbol{r}} = m\frac{d^2\boldsymbol{r}}{dt^2} \tag{2.83}$$

where $\boldsymbol{r}$ is the set of $3N_{atom}$ coordinates for each nucleus, and V is the potential energy of the system at which the nuclear motion is subjected at a specific configuration $\boldsymbol{r}$.

In MD, time is discretized on a finite grid with time step $\Delta t$, and at each step, the Newton equations are solved. By knowing the positions $\boldsymbol{r}$ and velocities $\boldsymbol{v}$ of the system at a given $t$, it is possible to compute the state of the system at a time $t + \Delta t$ and iterate this procedure until the desired

simulation time length is reached, usually of the order of a few picoseconds ($10^{-12}\ s$). The choice of the time step $\Delta t$ depends on the properties of the studied system, but should be smaller of the smallest time taken by some degrees of the systems to be changed. For instance, the time step should be smaller (at least one order of magnitude) of the time taken by the fastest motion to account for a vibration (typically the stretching vibration involving hydrogen atoms). Through a given set of positions at a step i ($\boldsymbol{r}_i$) it is possible to find the position $\boldsymbol{r}_{i+1}$ after, and before, a small time step $\Delta t$ (in general $10^{-15}\ s$)[54], via a Taylor expansion of $\boldsymbol{r}_i\,(t)$[1]:

$$\begin{aligned}\boldsymbol{r}_{i+1} &= \boldsymbol{r_i} + \frac{\partial \boldsymbol{r}}{\partial t}\,(\Delta t) + \frac{1}{2}\frac{\partial^2 \boldsymbol{r}}{\partial t^2}\,(\Delta t)^2 + \frac{1}{6}\frac{\partial^3 \boldsymbol{r}}{\partial t^3}\,(\Delta t)^3 + \cdots \\ \boldsymbol{r}_{i-1} &= \boldsymbol{r_i} - \frac{\partial \boldsymbol{r}}{\partial t}\,(\Delta t) + \frac{1}{2}\frac{\partial^2 \boldsymbol{r}}{\partial t^2}\,(\Delta t)^2 - \frac{1}{6}\frac{\partial^3 \boldsymbol{r}}{\partial t^3}\,(\Delta t)^3 + \cdots\end{aligned} \tag{2.84}$$

By adding both equations, we obtain the Verlet algorithm[55]:

$$\boldsymbol{r}_{i+1} = 2\boldsymbol{r}_i - \boldsymbol{r}_{i-1} + \boldsymbol{a}_i\,(\Delta t)^2 + \cdots \tag{2.85}$$

where $\boldsymbol{a}_i = \frac{\partial^2 \boldsymbol{r}}{\partial t^2}$ is the acceleration which can be computed through Eq. 2.83 as:

$$\boldsymbol{a} = -\frac{1}{m}\frac{\partial \boldsymbol{V}}{\partial \boldsymbol{r}} \tag{2.86}$$

The potential $V$ depends on the coordinates $\boldsymbol{r}_i$ and thus needs to be computed again for each steps. It can be either through a force field as in the classical MD, or it can be computed on the fly for each time step quantum mechanically usually using the DFT methods (AIMD).

In MD, systems can be describes by what is called statistical ensembles[56], also called thermodynamic ensembles, that depend on some observable parameters:

- Microcanonical ensemble ($NVE$ ensemble): systems in which the total energy (E), the number of particles (N), and the volume (V) are fixed.
- Canonical ensemble ($NVT$ ensemble): Same as $NVE$ but instead of the energy it is the temperature (T) that is kept constant.
- Grand canonical ensemble ($\mu VT$ ensemble): Here, the chemical potential ($\mu$) is kept fixed, with the volume and the temperature.

- Isobaric-isothermal ensemble ($NPT$ ensemble): The pressure is kept constant with the temperature and number of particles.
- isoenthalpic-isobaric ensemble ($NPH$ ensemble): Same as $NPT$ but instead of the temperature it is the enthalpy (H) that is kept constant.

The most widely used ensembles in MD simulations are $NVT$, $NPT$, and $NVE$.

In Astrochemistry, MD simulations are mostly used for computing sticking coefficients or simulate surface diffusion of the adsorbates of interest.

### 2.5.4 QM/MM and ONIOM method

Modeling large chemical systems can be highly demanding with QM methods. Semi-empirical methods exist to reduce the computational resources needed, but at the cost of reducing the accuracy. Through the years, other ways of dealing with large systems have been developed like hybrid methods, and more precisely the QM/MM method.

The QM/MM approach, formalized by 1976 by Warshel and Levitt[57], divides the system in two parts each of them treated at a different level of theory, as the name of this approach suggests. The fragment of chemical interest is treated with the QM high-level method of choice (depending on the chemical properties one is interested in treating), while the other part of the system with a MM low-level method. In the QM/MM method, the total energy of the system is based on ad additive approach:

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM-MM} \tag{2.87}$$

where $E_{QM}$ is the energy of the high-level part treated with the QM method, $E_{MM}$ is the energy of the low-level part treated classically, and $E_{QM-MM}$ is the energy of the interaction between the QM and MM parts. In the QM/MM scheme, the QM-MM interactions are treated essentially classically by the QM-MM Hamiltonian.

The ONIOM method, developed by Morokuma and coworkers and first introduced in 1995 by Maseras et al.[58], is based on a different idea. Contrary to the previous method, it is not limited to only two fragments and the fragments are not forced to be QM and MM. Each fragment of the system can be treated with any method of choice, being it QM, semi-empirical or MM. At variance with QM/MM, the ONIOM approach is not an additive

method, but a substractive one. For a two fragments system, it defines a "model" part of the system, where the properties of interest are present, and a "real" part, which is the entire system. Thus, the fragment of interest is the "model" fragment, and the fragment treated at a lower level of theory is a subtraction of the "model" part from the "real" part. For a "model" part treated as QM, and a "real" part treated as MM, the total energy of the system is written as:

$$E_{ONIOM} = E_{QM,model} + E_{MM,real} - E_{MM,model} \tag{2.88}$$

where $E_{QM,model}$ is the energy of the "model" fragment treated quantum mechanically, $E_{MM,real}$ is the energy of the entire system computed with the MM method (in this specific case), and $E_{MM,model}$ is the energy of the "model" fragment but computed at the MM level of theory. The subtraction of $E_{MM,model}$ to $E_{MM,real}$ allows the computation of the interactions between the QM and MM part, thus avoiding the need for a QM-MM Hamiltonian.

## 2.6 Extended tight binding: the xTB code

### 2.6.1 General presentation

As explained previously, xTB (which stands for eXtended Tight Binding) is comprised of a set of methods developed by Grimme's group at the Bonn University. The methods available in the xTB code are called GFN (Geometry, Frequency, Non-covalent interactions), representing the properties the code aims at treating accurately at a reduced computational cost. Among the GFNs there are 3 SQM methods (GFN0, GFN1, and GFN2) representing a semiempirical approximation to the Kohn-Sham DFT and, in the specific, they are an improvement over the DFTB formalism, and a force field method (GFN-FF). These methods are "general purpose" SQM methods, meaning that they are applicable to a wide range of systems and parametrized for the majority of elements in the periodic table (all *spd* block elements and the lanthanides up to Z = 86). Mainly, the target of the xTB methods is to: (i) find structural properties around the equilibrium (vibrational frequencies and non-covalent interactions); (ii) adopt an extended AO basis made of standard primitive Gaussian functions approximating Slater functions (STO-mG) with a better description of hydrogen bonding; (iii) use *d*-polarization functions for higher row elements; iv) keep the number of parameters to a minimum by

avoiding pair-specific potentials and focusing on global and element specific parameters, making it easy to parametrize[59].

The methods have no spin-dependent terms in their respective Hamiltonian and are formulated in a spin-restricted way leading to identical spatial MO and orbital energies for $\alpha$ and $\beta$ orbitals, but possibly different orbital occupation. This entails that the GFNn-xTB Hamiltonian will always favor low-spin configurations[59].

In the following section we will present succinctly the different GFN-xTB methods in the chronological order of their publication. Figure 2.1 shows a summary of the main components of the GFN-xTB methods.

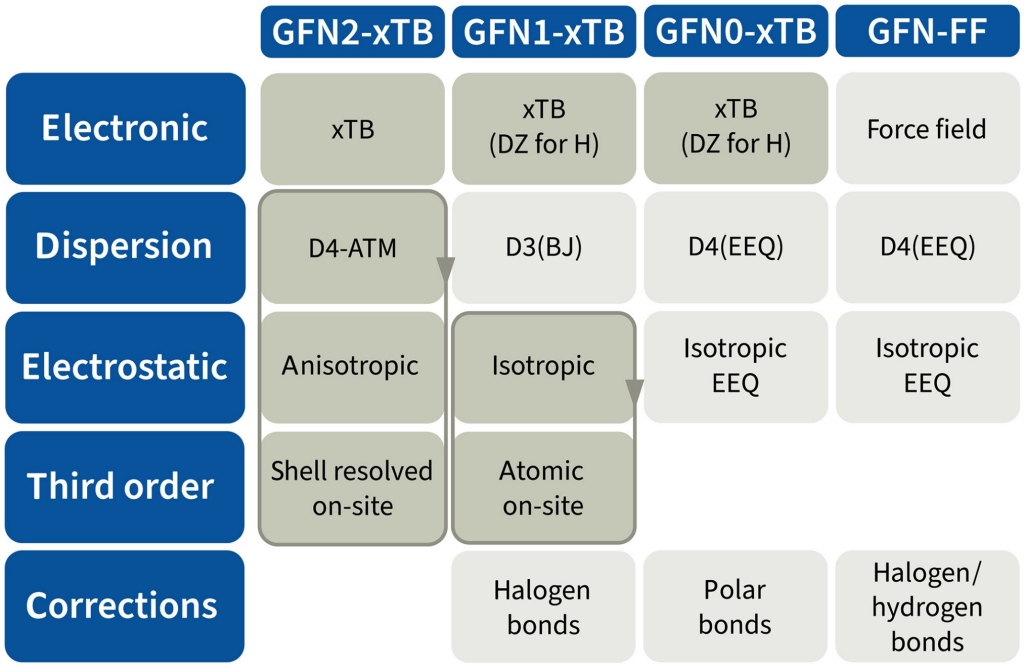



**Figure 2.1:** Overview of the GFN family of methods with main components and classification of the most important terms. Dark gray shaded areas denote a quantummechanical description while light gray parts indicate a classical or semi-classical description. The parts surrounded by the arrows are treated in an iterative, self-consistent fashion. Provided by Bannwarth et al. 2021[59] through the Creative Commons 4.0 license.

### 2.6.2 GFN1-xTB

GFN1-xTB[60] was the first method to be published by Grimme's group with the name of GFN-xTB at the time, as it was the only GFN method available. The total energy of GFN1-xTB follows the DFTB3 formalism and is defined as follows:

$$\begin{aligned} E_{GFN1} &= E_{rep}^{(0)} + E_{disp}^{(0)} + E_{XB}^{(0)} + E_{EHT}^{(1)} + E_{IES+IXC}^{(2)} + E_{IES+IXC}^{(3)} \\ &= E_{rep} + E_{disp}^{D3} + E_{XB}^{GFN1} + E_{EHT} + E_{\gamma} + E_{\Gamma}^{GFN1} \end{aligned} \tag{2.89}$$

where $E_{rep}$ is the atom pairwise repulsive term found in all GFN-type methods, $E_{disp}^{D3}$ is the dispersive contribution computed through the D3 dispersion model, $E_{XB}^{GFN1}$ is a corrective term for weak halogen bonds. $E_{EHT}$ is an extended Hückel type (EHT) term allowing for covalent bond formation. $E_{\gamma}$ is an isotropic electrostatic and exchange (XC) term originated from the second-order term in the tight-binding expansion (Eq. 2.67). $E_{\Gamma}^{GFN1}$ is a third order electrostatic/XC correction allowing for the stabilization partial charges for highly electronegative elements such as O. No additional hydrogen bonds corrective term is introduced in GFN1-xTB, apart an extra s-AO function on hydrogen atoms that better describe the hydrogen bond. Due to the monopole approximation of GFN1-xTB, it necessary neglects the dipole, quadrupole, and higher contribution to the differential charge densities $\delta\rho$. The dispersion energy is obtained via the well-known DFT-D3 method[33,34]:

$$E_{disp}^{D3} = E_{KS\_GFN} - E_{disp\_corr} \tag{2.90}$$

where $E_{KS\_GFN}$ is the self-consistent KS energy obtained through the GFN method, $E_{disp\_corr}$ is the dispersion energy correction defined as:

$$E_{disp\_corr} = E^{(2)} + E^{(3)} \tag{2.91}$$

*i.e.* the sum of two body and three body energies.

The molecular orbitals used to solve the Hamiltonian equation are linear combinations of atom-centered orbitals (LCAO):

$$\Psi_i = \sum_{\mu}^{N_{AO}} c_{\mu i} \phi\left(\zeta, STO - mG\right) \tag{2.92}$$

where the number of primitive $m$ depends on the AO and varies from 3 to 6. The AO basis set used is tied to the specific version of GFNn-xTB as

each basis set has been optimized *ad-hoc* for the corresponding GFNn-xTB parametrization. The variational minimization of the GFN energy with respect to the linear $c_{\mu i}$ coefficients leads again to the general eigenvalue problem $\mathbf{FC} = \mathbf{SC}\epsilon$ already seen in Eq. 2.64, *i.e.* the Roothaan-Hall equations. The matrix elements of the GFN1-xTB hamiltonian are computed as:

$$\begin{aligned}\langle\mu|\, F\,|\nu\rangle = \langle\mu|\, H_0\,|\nu\rangle + \frac{1}{2}S_{\mu\nu}\sum_C\sum_{l''}\left(\gamma_{AC,ll''}+\gamma_{BC,l'l'}\right)p_{l''}^C \\ + \frac{1}{2}S_{\mu\nu}\left(q_A^2\Gamma_A + q_B^2\Gamma_B\right)\ \left(\mu\in l\left(A\right),\nu\in l'\left(B\right)\right)\end{aligned} \tag{2.93}$$

where $\mu$ and $\nu$ are atomic orbitals with angular momenta $l$ and $l'$ respectively, $H_0$ is the zero-order hamiltonian defined as:

$$\begin{aligned}\langle\mu|\, H_0\,|\nu\rangle =& \frac{K_{AB}}{4}\left(k_l+k_{l'}\right)\left(h_A^l+h_B^{l'}\right)S_{\mu\nu}\left(1+k_{EN}\Delta EN_{AB}^2\right) \\ & \Pi\left(R_{AB,ll'}\right)\ \left(\mu\in l\left(A\right),\nu\in l'\left(B\right)\right)\end{aligned} \tag{2.94}$$

where $k_l$ and $k_{l'}$ are free parameters, $h_A^l$ and $h_B^{l'}$ are effective atomic energy levels, $\Delta EN_{AB}^2 = EN_A - EN_B$ is the electronegativity difference of atoms A and B, $K_{AB}$ is a scaling constant, $\Pi\left(R_{AB,ll'}\right)$ is a function dependent on distance and $l$.

All in all, GFN1-xTB possesses 16 global empirical parameters, and approximately 1000 element specific parameters for atomic number up to Z = 86. Most of the molecular structures used for obtaining these parameters were optimized at the PBEh-3c DFT level. PBE0-D3(BJ)/def2-TZVP levels were used for the rare gas systems. For each fitted parameter, 50 to 100 reference data points were used, including equilibrium geometries, slightly distorted structures, harmonic vibrational frequencies, and others.

GFN-xTB proved to be somewhat accurate for a large variety of chemical systems but a more robust method called GFN2 was published a few years later giving better results with respect to to GFN1.

### 2.6.3 GFN2-xTB

The GFN2-xTB[61] Hamiltonian is also based on DFTB3 and thus it is very similar to GFN1-xTB. It differentiates mostly for two points: first it does

not need halogen or hydrogen bonding corrections, and second, in the multipole expansion it includes electrostatic interactions and exchange-correlation effects up the second order. It can be written as such:

$$\begin{aligned} E_{GFN2} &= E_{rep}^{(0)} + E_{disp}^{(0),(1),(2)} + E_{EHT}^{(1)} + E_{IES+IXC}^{(2)} + E_{AES+AXC}^{(2)} + E_{IES+IXC}^{(2)} \\ &= E_{rep} + E_{disp}^{D4'} + E_{EHT} + E_{\gamma} + E_{AES} + E_{AXC} + E_{\Gamma}^{GFN2} \end{aligned} \tag{2.95}$$

Here most the terms are similar to the GFN1 Hamiltoniam seen in Eq. 2.89 In addition, it includes an anisotropic electrostatic term $E_{AES}$ and an anisotropic XC term $E_{AXC}$. The dispersion term is charge dependent and not based on the D3 model seen previously, but on a modified D4 dispersion model. The D4 dispersion is written as[62]:

$$E_{disp}^{D4} = E_{disp}^{(6,8)} + E_{disp}^{(9),ATM} \tag{2.96}$$

where $E_{disp}^{(6,8)}$ is a two-body London dispersion energy term including dipole-dipole ($n = 6$) and dipole-quadrupole ($n = 8$) interactions (this term is approximately equivalent to the $E_{disp}^{D3}$ of GFN1-xTB). $E_{disp}^{(9),ATM}$ is a three-body (triple-dipole, n=9) Axilrod-Teller-Muto (ATM) term. It is modified in the GFN2-xTB Hamiltonian as to depend on the partial atomic charge contrary to the normal D4 model.

### 2.6.4 GFN0-xTB

GFN0-xTB[63] is the latest entry in the SQM family of xTB methods. It does not includes terms higher than the first order in the tight-binding expansion, contrary to GFN2 and GFN1 that include terms up to the third order, making it akin to a DFTB1 method. Consequently, a self-consistent field procedure is not necessary and the method becomes 5 to 20 times faster then GFN2-xTB. Two corrective procedures are included to mitigate this approximation. The first one aims at correcting covalent bonds of heteroatoms using a classical pairwise potential. The second one describes classically the atom-wise isotropic electrostatic energy, presented as $E_{\gamma}$ in GFN1 and GFN2. For non-covalent interactions, the GFN0 method is slightly worse than its predecessor, even if still more reliable than most force-fields methods. By design, GFN0-xTB is more robust than GFN1 and GFN2, as no problems induced by self-consistent charge convergence or other convergence issue happens.

GFN0-xTB is more dependent on empirical parameters than the GFNn-xTB methods presented before, but it still does not contain pairwise parametrization. Its total energy is given by:

$$\begin{aligned} E_{GFN0} &= E_{rep}^{(0)} + E_{disp}^{(0)} + E_{EHT}^{(1)} + \Delta E^{0} \\ &= E_{rep} + E_{disp}^{D4} + E_{EEQ} + E_{EHT} + E_{srb} \end{aligned} \tag{2.97}$$

where $E_{EEQ}$, the electronegativity equalization energy, is the second correction presented before. The EEQ method allows to calculate partial charges between atoms dependent on connectivity and geometry. $E_{srb}$, short-range basis energy, is a short range correction term for covalent bond due to incompleteness of minimal basis sets. Both these terms represent the changes in the zeroth-order reference compared to the previous GFN methods, and they are contained in the $\Delta E^{(0)}$ term of Eq. 2.97. The $E_{disp}^{D4}$ based on the D4 dispersion model[62] does not include the three-body triple-dipole term contrary to GFN2-xTB (see Eq. 2.96).

### 2.6.5 GFN-FF

GFN-FF[64] is the last GFN method published by Grimme's group. It is a force field based method and, accordingly, very different from the previous ones. Its goal is to describe very large biomacromolecular systems (such as proteins), not affordable by the other GFN-xTB methods. Like the other GFN-xTB methods and contrary to regular FF methods, it is parametrized for a large set of elements (up to Z=86). It simplifies the last quantum mechanical terms of GFN0-xTB into classical terms, *i.e.* the zeroth order tight-binding terms are replaced with classical bond, bending angle, and torsion angle potentials. Again, no pairwise parameter is present in the parametrization. With all these simplifications compared to the already very simplified GFN0-xTB, GFN-FF is the most efficient among all the GFNn-xTB methods. In computing time, it reaches quadratic scaling for energy and forces, making it about two orders of magnitude faster than these other methods. The GFN-FF energy can be written as:

$$\begin{aligned} E_{GFN-FF} &= E_{cov} + E_{NCI} \\ &= \left(E_{bond} + E_{bend} + E_{tors} + E_{rep}^{bond} + E_{abc}^{bond}\right) \\ &\quad + \left(E_{IES} + E_{disp} + E_{HB} + E_{XB} + E_{rep}^{NCI}\right) \end{aligned} \tag{2.98}$$

The GFN-FF total energy is separated in two distinctive parts, where $E_{NCI}$ represent the non-covalent interactions described using the EEQ model of GFN0-xTB, and $E_{cov}$ represent the covalent interactions described by asymptomatically correct bonding, bending, and angular terms. A three body ($E_{abc}$) interaction correction is also added. Repulsive interactions are added for covalent and non-covalent interactions. In the non-covalent part of Eq. 2.98, the IES term is described using the EEQ scheme, the repulsive term is a simplified D4 term, hydrogen bonds ($E_{HB}$) and halogen bonds ($E_{XB}$) corrective terms are present due to their challenging description without electronic information.

# References


[1] Jensen Frank, *Introduction to Computational Chemistry, 3rd Edition — Wiley*. Wiley, 2017.

[2] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. Revised*. Dover Publications, 1989.

[3] E. Schrödinger, "Quantisierung als Eigenwertproblem," *Annalen der Physik*, vol. 385, pp. 437–490, jan 1926.

[4] I. N. Levine, *Quantum chemistry*. London, England, UK: Pearson, 1970.

[5] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, pp. 457–484, jan 1927.

[6] B. Sutcliffe, "The Nuclear Motion Problem in Molecular Physics," in *Advances in Quantum Chemistry*, vol. 28, pp. 65–80, Academic Press, jan 1997.

[7] J. C. Slater, "The Theory of Complex Spectra," *Physical Review*, vol. 34, pp. 1293–1322, nov 1929.

[8] 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*, vol. 24, pp. 89–110, jan 1928.

[9] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 111–132, jan 1928.

[10] D. R. Hartree, "The Wave Mechanics of an Atom with a non-Coulomb Central Field. Part III. Term Values and Intensities in Series in Optical Spectra," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 426–437, jul 1928.

[11] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part IV. Further Results relating to Terms of the Optical Spectrum," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 25, pp. 310–314, jul 1929.

[12] V. Fock, "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems," *Zeitschrift für Physik*, vol. 61, pp. 126–148, jan 1930.

[13] C. C. J. Roothaan, "New Developments in Molecular Orbital Theory," *Reviews of Modern Physics*, vol. 23, pp. 69–89, apr 1951.

[14] G. G. Hall, "The molecular orbital theory of chemical valency VIII. A method of calculating ionization potentials," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 205, pp. 541–552, mar 1951.

[15] F. Coester, "Bound states of a many-particle system," *Nuclear Physics*, vol. 7, pp. 421–424, jun 1958.

[16] F. Coester and H. Kümmel, "Short-range correlations in nuclear wave functions," *Nuclear Physics*, vol. 17, pp. 477–485, jun 1960.

[17] J. Čížek, "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods," *The Journal of Chemical Physics*, vol. 45, pp. 4256–4266, dec 1966.

[18] J. Čížek and J. Paldus, "Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methodst," *International Journal of Quantum Chemistry*, vol. 5, pp. 359–379, jul 1971.

[19] I. Y. Zhang and A. Grüneis, "Coupled Cluster Theory in Materials Science," *Frontiers in Materials*, vol. 6, p. 123, jun 2019.

[20] B. Liu and A. D. McLean, "Accurate calculation of the attractive interaction of two ground state helium atoms," *The Journal of Chemical Physics*, vol. 59, pp. 4557–4558, oct 1973.

[21] S. Boys and F. Bernardi, "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors," *Molecular Physics*, vol. 19, pp. 553–566, oct 1970.

[22] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, "State of the Art in Counterpoise Theory," *Chemical Reviews*, vol. 94, pp. 1873–1885, nov 1994.

[23] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review*, vol. 136, pp. B864–B871, nov 1964.

[24] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Review*, vol. 140, pp. A1133–A1138, nov 1965.

[25] J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," *Physical Review B*, vol. 45, pp. 13244–13249, jun 1992.

[26] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, oct 1996.

[27] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, "Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids," *Physical Review Letters*, vol. 91, p. 146401, sep 2003.

[28] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, "Accurate Density Functional with Correct Formal Properties: A Step Beyond the Generalized Gradient Approximation," *Physical Review Letters*, vol. 82, pp. 2544–2547, mar 1999.

[29] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, apr 1993.

[30] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, pp. 785–789, jan 1988.

[31] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Canadian Journal of Physics*, vol. 58, pp. 1200–1211, aug 1980.

[32] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism

Spectra Using Density Functional Force Fields," *The Journal of Physical Chemistry*, vol. 98, pp. 11623–11627, nov 1994.

[33] 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*, vol. 132, p. 154104, apr 2010.

[34] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *Journal of Computational Chemistry*, vol. 32, pp. 1456–1465, may 2011.

[35] A. S. Christensen, T. Kubař, Q. Cui, and M. Elstner, "Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications," *Chemical Reviews*, vol. 116, pp. 5301–5337, may 2016.

[36] R. G. Parr, "A Method for Estimating Electronic Repulsion Integrals Over LCAO MO'S in Complex Unsaturated Molecules," *The Journal of Chemical Physics*, vol. 20, pp. 1499–1499, sep 1952.

[37] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties," *Physical Review B*, vol. 58, pp. 7260–7268, sep 1998.

[38] M. Gaus, Q. Cui, and M. Elstner, "DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)," *Journal of Chemical Theory and Computation*, vol. 7, pp. 931–948, apr 2011.

[39] U. Burkert and N. L. Allinger, "Pitfalls in the use of the torsion angle driving method for the calculation of conformational interconversions," *Journal of Computational Chemistry*, vol. 3, no. 1, pp. 40–46, 1982.

[40] A. K. Rappé and C. J. Casewit, *Molecular Mechanics Across Chemistry*. University Science Books, 1997.

[41] J. W. Ponder and D. A. Case, "Force Fields for Protein Simulations," in *Advances in Protein Chemistry*, vol. 66, pp. 27–85, Academic Press, jan 2003.

[42] A. D. Mackerell, M. Feig, and C. L. Brooks, "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations," *Journal of Computational Chemistry*, vol. 25, pp. 1400–1415, aug 2004.

[43] U. Dinur and A. T. Hagler, "New Approaches to Empirical Force Fields," in *Reviews in Computational Chemistry, Volume 2*, pp. 99–164, John Wiley & Sons, Ltd, jan 2007.

[44] J. E. Jones, "On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 106, pp. 441–462, oct 1924.

[45] T. L. Hill, "Steric Effects. I. Van der Waals Potential Energy Curves," *The Journal of Chemical Physics*, vol. 16, pp. 399–404, apr 1948.

[46] W. L. Jorgensen, "Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water," *Journal of the American Chemical Society*, vol. 103, pp. 335–340, jan 1981.

[47] W. L. Jorgensen, "Revised TIPS for simulations of liquid water and aqueous solutions," *The Journal of Chemical Physics*, vol. 77, pp. 4156–4163, oct 1982.

[48] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, pp. 926–935, jul 1983.

[49] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *The Journal of Chemical Physics*, vol. 112, pp. 8910–8922, may 2000.

[50] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*. West Sussex, England: John Wiley & Sons, Ltd, 2nd ed., 2004.

[51] B. J. Alder and T. E. Wainwright, "Phase Transition for a Hard Sphere System," *The Journal of Chemical Physics*, vol. 27, pp. 1208–1209, nov 1957.

[52] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *The Journal of Chemical Physics*, vol. 31, pp. 459–466, aug 1959.

[53] R. Car and M. Parrinello, "Unified Approach for Molecular Dynamics and Density-Functional Theory," *Physical Review Letters*, vol. 55, pp. 2471–2474, nov 1985.

[54] S. A. Hollingsworth and R. O. Dror, "Molecular Dynamics Simulation for All," *Neuron*, vol. 99, pp. 1129–1143, sep 2018.

[55] L. Verlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," *Physical Review*, vol. 159, pp. 98–103, jul 1967.

[56] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*. Cambridge University Press, sep 2010.

[57] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, pp. 227–249, may 1976.

[58] F. Maseras and K. Morokuma, "IMOMM: A new integratedab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states," *Journal of Computational Chemistry*, vol. 16, pp. 1170–1179, sep 1995.

[59] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, "Extended tight-binding quantum chemistry methods," *WIREs Computational Molecular Science*, vol. 11, p. e1493, mar 2021.

[60] S. Grimme, C. Bannwarth, and P. Shushkov, "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems

Parametrized for All spd-Block Elements (Z = 1-86)," *Journal of Chemical Theory and Computation*, vol. 13, no. 5, pp. 1989–2009, 2017.

[61] C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions," *Journal of Chemical Theory and Computation*, vol. 15, no. 3, pp. 1652–1671, 2019.

[62] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme, "A generally applicable atomic-charge dependent London dispersion correction," *The Journal of Chemical Physics*, vol. 150, p. 154122, apr 2019.

[63] P. Pracht, E. Caldeweyher, S. Ehlert, and S. Grimme, "A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules," *ChemRxiv*, no. preprint, pp. 1–19, 2019.

[64] S. Spicher and S. Grimme, "Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems," *Angewandte Chemie - International Edition*, vol. 59, no. 36, pp. 15665–15673, 2020.

# Chapter 3

# History of the icy grain modeling

# Abstract

In the chapter we will present a brief history of the modeling of interstellar icy grains both experimental and theoretical. We will pay a close attention to computational chemistry modeling of icy grains, with presentation of both recent periodic and cluster models. A small part we will dedicated to bare dust grains. As a second part of this chapter we will present a benchmark of the GFNn-xTB methods, the method of choice used in the remaining of this thesis, by (i) comparing the results obtained on small water clusters, of 2 to 10 molecules, by xtb methods and more accurate methods (MP2 and coupled cluster), and (ii) comparing the adsorption geometries of several interstellar relevant molecules obtained by GFN2 and DFT.


## 3.1 Experimental models

Experimental studies of interstellar grain analogues are paramount to understand the chemical composition of the ISM. Laboratory examinations of ices and dust grains analogues under interstellar physical chemical conditions have allowed researchers to identify multitudes of chemical species via comparison with observed IR adsorption spectra. Nevertheless, like for computational studies, experimental studies suffer from limitations. Interstellar conditions such as very low pressures of $10^{-5}$ bar coupled with temperatures of 10 K in dense interstellar clouds indicate time scales impossible to study on a human life-span for some interstellar processes. In this section we will present a short summary of past and present experimental studies of chemical processes happening on dust and ice covered grains.

### 3.1.1 Short history

At the start of the previous century interstellar grains were not yet discovered. Holes in the ISM observation were thought to be mainly the result of space void and not extinction from particles contained in interstellar gas clouds. After the proof of their existence in 1930 by R. J. Trumpler[1], their chemical composition was mostly a mystery and firsts laboratory experiments were done to compare infra-red spectra of molecules in interstellar conditions to IR spectra obtained via observation of the ISM. To do that they would either produce the IR spectra of one molecule, or produce an ice grain or dust

grain analog (or both) based on theoretical proposed models and compute the IR spectra. Comparing single molecules IR spectra to observed interstellar grain spectra is quite tricky, since molecules on interstellar grains are mostly in interactions between each others. Therefore, comparing them independently implicates either a limited interaction between them, or a need for correction in the IR spectra, as position and intensity would be different between observed and laboratory experiments. Firsts models of interstellar icy grains (see Van de Hulst 1949[2] and Whipple 1950[3]) were made analogous of the believed composition of comets at the time, mostly "icy-conglomerate" composed of water, ammonia, molecular hydrogen, and methane, but these aggregates were found to be unstable in interstellar conditions, due to cosmic ray impacts common in the ISM.

Today, experimental studies of molecular adsorption, and desorption, on interstellar grains analogs are mainly done using a a ultra-high vacuum (UHV) chamber capable of obtaining pressures close to those present in the interstellar medium and dense clouds, i.e. up to $10^{-11}$ mbar or $10^6$ particles per $cm^3$. Molecules, mainly water if an interstellar icy grain surface is modeled, are deposited onto a substrate, or spectrometer window, at very low temperatures (a few Kelvin). Crystalline ices and amorphous ices can be obtained depending on the deposition temperature; water will be deposited in a glass structure (amorphous) at low temperature, and in a crystalline structure at temperatures around 80 K. To have more control on the crystalline or amorphous structure of the icy grain analog, annealing of the ice to crystallization is performed, *i.e.* heating the ice at temperatures close to 300 K to make its structure amorphous, then cool it down to low temperatures slowly (cooling the ice quickly gives the amorphous structure). The deposited ice can be be either: (i) submonolayered (the adsorbates do not cover the entire substrate); (ii) monolayered (the absorbates cover the entire substrate with one layer of molecules); (iii) multilayered (the adsorbates cover the entire substrate with more than one layer). In the multilayered conditions, if the molecules deposited are identical and not a mixture, the first layer of molecule is bonded to the substrate while the following layers are bonded to other identical molecules. In that case, the desorption of molecules from identical molecules is called a "zero-order" desorption, while the desorption of molecules adsorbed at the substrate is called a "first-order" desorption. The widespread technique used to study molecular desorption is the Temperature Programmed Desorption (TPD), it involves depositing a molecule of interest onto a substrate and study its desorption by heating the

substrate at a rate of, typically, a few Kelvin per minute. The mass of the desorbed molecules is then measured using, generally, a Quadrupole Mass Spectrometer (QMS). From the measured desorption rates, one can obtain a thermal desorption curve from which we can derive a desorption energy value using the Arrhenius, or Polanyi-Wigner, equation:

$$r(\theta) = \frac{d\theta}{dt} = \nu(\theta)\,\theta^n e^{\frac{-E(\theta)}{RT}} \tag{3.1}$$

where $r$ is the rate of desorption, $\theta$ is the surface coverage, $\nu$ is the pre-exponential factor, $n$ is the order of desorption, $E$ is the desorption energy, also called binding energy (BE) or activation energy (the energy needed to separate the molecule from the substrate), $R$ is the gas constant, and $T$ is the absolute temperature.

## 3.2 Computational models

Interstellar dust grains are submicrometric particles composed of a silicate/-carbonaceous materials (the *core*). In the cold parts of the interstellar medium ($\leq 50$ K) the core is covered by an amorphous *mantle* phase, mostly made of water with a minor fraction of other molecules present in the gas phase surrounding their environment (CO, $H_2$, $NH_3$, etc. for dense molecular clouds). Water molecules belonging to the mantle are formed *in situ*, through surface recombination of pre-adsorbed oxygen and hydrogen atoms. When computational chemistry is adopted to elucidate some of the grain details, one is facing limiting factors like: (i) computing resources; (ii) accuracy of the available computational methods; (iii) weakness in the grain structural models representing the real system. In the specific case of the interstellar grains, even an accurate experimental understanding of the fine structural details is missing due to the difficulties of investigating *in situ* (*i.e.* in the ISM) their features. Therefore, results from computer modeling are usually contrasted with experiments carried out in the terrestrial laboratory, in which many operative assumptions are forcibly applied to allow results on a reasonable time scale. For instance, the adsorption and reactions at the real interstellar grains occur on a very long time scale and in a highly diluted situation, at variance with the laboratory experiments whose outcome is expected in minutes or hours. Due to these limitations, each computer representation of the interstellar grains has, *a priori*, a degree of arbitrari-

ness. These approximations result in models exhibiting different properties such as level of crystallinity, size, or composition.

Interstellar grain models are numerous in the literature and the comings sections summarize the literature of grain models used starting from the inference of interstellar dust and the infancy of computational chemistry, up to the present days.

## 3.3 A historic perspective of grain models

One of the earliest theoretical model of interstellar icy grain was proposed in 1963 by Gould and Salpeter[4]. At the time, icy grains entirely made of hydrogen bounded molecules were still a widely accepted concept, and their most widespread composition was the one proposed by Van de Hulst[2] made roughly of 100 molecules of $H_2O$, 30 molecules of $H_2$, 20 molecules of $CH_4$, 10 molecules of $NH_3$, and 5 molecules of MgH. The icy grain surface was modeled around a cubic periodic lattice with a molecular mechanics level of theory via a Lennard-Jones potential. Gould and Salpeter[4] were able to compute the adsorption energies of molecular hydrogen and the hydrogen atom on the surface of an idealized icy grain, up to the $H_2$ formation. While this approach had several drawbacks, like the use of molecular mechanics for an intrinsic quantum mechanical process, or the highly idealized grain surface, it was quite impressive for the time and opened up the route for more sophisticated approaches. Their research was continued and improved by the work of Knaap et al. 1966[5] using the same grain composition and periodic model. Parallel to this work, Stecher et al.[6] also studied molecular formation on the surface of interstellar grains using van der Waals potentials, while adopting a graphite surface instead of dirty ice, simulating processes at the core's grain.

In 1970, a pioneering study was made by Hollenbach and Salpeter[7]. Though the surface of the grain was still modeled via a periodic cubic crystal system, a quantum mechanical approach was adopted to find the adsorption energy of different molecules. To deal with the computational limitations of the times, the tight binding approximation was used in its "extreme form", where the adsorption potential is replaced by an harmonic oscillator potential, approximating the periodic lattice as a manifold of harmonic oscillators. Possible adsorption sites were also crudely defined, as only three different types were characterized: one between any two molecules next to each other,

one between any four molecules forming a square next to each other, and one on top of any molecule. The atoms adsorbed were observed to hop between these surface points until they either got stuck or escaped. Even though severe approximations were made, their results proved to be highly accurate when compared to the experimental data of the time.

By going back to a completely classical mechanical level of theory, Burke and Hollenbach[8] modeled, in 1983, the interstellar icy grain surface as a one-dimensional "soft-cube" model[9,10] to study the interaction between gas and surface atoms. The surface was represented as a line of successive cubical atoms, each attached to one spring bound to an infinite mass. Gas particles were launched toward the surface in a linear trajectory. When close to the surface, the gas particles' trajectory were changed due to an attractive potential linked to the surface. After colliding with the surface atoms, a repulsive potential was encountered, making the particles either rebound or hop until stuck on the surface or ejected from it. From this study, the authors derived the thermal energy accommodation coefficient and the trapping fraction (important parameters to study the chemistry of interstellar gas) for water ice, silicates, and graphite. While far from being an accurate representation of the process emulated, this dynamical method is extremely cheap in computational power, and Burke and Hollenback[8] found results in moderate agreement with the experiments.

In 1985, Leitch-Devlin and Williams[11] proposed an entirely quantum mechanical model for studying the sticking of gas particles onto a grain surface. Four different lattice were used, classed by softness depending on their bond strength: (i) a metal oxide; (ii) graphite; (iii) silicate; (iv) graphite with a physisorbed water monolayer. All models were representing interstellar grain material,s as they were proposed at the time of the study. The sticking coefficients, S, were computed at different temperatures for the gas and the grain. For almost all grain types and temperatures, S was found to follow approximately a bell shape curve for all the lattices, rising from 0 at a gas temperature of 0 K, to a maximum value when increasing the temperature (generally between $10 - 100$ K), to a lower value again, when increasing T further.

Following Hollenbach and Salpeter[7], Burke and Hollenbach[8], and Leitch-Devlin and Williams[11], Buch and Zhang[12] continued in 1991 the theoretical study of sticking coefficient on top of an icy grain surface. Previous studies were based on somehow crude assumptions, the most important ones being the approximation of surface molecules as spherical, and the modeling of

the grain surface as mostly flat and crystalline. Buch and Zhang[12], instead, proposed an amorphous surface, composed of 115 water molecules gathered in the form of a cluster built by slow condensation of gaseous $H_2O$. Indeed, the amorphous phase of the ice is very important for the prediction of S, as it implies a larger number of binding sites of different kind compared to crystalline surfaces, resulting in a sensible distribution of binding energies. Furthermore, the complexity of an atom trajectory on an amorphous surface cannot be represented by the simple trajectory occurring on a flat crystalline surface, where atoms can only either become adsorbed or ejected back to the gas phase. Contrary to Leitch-Devlin and Williams[11], this study was done entirely adopting classical mechanics, via the TIPS2 potential[13], explicitly developed for water/water interactions, while adopting pairwise-additive potentials for interactions between water and gas particles. Furthermore, the water molecules were considered structurally rigid.

Continuing with the study of the sticking coefficient, Masuda et al.[14] also used a classical molecular method in the form of molecular dynamics. The modeled icy grain surfaces were made out of amorphous periodic slabs generated by randomly placing 1000 water molecules in a cubic $40 \times 40 \times 40$ Å unit cell. Molecular dynamic lasting 10 ps were run, until water ice slabs of 20 Å in thickness were generated and considered at equilibrium. Two temperatures, 10 K and 70 K, were used to generate the slabs to simulate different possible densities of ice water. Alike Buch and Zhang[12], Masuda et al.[14] used rigid water molecules and a TIPS2 potential[13] for water/water interactions. For the interaction between H gas atoms and water, a Zhang-Sabelli-Buch potential[15], based partly on Lennard-Jones potential, was used. One main difference from the Buch and Zhang[12] study, other than the use of periodic boundary conditions instead of a cluster, was the size of the surface model. While Buch and Zhang[12] used a classical potential to represent their surface, they still decided to adopt a somewhat small grain to represent the surface, maybe due to the computational limitations at that time, while Masuda et al.[14] were able to treat a slab of 1000 water molecules, allowing more interesting trajectories to model the atoms' sticking.

In 1999, Woon[16] studied chemical reactions on astrophysical ices using an *ab initio* quantum mechanical method, namely MP2 with a 6-31+G** basis set and including the ZPE correction. As this level of theory is highly accurate but rather costly, the water ice grain surface was modeled as a passive influence through a continuous solvation model. Active influence of water molecules were still studied during dimerization and trimerization of

$H_2CO$, by including them explicitly to study their catalytic nature; the same approach was followed with $NH_3$.

Finally, Bergeron et al.[17] studied in 2008 the formation of OH on a dust grain modeled by a single graphite layer, i.e. a graphene sheet. The graphene itself was simulated as a cluster model of increasing size, called polycyclic aromatic hydrocarbon (PAH); the sequence of clusters envisaged benzene ($C_6H_6$), pyrene ($C_{16}H_{10}$), coronene ($C_{24}H_{12}$), circumpyrene ($C_{42}H_{16}$), and circumcoronene ($C_{54}H_{18}$). Unlike Woon[16], the level of theory was similar to the previously quoted methods (*i.e.* molecular mechanics), adopting quasi classical trajectory (QCT) calculation, as the authors indicated it as a suitable approximation for the quantum Langmuir-Hinshelwood dynamics.

## 3.4 Recent works

More recently, numerous approaches to model interstellar grains have been proposed, for both the dust core and the icy mantle. With modern computational schemes and technology available, theoretical methods of high quality can be easily applied. Still, very few of these models were built following the processes supposed to be operative in the ISM (*vide supra*). In this section, we briefly review the most recent grain modeling work. For a better legibility, we have separated the cases in which the grain mantle is modeled by a periodic approach from those in which a cluster approach was adopted. As for the bare dust grains, we report recent advance in a separate section.

### 3.4.1 Bare dust grain models

Like ice mantles, bare dust grains are extensively investigated in the astrochemical community. In this small subsection we will present some cosmic dust models developed to study the formation of dust grains, as well as different processes happening at their surfaces.

Goumans & Bromley[18] studied the structure stability of the first eight members of the $(MgO)_n(SiO_2)_n$ (enstatite) and $(MgO)_{2n}(SiO_2)_n$ (fosterite) series in order to investigate the adsorption of O and H atoms at the surface of these nanosilicates. They generated hundreds of candidate structures through a force field method to establish the ground-state geometry of the nanoclusters and selected $20-40$ of the most promising ones to get optimized using the B3LYP functional[19–22] and 6-31G* basis set. The most stable struc-

ture was finally found to be Nanopyroxene ($Mg_4Si_4O_{12}$). They found that the nanostructure chemisorbed H atoms without barriers and acted as a catalyst for the $H_2$ formation. Oxygen atom was also found to be chemisorbed and subsequent addition of H atoms was incorporated into the structure instead of reacting to form $H_2$. Water molecules were only formed after that at least four oxygen atoms were incorporated into the nanocluster. These results could partially explain the large oxygen depletion observed in diffuse cloud.

Adopting the same approach, the same authors[23] studied the nucleation of silicates via heteromolecular homogenous condensation. Clusters were formed via aggregation of SiO, Mg, and $H_2O$ to give magnesium silicates. From these clusters, 25 promising structures were selected and optimized using the B3LYP functional[19–22] and 6-31G* basis set. They found that condensation of pure SiO at high temperature is unfeasible, contrary to the oxidation of SiO dimer and the addition of magnesium and further oxygen atoms. This study provides a possible pathway towards the formation of magnesium silicates in stellar outflows conditions.

Following the previous paper in which the nucleation of SiO was predicted to be unfavorable at high temperatures, Bromley et al.[24] investigated in which conditions their nucleation was feasible. Via an in-house computational method, they produced several $(SiO)_n$ clusters (optimized at B3LYP functional[19–22] and 6-311+G(2d,p) basis set), exhibiting segregation between oxygen-rich and silica parts for $n \leq 7$. Kinetic Nucleation Theory (KNT), Classical Nucleation Theory (CNT), and a full time-resolved kinetic model were used to predict the nucleation rate. A range of pressure and temperature have been predicted for which nucleation occurs and found that the full kinetic model is in agreement with experimental data on the critical pressure to achieve SiO nucleation.

Zamirri et al.[25] modeled pure forsterite ($Mg_2SiO_4$) nanoparticles of varying sizes between $\sim 1.2$ and $\sim 4.6$ nm to study the adsorption of CO. They followed a top-down approach where a crystalline forsterite optimized at PBE/OIBS level of theory[26] was cut along the crystallographic directions representing the most energetically stable surfaces. The resulting nanoparticles where then relaxed at PBE/OIBS for the smallest ones (hundreds of atoms), while adopting a force field for the largest ones (thousands of atoms). Amorphization via thermal annealing by MD run at 1800 K with 10 ps of equilibration and 20 ps of production time (200 ps for the largest nanoparticles). A final geometry optimization was performed after the MD run. A bottom-up approach was also used to generate cluster from coagu-

lation of pyroxene ($MgSiO_3$) and MgO nanoclusters. The stoichiometry of the nanoparticles was kept equal to that of fosterite. After coagulation, a MD computation at 1800 K was performed and the clusters were optimized at the PBE/OIBS level of theory. IR spectra were computed for all these nanoparticles (bulk cut before amorphization, bulk cut after amorphization and coagulated clusters), to check whether spectral differences between objects coming from crystalline forsterite were indeed different from that of the amorphous ones. The surprising result was that it was difficult to distinguish between the studied cases on the sole base of their IR features. This implies that estimating the degree of crystallinity of the silicate grains in the ISM is difficult by the analysis of the IR spectral features. Features of coagulated grains were somehow different when compared to the clusters made with the top-down approach.

Rimola & Bromley[27] study the aggregation of small silicate nanoclusters by adopting a periodic approach. The aggregation was simulated by introducing into a cubic unit cell the silicate ($Mg_6Si_3O_{12}$) cluster to be used as a seed for further aggregation. Geometry optimization of the internal atomic coordinates at B3LYP level was performed, while keeping the unit cell parameters fixed. After the first optimization, the unit cell size was progressively reduced and relaxed until the interactions between adjacent clusters were deemed sufficient to unfix the lattice cell parameters during optimization. After this process, several amorphous silicate were made, with different structures and densities (*i.e.* porous and non porous). After analysis, the aggregation process was found to be energetically favorable. IR spectra were computed and found to match well with the observed astronomical IR spectra, and in addition, for the porous system, a specific vibrational feature appeared in the spectrum.

### 3.4.2 Periodic icy mantle grain models

Mathis et al.[28], (usually referred as MRN) showed that a mixture of silicate and graphite dust, with a grain size distribution (number of grains per grain radius) proportional to $a^{-3.5}$, where $a$ is the grain radius ($a \approx 0.001 - 0.25\mu$m or $1 - 100$ nm), reproduces the Milky Way extinction curve. The distribution curve reveals a prominence of smaller grain sizes, which is fortunate for computer simulations. From an atomistic viewpoint, representing their surface through a periodic model can be physically sound. Contrary to the size limitation of the clusters, periodic boundary conditions allow the imi-

tation of long range effects of the grain (particularly for the larger sizes), affecting the diffusion and the adsorption features of the adsorbed molecules. As a downside, the richness of surface different sites is, in general, inferior to that of a medium sized cluster, as only a small fraction of the unit cell atoms is at the surface. Hence, the gain of long range effects is balanced by the loss of surface variety, impacting the number of binding sites available for surface processes. The variety of surface sites can be enhanced by using periodic amorphous models instead of a crystalline ones, but the building of an amorphous slab is somehow a subjective and non-trivial operation. It is easy to prove that a square unit cell of length $b$ has the same surface area of a corresponding spherical cluster with radius of $\approx b/4$. For the unit cell slab to have the same volume of the cluster, its thickness should be $h \approx b/16$. For instance, a reasonable value of $b = 20\,\text{Å}$, would imply a very small $h = 1.25\,\text{Å}$ as a thickness. Clearly, in order for $h$ to define a reasonable slab thickness (involving at least 4 $H_2O$ molecules contributing to the thickness), it should be much larger than the theoretical limit giving the same spherical cluster volume. This implies that the number of atoms needed to simulate an icy surface via periodic boundary conditions is definitely larger than a cluster with the same surface area. Another advantage of using a cluster approach for simulating ISM grains is in a more general applicability of a much larger variety of quantum mechanical methods than for systems under PBC. In the latter, basically only DFT can be routinely used, while for clusters all the accurate wavefunction based methods can be used, provided enough computer resources are available.

In Zamirri et al. 2017[29] the interstellar core dust grain was modeled with fosterite ($Mg_2SiO_4$) periodic crystalline surfaces for the study of carbon monoxide adsorption, a challenging endeavor, as CO is notoriously difficult to study even with accurate DFT functionals. The surfaces were generated from bulk and surface models optimized at the B3LYP level of theory. Seven slab were generated from the most natural occurring fosterite surfaces, and re-optimized at the B3LYP-d* level. The considered surfaces were designed to avoid a dipole moment across the slab, or breaking the stoichiometric ratio of $SiO_4$. For all the seven considered surfaces, a single CO was first adsorbed on each available Mg ion. Different surface coverage were also considered up to cover all available Mg ions. For a selected number of surfaces, an intermediate coverage was tested. For all three types of coverages CO resulted adsorbed through weak interactions. Two CO molecules per Mg ion were also considered, revealing inter-CO repulsion. Only is some cases, the attraction

with the Mg ions overcomes the CO repulsion. The cristallinity of the surfaces is an important limitation of this study, as silicate in interstellar clouds are believed to be of amorphous nature. Nevertheless, the reported results may be useful to interpret data in other regions of the ISM where crystallinity is expected to be higher due to a different thermal evolution of the grains.

In a following paper from Zamirri et al. 2018[30], two different periodic crystalline models, P-ice and ice XI, were used to study the interaction of water ice with CO and its resulting IR spectral fingerprint. P-ice is selected for its lack of dipole component perpendicular and parallel to its (001) and (010) surfaces and, like ice XI, possesses 12 molecules per unit cell. As a complement, and to better understand the resulting IR spectra of CO on water adsorption, pure crystalline CO ice, as well as CO trapped in ice water cage to simulate filled pores in icy grains were studied. These pure crystalline CO ices can be representative of CO islands growing on the dust grain mantle in interstellar conditions. As a single CO molecule adsorbed on a periodic crystalline slab is not representative of the true ratio between CO and $H_2O$ in interstellar ices, Zamirri et al.[30] studied models with higher CO loading. The system was then amorphized via an *ab initio* molecular dynamics simulation at 300 K for 15 ps, and "snapshots" (structures at different moments of the AIMD simulation) were subsequently optimized at the DFT PBE-D2 level. While the periodic crystalline model is not representative of interstellar ices in dense molecular clouds, this study showed interesting results involving CO interaction with water in small ice pores, and its tendency to clusterize with increasing concentration.

Ferrero et al. 2020[31] examined the adsorption of 21 astrochemically relevant molecules using both crystalline and amorphous water surfaces. The crystalline water surface was identical to the P-ice one used by Zamirri et al. 2018[30], and structures were optimized at the DFT/A-VTZ* level (more information regarding this study are provided in section 3.5.2). The amorphous slab unit cell is composed of 60 water molecules, 48 more than for the crystalline one, constructed by aggregating three amorphous clusters of 20 water molecules each. As the amorphous slab envisages a unit-cell five times more water rich than the crystalline one, the same level of theory was not manageable, and the fast and less accurate HF-3c method was used instead. Despite the amorphous nature of the ice (and therefore closer to the ice in ISM), the definition of the periodic amorphous slab was entirely arbitrary. Furthermore, to speed up the search for possible adsorption site and reduce the computational burden, the considered adsorbates were manually

positioned at the ice sites by maximizing the hydrogen bond strength. This clearly bias the computed binding energies towards higher values. Additionally, the amorphous slab structure was rather fragile, as the authors observed important restructuring of the ice during the geometry optimization by the adsorption process.

### 3.4.3 Cluster icy mantle grain models

Cluster models are defined as an aggregate of particles maintained together by their respective attraction. Contrary to periodic models, cluster models do not emulate long range but take them into consideration by enlarging the cluster size. As will be showed in this section, the cluster models used in astrochemistry studies are mostly of very small sizes.

A highly quoted paper by Wakelam et al.[32] proposed in 2017, adopted a single water molecule to represent an icy grain. Dozens of interstellar relevant molecules, from a simple hydrogen atom to more complex molecule like $C_2H_5CHO$ have been put in interaction with the single water molecule. Due to the very small model size, the level of calculus was either M06-2X/cc-pVTZ level without ZPE and BSSE corrections or MP2/aug-cc-pVTZ. To account for the deficiencies of a single water molecule to represent the ice mantle of a micrometric icy grain, and the missing hydrogen bond cooperativity exhibited in large water clusters, a correction factor for the computed BEs was derived by fitting a selected values of the BEs against the best experimental values. The scaling factor is then used for all the other cases, assuming to be valid for a large variety of cases. This clever approach has the advance of being extremely computationally cheap at the price of a high empiricism. Furthermore, no physical insights and fine details for understanding the resulting BEs can be extracted by this approach.

As an extension of the Wakelam et al.[32] work, Das et al.[33] computed the BEs of a set of molecules taken from the previous work, as a function of water clusters size. Five water clusters made of $1-6$ water molecules were used, and adsorptions studied at MP2/aug-cc-pVDZ level. By comparing their results and those of Wakelam et al.[32], using both DFT and MP2 structures for the latter, with experimental results from many different literature sources, they found that, on average, the percentage of deviation between the theoretical BEs and the experimental ones were decreasing with the increasing of the cluster sizes: from a bit more than 40% deviation for a single water molecule, to only 20% for the pentamer and hexamer clusters. This showed

the importance of increasing the interstellar grain size to better match the experimental trend.

Sameera et al.[34] adopted two kinds of crystalline ice surface clusters and an amorphous one for studying the $CH_3O$ adsorption. The crystalline clusters were made from a $I_h$ ice structural model, while the amorphous clusters were made from a previous work by Anderson et al.[35], who used an amorphous periodic icy slab obtained via molecular dynamics evolution of a crystalline one. The three different types of cluster were optimized through the ONIOM method, with the low-level computations at the molecular mechanics level using the AMBER force field and the AMOEB09 polarizable force field, and the high-level computations at the wB97X-D/def2-TZVP level. To maintain the structural memory of the crystalline cluster, the structure treated with the low-level method was frozen during optimizations. The two kind of crystalline clusters were made of 162/156 water molecules, 48/44 in the QM region and the remaining 114/112 to the MM region, while the amorphous cluster included 162 water molecules, 49 for the QM part and 113 for the MM part. In total, $CH_3O$ was adsorbed on 16 different binding sites for the two kind of crystalline slabs, and on 10 sites for the amorphous one.

For studying the adsorption energy of several single atom species, Shimonishi et al.[36] produced 9 different amorphous water clusters made of 20 water molecules each, using classical molecular dynamics calculation at 300 K followed by a run at 10 K. The 9 structures were then optimized at the $\omega$-B97XD/6-311+G level. The adsorbates were randomly placed on the clusters the resulting largest BEs were averaged.

To study chemical reactions, on top of grain surfaces, creating interstellar complex organic molecules from simple radical species, Rimola et al.[37], Duvernay et al.[38], and Enrique-Romero et al.[39] used small water clusters ranging from 18 to 32 water molecules. Various DFT methods were used, namely BHLYP/6-311++G(d,p) for Rimola et al.[37], B3LYP/6-31+G(d,p) for Duvernay et al.[38] and M06-2X-D3/6-311++G(d,p) for Enrique-Romero et al.[39]. In the case of chemical reactions on top of a grain surface, the real computational burden is no longer in the calculation of the BE for various sites but in the reconstruction of the reactive potential energy surface.

To study the dynamics of nitrogen atoms on an amorphous water surface Molpeceres et al.[40] used an hemisphere of 499 water molecules as an interstellar icy grain model (also used in previous works by Song and Kästner[41], Lamberts and Kästner[42], and Molpeceres and Kästner[43]). The hemisphere was cut out from a unit cell including 18937 water molecules optimized with

the TIP3P method, equilibrated at 300 K for 100 ps, and suddenly quenched at 10 K for 50 ps using the NAMD code in the NVT ensemble and with the Langevin thermostat. Afterwards, the ice was cut in an hemispherical way and, during the subsequent molecular dynamic simulations, 204 external water molecules were kept fixed, to provide a boundary to the outer layer of the hemispherical cluster, while the remaining 295 water molecules were free to move and interact with the studied atoms. While this method is an interesting way of producing amorphous clusters for mimicking interstellar icy grain surfaces, the high temperature adopted in the amorphization procedure does not account of the physical chemistry conditions of the ISM.

In summary, recent proposal for modeling the interstellar grains adopted either too small clusters, unable to accurately account for the hydrogen bond cooperativity and binding sites variety, or built following a procedure that implies physical chemistry conditions not found in the dense molecular clouds of the ISM. Indeed, the icy mantle is made by subsequent oxygen and hydrogen reactions to give water on the dust grain or on an already formed ice mantle.

In the next chapter we propose a new approach to build up amorphous icy grain models made by several hundreds of water molecules. Our method lies on the following essential features: (i) the grain should be larger than $20-30$ water molecules, as usually found in the literature, as that size cannot cope with the H-bonding cooperativity of a real icy mantle; (ii) the hydrogen bond features within the icy grain should be accurately represented by adopting a proper quantum mechanical method; (iii) the ice should be amorphous by construction, avoiding the usual approach of heating at high T and sudden cooling of a crystalline ice model, which has no counterpart in the grain evolution in the cold (10 K) condition of the ISM; (iv) to be generally available to the astrochemical community to extend the adsorption of important complex organic molecules but also reactive process at its surface.

To arrive to the definition of the "realistic icy grain" the starting point was to choose the quantum mechanical method, which should be highly accurate for dealing with H-bonds, while being cheap enough to handle large water cluster, as described in the next chapter.

## 3.5 GFN-xTB as a method of choice for icy grain simulation

The GFN-xTB family of methods[44–46], presented extensively in chapter 3 of this thesis, has been developed primarily to treat non covalent interactions including hydrogen bonds, the most important interaction in water clusters. Furthermore, at variance with many literature force fields, these methods have been extensively parametrized for almost all elements of the periodic table (up to Z=86), rendering the GFN-xTB family "neutral" with respect the different molecules to be considered as adsorbates, ensuring a good accuracy in the computing BEs for the large variety of complex organic molecules of astrochemical interest.

To verify which of the three methods (GFN0, GFN1 and GFN2) available in the GFN-xTB suite was the best, we performed two kind of benchmark: one for water ice structure and energetic, contrasting the GFN-xTB results with that by Temelso et al.[47], and a second one in which we compared the GFN-xTB BEs with the ones by Ferrero et al.[31] on periodic icy models for 21 molecular species.

### 3.5.1 Water cluster study

To benchmark water ice structures and energetic, we used the set of water clusters by Temelso et al.[47], containing 38 clusters made from 2 up to 10 water molecules. The variety of the clusters was due to the different spatial configurations that can be afforded by the different H-bonding organization between the water molecules. By counting the number of cluster $n$ of a given water nuclearity $m$ using the $(n/m)$ notation, Temelso et al considered (1/2), (2/3), (3/4), (7/5), (8/6), (11/7), (2/8), (2/9) and (2/10) clusters. As said, group of clusters containing the same number of water molecules exhibit different H-bond configurations between the same number of water molecules. The geometries of the clusters were optimized by Temelso and al.[47] at RI-MP2/aVDZ and then used to refine the total energy by a single-point calculation at CCSD(T)/CBS/CBSnocp (Coupled Cluster Single Double Triple within the complete basis set approximation).

Here, the "structural energy" refers to the change in energy resulting from the relaxation of the water cluster structure compared to the isolated water molecules. It is computed on a per water molecules and is defined as the

difference between the energy of an isolated water molecule $E_{H2O}$ multiplied by the number $N_{H2O}$ of water molecules and the energy of the interacting cluster $E_{WC}$:

$$E_N^{struct} = \frac{E^{struct}}{N_{H2O}} = \frac{E_{H2O} \times N_{H2O} - E_{WC}}{N_{H2O}} \tag{3.2}$$

where $E^{struct}$ is the non-normalized structure energy.

To benchmark GFN-xTB we reoptimized the optimized structures of Temelso et al.[47] using GFN0, GFN1, and GFN2, and to compare both the structure energy of GFNn-xTB with the CCSD(T) ones as well as the resulting geometries.

#### 3.5.1.1 Water cluster structure energy prediction

The results for the structure energies have been plotted in Fig. 3.1 for the three semi-empirical GFN methods. The color code of the dots indicates the number of water molecules in the clusters, the black colored line is the best fit line for all the points, and the gray colored line is the "ideal" line representing perfect agreement between GFN-xTB and CCSD(T) results.

In term of structure energy, GFN0, the cheapest semi-empirical method, is the least accurate, while GFN1 and GFN2 show a best fit very close to the CCSD(T) results, with the latter showing an overall better agreement. These results are in agreement with previous benchmarks by Grimme's group[44–46] and confirmed the impressive performance for structure energy of GFN2, considering the orders of magnitude speedup of GFN2 compared to the CCSD(T)//RI-MP2 one.

In Fig. 3.2 we show the absolute percentage of difference ($APD$) between the structure energy computed at CCSD(T) level and those computed at the GFNn-xTB level, defined as:

$$APD(n) = \left| \frac{E_{CCSD(T)}^{struct} - E_{GFNn-xTB}^{struct}}{E_{CCSD(T)}^{struct}} \right| \times 100 \tag{3.3}$$

where $E_{CCSD(T)}^{struct}$ is a structure energy computed via coupled cluster theory, and $E_{GFNn-xTB}^{struct}$ is a structure energy computed via GFNn-xTB.

As for data showed in Fig. 3.1, the GFN2 results showed errors almost half of those of GFN1 and almost a third of GFN0. At variance with both GFN0 and GFN1 the decrease of the $APD(2)$ is remarkable for clusters with

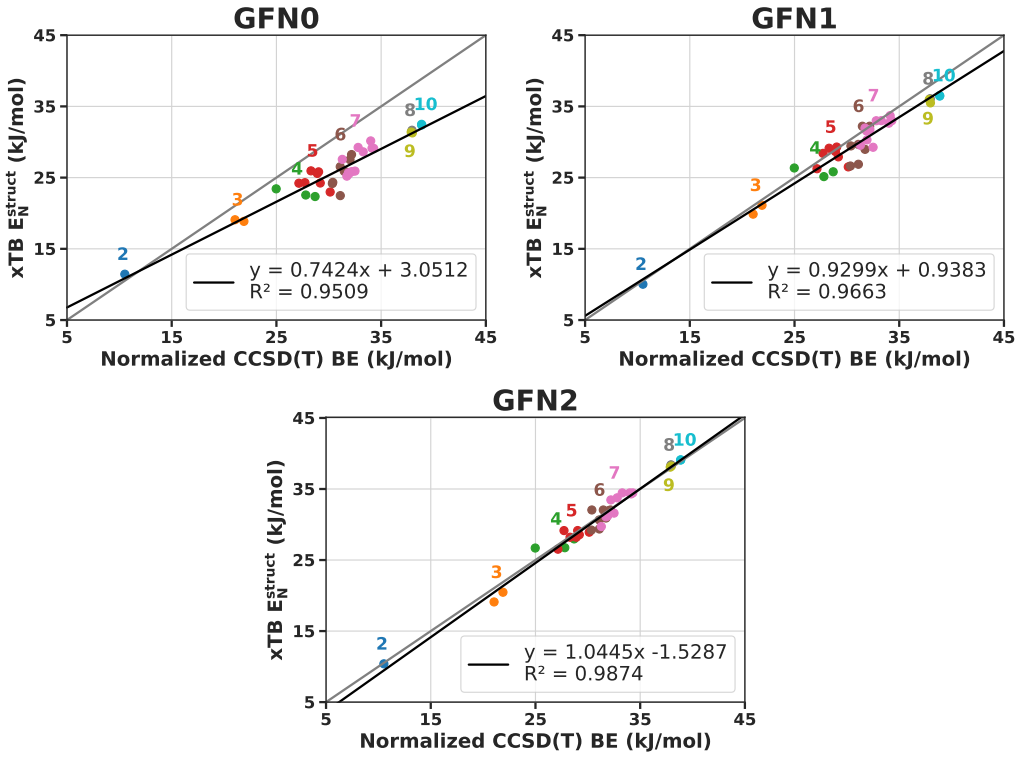



**Figure 3.1:** Comparison between normalized $E_N^{struct}$ values of the 38 water clusters computed by GFN-xTB and CCSD(T). GFN0, GFN1, GFN2, refer to the level of GFN-xTB parametrization (see Chapter 2 for details). Labels indicate the nuclearity $m$ of each water cluster.

more than 7 water molecules. This result is extremely encouraging for the simulation of large interstellar amorphous water ice clusters. However, as the sample of cluster with more than 7 water molecules is so few, we concede that a definitive conclusion about this matter is not possible.

If we compare the min/max/average APD values for the GFN methods, we obtain 0.2/9.2/2.6 % for GFN2, 0.1/13.6/4.7 % for GFN1, and 6.2/27.7/15.8 % for GFN0, showing the GFN2 method the method of choice.

#### 3.5.1.2 Water cluster geometry prediction

We compared the RI-MP2/aVDZ optimized geometries with the one at GFN-xTB levels. The set of Temelso et al.[47] contains 38 structures, a number too

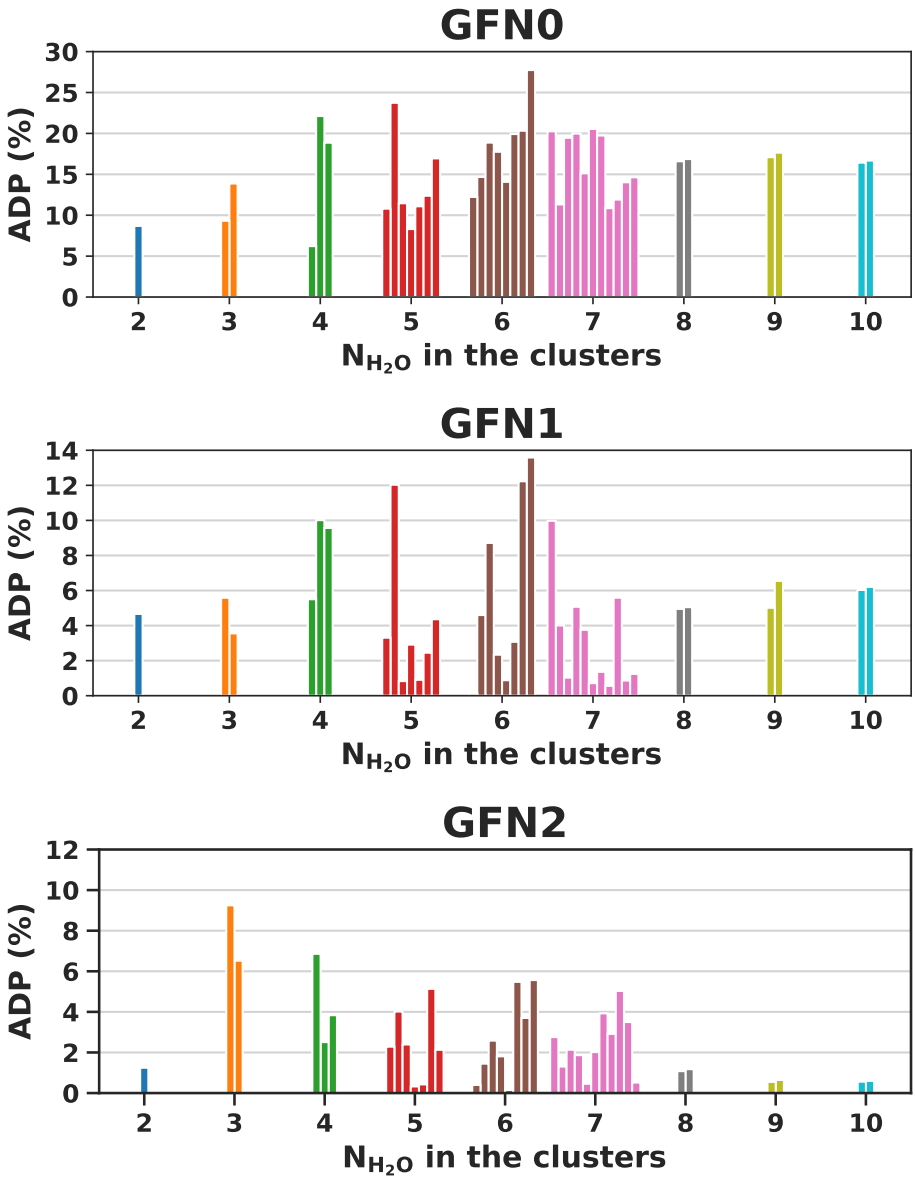



**Figure 3.2:** APD between BE GFN-xTB and CCSD(T). Note that the y-axis scale changes between the three plots.

large to a one by one comparison, therefore we relied on a global indicator for the similarities in geometries based on the well known root mean squared deviation (RMSD).

The RMSD indicates the error, through the mean square of the displacements, between two structures. It is computed using the Kabsch algorithm[48] where we define two coordinates matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, with the latter being the reference structure and $\boldsymbol{P}$ being the actual structure. First, the nuclei geometric center of both structures need to put in coincidence by a simple translation; second, the covariance matrix $\boldsymbol{H}$ of $\boldsymbol{P}$ and $\boldsymbol{Q}$ is computed:

$$\boldsymbol{H} = \boldsymbol{P}^T \boldsymbol{Q} \tag{3.4}$$

Third, we compute the optimal rotation matrix $\boldsymbol{R}$:

$$\boldsymbol{R} = \left(\boldsymbol{H}^T \boldsymbol{H}\right)^{\frac{1}{2}} \boldsymbol{H}^{-1} \tag{3.5}$$

After rotating the $\boldsymbol{P}$ matrix using $\boldsymbol{R}$ allows to compute the root mean squared deviation:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{P}_i^R - \hat{\boldsymbol{Q}}_i\right)^2} \tag{3.6}$$

where $\boldsymbol{P}_i^R$ is the i$^{\text{th}}$ point of the rotated matrix $\boldsymbol{P}$, $\boldsymbol{Q}_i$ is the i$^{\text{th}}$ point of the reference matrix $\boldsymbol{Q}$, and $n$ the number of points.

Obviously, the smaller the RMSD the better is the prediction by the GFN-xTB method.

Fig. 3.3 shows the results for each ($m$,$n$) cluster (*vide supra*). From the results obtained we can see that, in contrast with the structure energy results, GFN0 performs slightly better than GFN1, with RMSD values that are less spread-out for the different nuclearity clusters. Again, GFN2 is the most robust of the three, with results a lot less spread-out than both GFN0 and GFN1, as shown by the average RMSD values of 0.30/0.38/0.36 Å for GFN2/GFN1/GFN0. For all three methods, the structures are almost identical to the RI-MP2 ones when the number $n$ of water molecules is $n > 7$. This is related to the much-reduced configuration space of water molecules for clusters at higher nuclearity compared to the smaller ones.

Fig. 3.4 showed the comparison, for the same cluster of seven water molecules, between the RI-MP2 geometry and those optimized with the three GFNn-xTB methods.

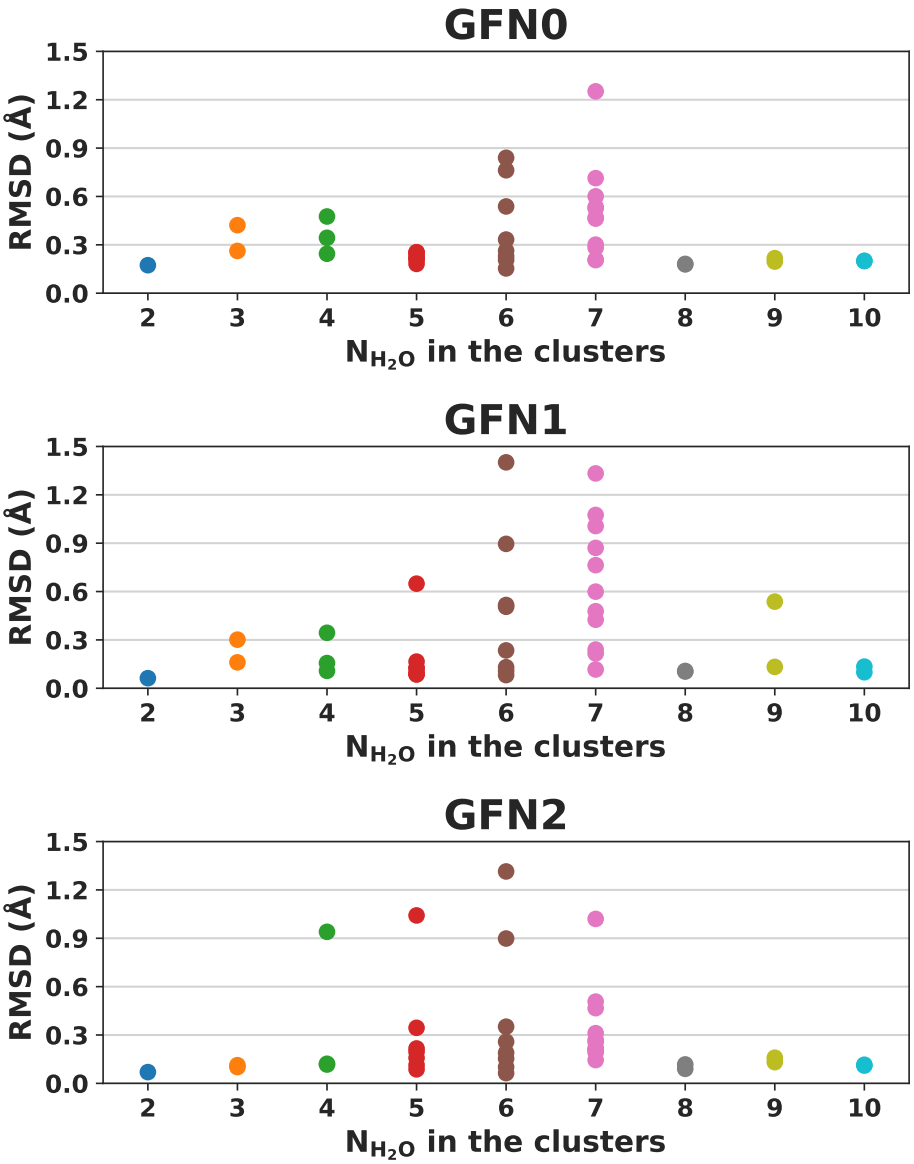



**Figure 3.3:** RMSD for the water clusters. Different points for each cluster nuclearity $m$ represent different water organization within the cluster itself.

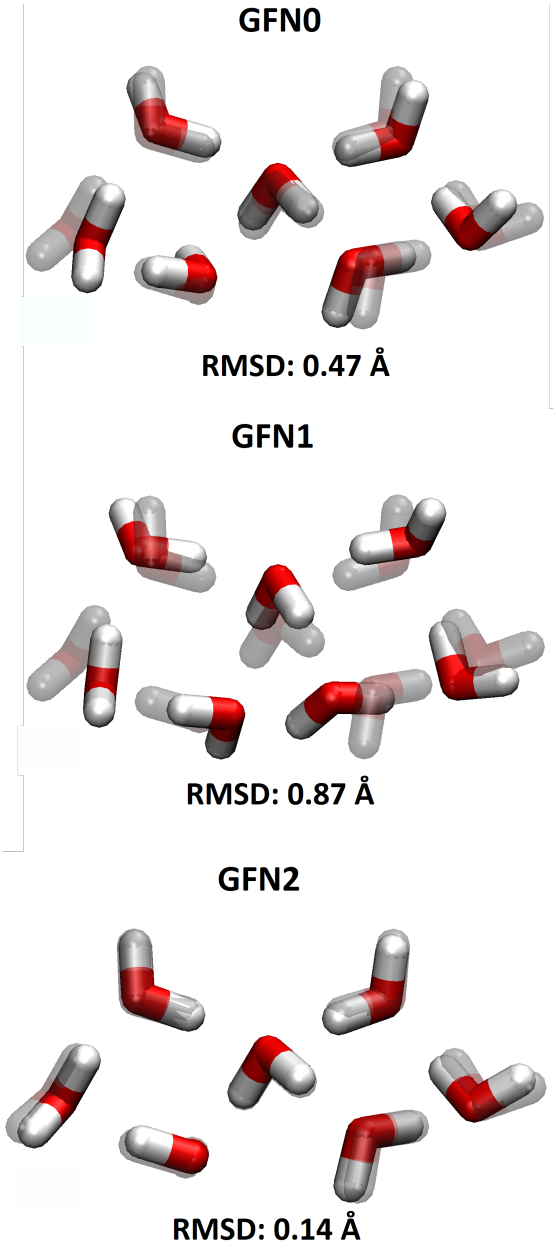



**Figure 3.4:** Superposition of the GFN structures (full body) with the ones obtained at high level of theory (fade body) for one configuration of the water cluster exhibiting seven water molecules. RMSD values in unit of Å. Taken from Germain et al. 2020[49].

In conclusion, the GFN2 method is the one behaving more closely to the accurate *ab initio* results, at least as far as the pure water clusters are considered.

### 3.5.2 Benchmarking the GFN2 BEs with ab initio periodic results

The next step in benchmarking the GFN2 method was to contrast the GFN2 BEs with that resulting from the work by Ferrero et al.[31] in which 21 molecules of interstellar relevance were studied on both crystalline and amorphous periodic slabs in an *ab initio* fashion.

The crystalline slab model of Ferrero et al.[31] is a periodic (010) proton-ordered ice (P-ice) surface containing 12 water molecules in its unit cell. The BEs were computed using a M06-2X/A-VTZ* method, for the 4 radical species of the considered set, and the B3LYP-D3/A-VTZ* for the remaining closed shell molecules. Even if Ferrero et al.[31] also adopted an amorphous periodic model (*vide supra*), we only consider the results coming from the crystalline model for the comparison with the GFN2 method.

#### 3.5.2.1 Benchmarking GFN2 against crystalline slab model results

Ferrero et al.[31] computed the binding energies of 21 species on top of a crystalline periodic slab whose unit cell contained only one type of binding site, as shown in Fig. 3.5. The binding site exhibits a dangling hydrogen (dH) and a dangling oxygen (dO), periodically repeated. To minimize the lateral interactions between species of sensible size (*i.e.* $CH_3CN$, $CH_3OH$, $H_2CO$, $HCONH_2$, and HCOOH), Ferrero et al.[31] adopted a supercell approach, in which one of the cell edges was doubled.

As stated in the Method section, GFN-xTB does not support periodic boundary conditions with the GFN2 method, so that a straightforward comparison with the Ferrero et al.[31] results cannot be carried out. Therefore, we use a large molecular icy cube, cut out from the periodic P-ice slab of Ferrero et al.[31], including 84 water molecules. The cut was achieved to ensure that both the dangling hydrogens and oxygens (the main binding sites of the crystalline slab), were properly embedded within surrounding layers of water molecules, allowing the bound molecules to move on the cluster's surface, and

to feel the long range interactions of the surrounding water molecules. Obviously, fully optimizing the cluster geometry will bring the whole structure to collapse in a very different shape, due to the need of each water molecule to maximize the H-bond contacts. To circumvent this problem, while keeping the structural memory of the crystalline ice, we fixed the coordinates of the H atoms of every water molecules at the border of the cluster pointing in the former $\boldsymbol{a}$ and $\boldsymbol{b}$ unit cell directions, leaving the H atoms pointing in the $\boldsymbol{z}$ direction free to move as well as for the water molecules in the interior of the cluster. This ensures that the dH/dO adsorbing sites are free to move in response to the adsorbate interaction. The geometry optimization of the slab was made with GFN2-xTB using the *extreme* level of optimization, ensuring a robust and stable optimized structure.

#### 3.5.2.2 BEs benchmark

To start with, we recreate, as close as possible, the initial positions of the molecules on top of the periodic slab used by Ferrero et al.[31], followed by the GFN2 optimization. If the final GFN2 geometry is different from the reference one, i.e. different bonded atoms, different H-bond length, different binding site, etc., the starting position is changed to match as close as possible the final relaxed geometry of Ferrero et al.[31] and the geometry optimization is started again. This allows to check if the GFN2 method is able to reproduce the relaxed DFT/A-VTZ* position.

The binding energies for each species were computed as follow:

$$BE_{mol} = E_{mol} + E_{slab} - E_{mol+lab} \tag{3.7}$$

where $E_{mol}$ is the energy of the isolated studied molecule, $E_{slab}$ is the energy of the isolated icy cluster, and $E_{mol+slab}$ is the energy of molecule interacting with the cluster. It is important to note that the deformation energy of the slab and molecule are embedded into $E_{mol+lab}$ and not explicitly shown in Eq. 3.7.

To confirm that the geometries obtained are true stable states of the potential energy surface, we perform a vibrational frequency calculation using GFN2 and check that all the harmonic vibrational frequencies were real, as the presence of an imaginary frequency in a geometry would results in a transition state structure. In cases where a imaginary frequency is found, GFN-xTB program provides an option to slightly perturb the geometry escaping the saddle point and let falling the structure in a nearby minimum.

The computation of the harmonic frequencies also gave the zero point energy (ZPE), as seen in Chapter 2, a correction to the BEs useful for comparison with experimental data:

$$\Delta ZPE = ZPE_{slab+mol} - ZPE_{mol} - ZPE_{slab} \tag{3.8}$$

where $ZPE_{slab}$ is the ZPE of the isolated ice cluster, $ZPE_{mol}$ is the ZPE of the isolated molecule, and $ZPE_{slab+mol}$ that of the ice cluster and the molecule in interaction.

From $\Delta ZPE$ we can obtain the zero point corrected binding energy $BE\,(0)$, for each species studied, by subtracting it to the original $BE_{mol}$:

$$BE\,(0) = BE_{mol} - \Delta ZPE_{mol} \tag{3.9}$$

As $\Delta ZPE$ is generally a positive quantity, the zero point corrected binding energy $BE\,(0)$ is always smaller to the uncorrected one.

#### 3.5.2.3 Benchmarking the final geometries

As the GFN2 method used does not incorporate PBC, some drawbacks are to be expected. For example, even with the added geometrical constraints to emulate the PBC, the hydrogen bond cooperativity of the real crystalline material is necessary underestimated in the cluster, possibly impacting the final structure and energetic of the adsorbed species. Furthermore, in PBC, the dipole moment within the infinite crystal plane is canceled out, which is not the case for the cluster, in which a dipole moment as large as 14.4 Debyes was computed in the a-b plane aligned along the $\boldsymbol{a}$ axis. This dipole moment could also impact the final positions and the BEs of the adsorbed species.

Table 3.1 shows the BE data for GFN2 and DFT from Ferrero et al.[31]. The GFN2 results are in surprisingly good agreement with the DFT ones, except for $CH_3$, CO, HCO, OH, and $O_2$. Leaving the special case of CO, all the above mentioned species are radicals, making them highly reactive. As the GFN-xTB family of methods are mainly developed for closed shell systems, the present deviations are not surprising.

A special mention is deserved to the CO molecule. Even if not a radical, its electron configuration with a triple bond and a very small dipole moment is critical for the GFN2 method, for reasons probably intrinsically bound to the tight binding parametrization. The experimental dipole moment is 0.11 D[51], with the negative pole located on the C atom [52]. GFN2 predicted a value

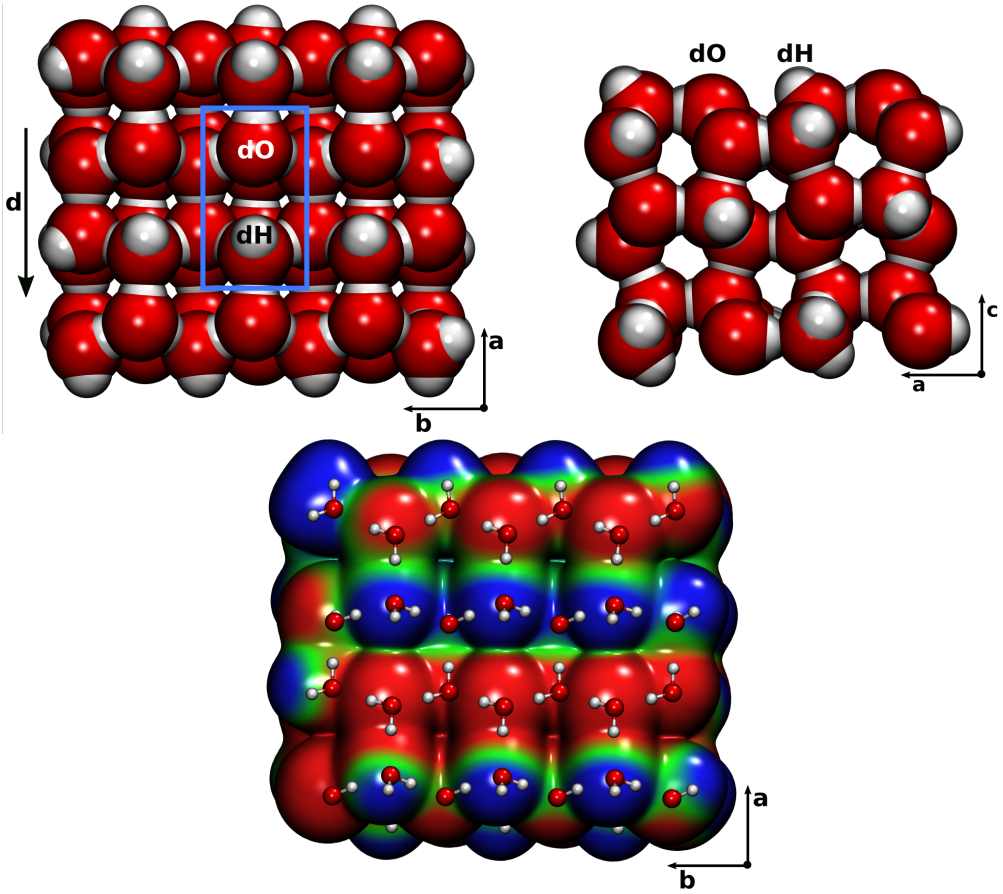



**Figure 3.5:** Slab cut from the proton-ordered P-ice crystal used by Ferrero et al.[31]. Top: Slab seen from the ab plane with the dO and dH binding sites indicated inside the blue rectangle. The net dipole moment is indicated by a black arrow. Center: Slab seen from the ac plane. Bottom: Electrostatic potential mapped on the electron density of the slab. Negative regions are in red and positive regions are in blue. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

**Table 3.1:** GFN2 BE, zero point energy corrected BE(0) and ΔZPE values (in kJ/mol) for all considered species.

| Species | BE | | BE (0) | | ΔZPE | |
|---|---|---|---|---|---|---|
| | GFN2 | DFT | GFN2 | DFT | GFN2 | DFT |
| $H_2$ | 8.47 | 9.90 | 3.48 | 2.40 | 5.0 | 7.5 |
| $O_2$ | 40.6 | 8.50 | 47.8 | 6.80 | -7.1 | 1.7 |
| $N_2$ | 16.8 | 13.0 | 12.6 | 10.4 | 4.2 | 2.6 |
| $N_2$-end | 15.6 | 13.0 | 13.1 | 10.4 | 2.5 | 2.6 |
| $CH_4$ | 21.9 | 14.0 | 16.2 | 11.2 | 5.8 | 2.8 |
| CO | 39.6 | 19.6 | 35.5 | 15.7 | 4.1 | 3.9 |
| $CO_2$ | 32.7 | 28.6 | 29.4 | 25.6 | 3.3 | 3.0 |
| OCS | 27.8 | 28.9 | 25.4 | 26.5 | 2.4 | 2.4 |
| HCl | 66.4 | 54.1 | 58.2 | 47.4 | 8.2 | 6.7 |
| HCN | 35.5 | 42.6 | 31.2 | 41.6 | 4.2 | 1.0 |
| $H_2O$ | 70.0 | 70.1 | 58.1 | 56.3 | 11.9 | 13.8 |
| $H_2S$ | 44.2 | 47.2 | 35.0 | 38.2 | 9.2 | 9.0 |
| $NH_3$ | 59.6 | 61.3 | 50.0 | 50.1 | 9.6 | 11.2 |
| $CH_3CN$ | 52.9 | 62.8 | 47.0 | 56.6 | 6.0 | 6.2 |
| $CH_3OH$ | 70.7 | 72.2 | 62.7 | 62.6 | 8.0 | 9.6 |
| $H_2CO$-SC1 | 53.0 | 48.8 | 42.5 | 39.6 | 10.5 | 9.2 |
| $H_2CO$-SC2 | 43.8 | 53.0 | 36.0 | 43.9 | 7.8 | 9.1 |
| $HCONH_2$-SC1 | 84.0 | 79.9 | 75.1 | 69.5 | 8.9 | 10.4 |
| $HCONH_2$-SC2 | 80.8 | 83.8 | 71.2 | 71.3 | 9.6 | 12.5 |
| HCOOH | 86.4 | 78.5 | 81.5 | 71.7 | 4.9 | 6.8 |
| OH | 115 | 54.4 | 127 | 43.3 | -11.8 | 11.1 |
| HCO | 38.5 | 28.9 | 31.4 | 20.2 | 7.1 | 8.7 |
| $CH_3$ | 35.8 | 21.3 | 29.8 | 14.3 | 6.0 | 7.0 |
| $CH_3$-end | 23.8 | 21.3 | 17.4 | 14.3 | 6.4 | 7.0 |

Note – For $H_2CO$ and $HCONH_2$ "SC1" and "SC2" refer to different initial positions used for the molecules on the binding sites by Ferrero et al. 2020[31]. Table taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

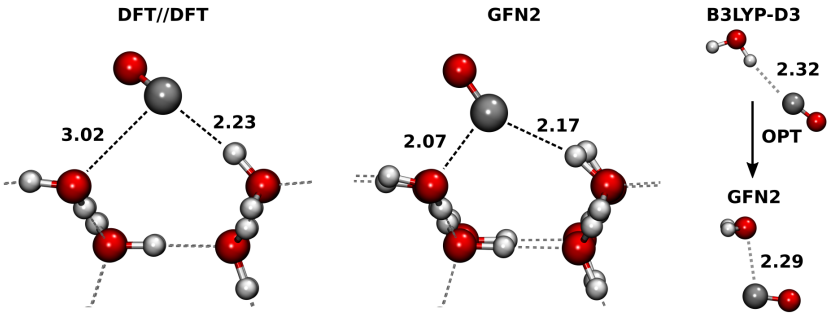



**Figure 3.6:** Left: DFT//DFT results for CO. From Ferrero et al.[31]. Center: GFN2 results for CO. Right: $H_2O - CO$ dimer with the starting geometry from B3LYP- D3/aug-cc-pVTZ on top and the GFN2 optimized geometry at the bottom. Dotted lines were added to emphasize the H--C and C--O interactions. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

as large as 0.615 D with the oxygen charge being 0.0049$e$ more negative than that of the carbon atom. This, in turn, let CO to interact through the oxygen atom when engaged in H-bonding interaction, at variance high level calculations predicting the most favorable interaction through the C-end (see Fig. 3.6-Left). Even though also the C atom seems to form a H-bond with the electropositive dH of the ice cluster, it is more attracted by the electronegtive dO as shown in Fig. 3.6-Center): this is a clear and serious artifact not accounted for by any high level calculations. To better elucidate this point, we compared the GFN2 results of CO interacting with only one water molecule with the reference structure optimized at B3LYP-D3/aug-cc-pvTZ. As shown in Fig. 3.6-Right, the C atom is more attracted to the dO than by the hydrogen atoms of the water molecule, while the reference geometry has the C atom correctly partaking in a H-bond with the water molecule. Therefore, CO molecule cannot be included in our approach, as the GFN2 method provides a physically incorrect picture of its binding with the ice surface. This is unfortunate, as CO is the second most abundant molecule in the gas-phase and the solid phase of the ISM[53–55]. Work is in progress to find out a cheap but reliable method to substitute GFN2 for the modeling of CO interacting with the icy grain.

Fig. 3.8 shows a comparison of GFN2 BEs against the DFT ones, leaving aside the $O_2$, CO, OH, and $CH_3$ cases already described. The comparison is

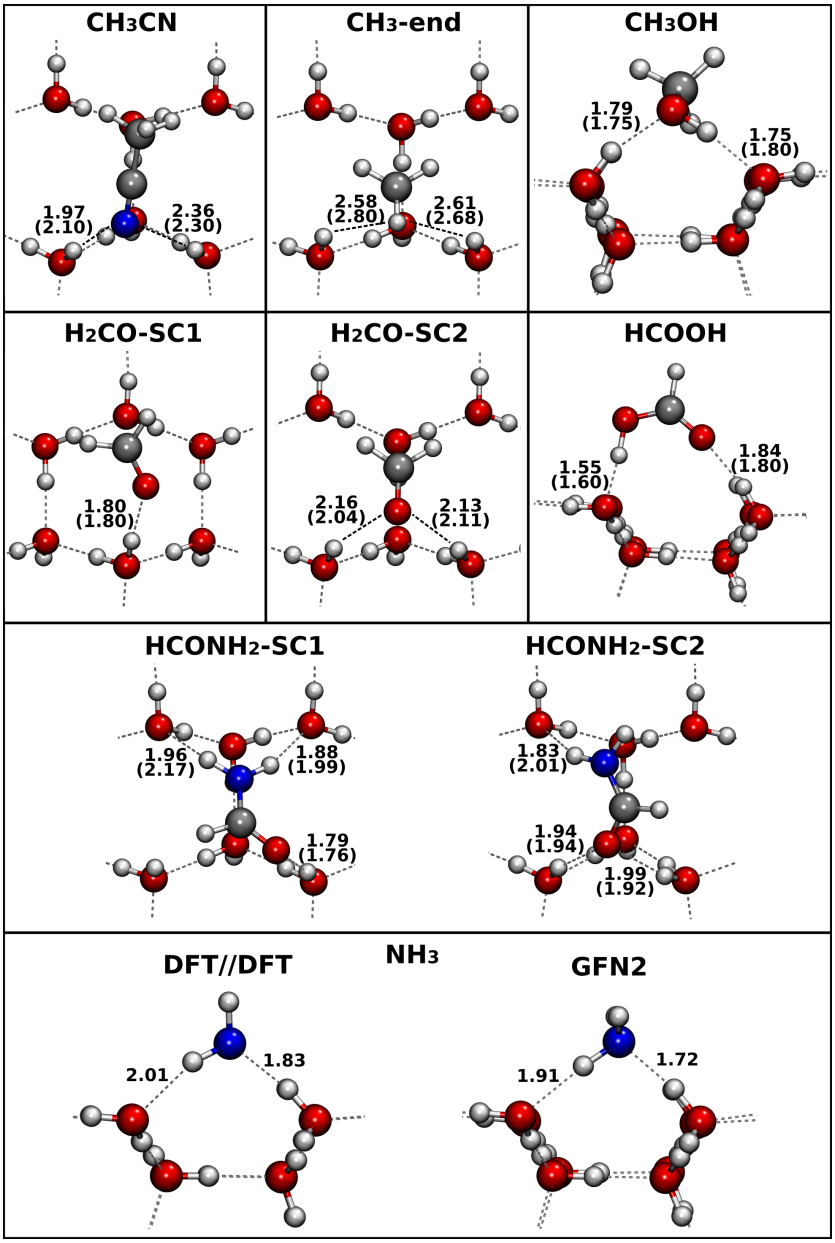



**Figure 3.7:** Comparison between DFT structures (in parenthesis) from Ferrero et al.[31] and GFN2 results for selected molecules. Distances in unit of Å. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

done for both ZPE corrected and uncorrected BES. Each point represents a molecule of the set and, as for Fig. 3.1, the black line represent the best fit line while the gray line represents the ideal line. For both BE and BE(0) the best fit line is close to the ideal line, showing that the results obtained with GFN2 are in general good agreements with the DFT ones. This agreement is a good indication that the relatively high dipole moment of the cluster does not significantly affect the GFN2-xTB binding energies.

Some larger deviation from the reference DFT//DFT results are shown by a finer analysis of the data of Fig. 3.8:

- HCl: +12.2 kJ/mol (22.6%) for the BE and +10.8 kJ/mol (22.7%) for the ZPE corrected BE(0)
- HCN: -7.14 (16.8%) for the BE and -10.4 (24.9%) for the BE(0)
- $CH_3CN$: -9.85 (15.7%) for the BE and -9.64 (17.0%) for the BE(0)
- $H_2CO$-SC2: -9.22 (17.4%) for the BE and -7.95 (18.1%) for the BE(0)
- HCOOH: +7.86 (10.0%) for the BE and +9.76 (13.6%) for the BE(0)

The geometries obtained with GFN2 are shown in Fig. 3.9 for HCN and Fig. 3.7 for the other molecules of the list. Only HCN shows significant difference in the final structure compared to that by Ferrero et al.[31]. While the nitrogen atom still is engaged in H-bond with the dH at a distance close to the DFT one, the hydrogen atom is not engaged with the dO binding site, instead preferring an adjacent dO. This due to the delicate balance between the H-bond donor/acceptor geometrical features which are handled differently between DFT and GFN2-xTB. Furthermore, some similarities of the electronic structure with those of the critical CO (*vide supra*), render also HCN a borderline molecule for the GFN2 method.

In Ferrero et al.[31] the vibrational frequencies are not computed for the amorphous slab as the number of water molecules in the corresponding unit cell is too large to be treated computationally for the whole set of cases. Instead, Ferrero et al.[31] carried out a best linear fit between the DFT crystalline BE(0) and the corresponding uncorrected BEs. The resulting parameters were used to correct directly the BEs coming from the adsorption on the

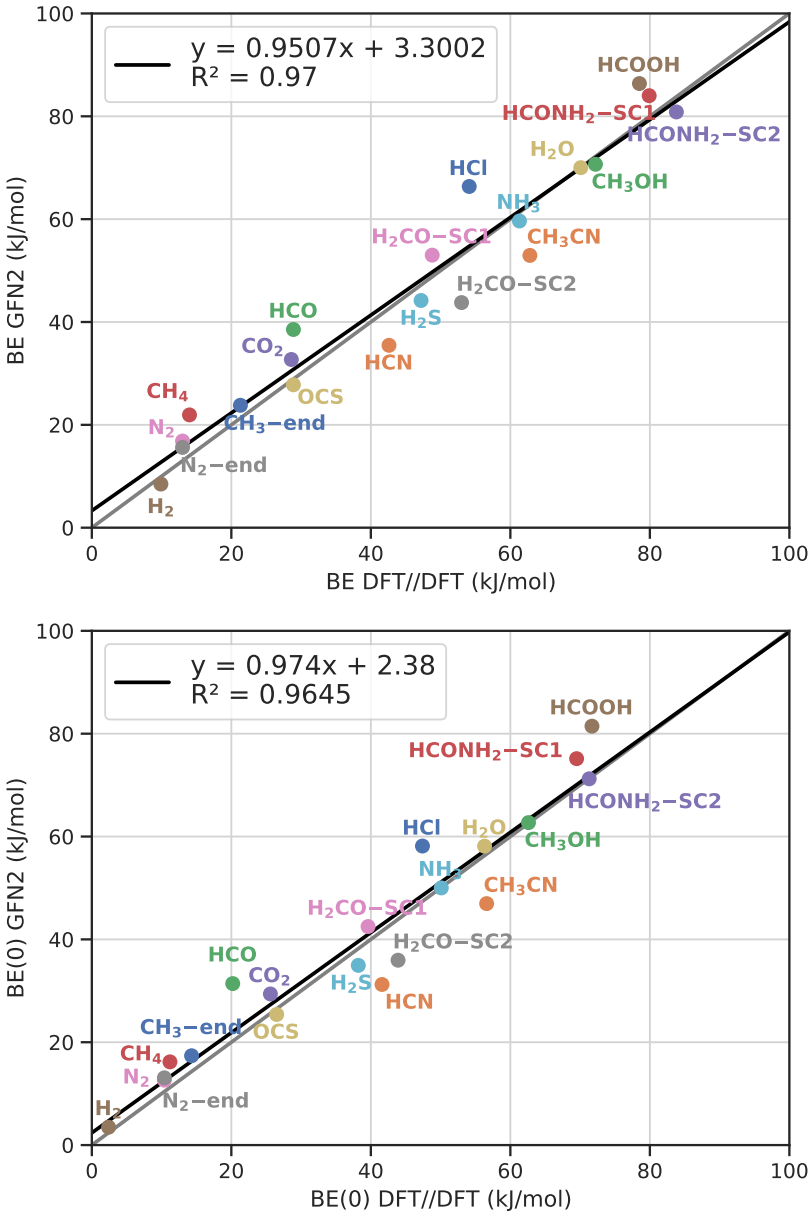



**Figure 3.8:** Top: Linear fit (black line) between BEs using DFT//DFT results available in Ferrero et al.[31] and BEs using GFN2. The gray line represents the "ideal line". Bottom: Same as up except the BEs are ZPE corrected in both cases. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

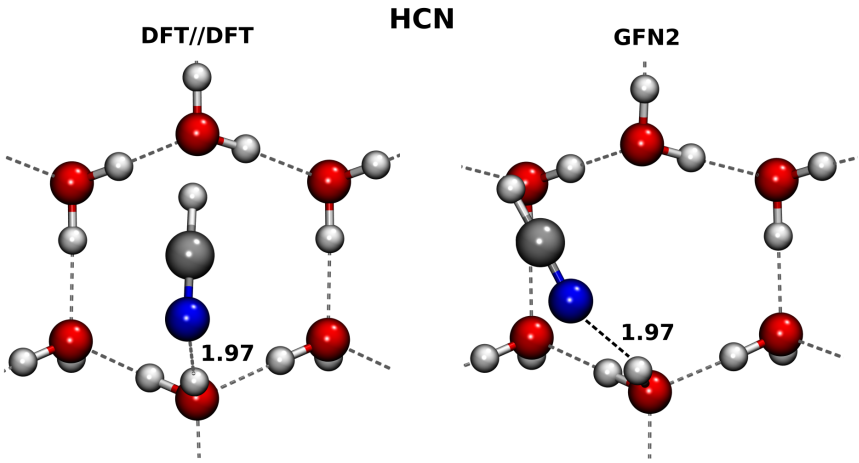



**Figure 3.9:** Left: DFT//DFT results for HCN. From Ferrero et al.[31]. Right: GFN2 results for HCN. Distances in unit of Å. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

amorphous slab. This simple approach was also successful when carried with GFN2 results, as shown in Fig. 3.10-Top, providing good fit and a huge computing saving. Unfortunately, the correlation between the ΔZPE(GFN2) with the ΔZPE(DFT//DFT) ones is only coarse, as shown by Fig. 3.10-Bottom, and is not useful.

In conclusion, also for this, more stringent benchmark, the GFN2 shows excellent performance for both structures and BEs, with some important exceptions regarding radical species and CO. The relative bad behavior for radicals is not surprising, as already many functionals fails when dealing with open shell systems. Unfortunately, GFN2 seems to have intrinsic weakness when applied to study CO and, to lesser extent, HCN molecules, both relevant in the ISM budget of molecules.

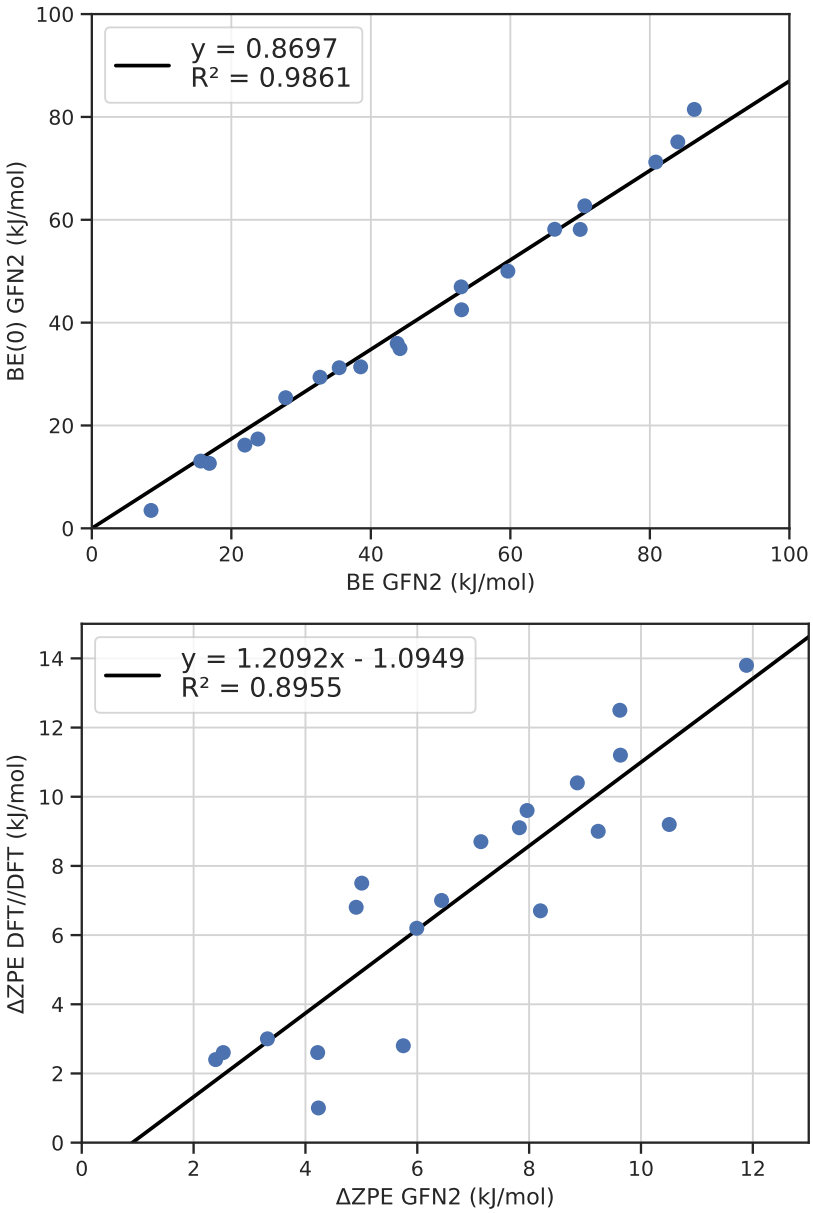



**Figure 3.10:** Top: Linear fit between GFN2 BE and BE (0). Bottom: Linear fit between GFN2 ΔZPE and DFT//DFT ΔZPE. Taken from Germain et al. 2021[50] through the Creative Commons 4.0 license.

# References


[1] R. J. Trumpler, "Absorption of Light in the Galactic System," *Publications of the Astronomical Society of the Pacific*, vol. 42, p. 214, aug 1930.

[2] H. C. van de Hulst, "The Solid Particles in Interstellar Space," *Rech. Astron. Obs. Utrecht*, no. 11 part 2, 1949.

[3] F. L. Whipple, "A comet model. I. The acceleration of Comet Encke," *The Astrophysical Journal*, vol. 111, p. 375, mar 1950.

[4] R. J. Gould and E. E. Salpeter, "The Interstellar Abundance of the Hydrogen Molecule. I. Basic Processes.," *The Astrophysical Journal*, vol. 138, p. 393, aug 1963.

[5] H. F. P. Knaap, C. J. N. van den Meijdenberg, J. J. M. Beenakker, and H. C. van de Hulst, "Formation of molecular hydrogen in interstellar space," *Bulletin of the Astronomical Institutes of the Netherlands*, vol. 18, p. 256, 1966.

[6] T. P. Stecher and D. A. Williams, "Interstellar Molecule Formation," *The Astrophysical Journal*, vol. 146, p. 88, oct 1966.

[7] D. Hollenbach and E. E. Salpeter, "Surface Adsorption of Light Gas Atoms," *The Journal of Chemical Physics*, vol. 53, pp. 79–86, jul 1970.

[8] J. R. Burke and D. J. Hollenbach, "The gas-grain interaction in the interstellar medium - Thermal accommodation and trapping," *The Astrophysical Journal*, vol. 265, p. 223, feb 1983.

[9] R. M. Logan and J. C. Keck, "Classical Theory for the Interaction of Gas Atoms with Solid Surfaces," *The Journal of Chemical Physics*, vol. 49, pp. 860–876, jul 1968.

[10] F. O. Goodman, "Application of a one-dimensional classical model of atom-oscillator scattering to atom-surface accommodation coefficient theory," *Surface Science*, vol. 60, pp. 45–64, nov 1976.

[11] M. A. Leitch-Devlin and D. A. Williams, "Sticking coefficients for atoms and molecules at the surfaces of interstellar dust grains," *Monthly Notices of the Royal Astronomical Society*, vol. 213, pp. 295–306, mar 1985.

[12] V. Buch and Q. Zhang, "Sticking probability of H and D atoms on amorphous ice - A computational study," *The Astrophysical Journal*, vol. 379, pp. 647–652, 1991.

[13] W. L. Jorgensen, "Revised TIPS for simulations of liquid water and aqueous solutions," *The Journal of Chemical Physics*, vol. 77, pp. 4156–4163, oct 1982.

[14] K. Masuda, J. Takahashi, and T. Mukai, "Sticking probability and mobility of a hydrogen atom on icy mantle of dust grains," *A&A*, vol. 330, pp. 773–781, 1998.

[15] Q. Zhang, N. Sabelli, and V. Buch, "Potential energy surface of $H\cdots H_2O$," *The Journal of Chemical Physics*, vol. 95, pp. 1080–1085, jul 1991.

[16] D. Woon, "Ab Initio Quantum Chemical Studies of Reactions in Astrophysical Ices 1. Aminolysis, Hydrolysis, and Polymerization in $H_2CO/NH_3/H_2O$ Ices," *Icarus*, vol. 142, pp. 550–556, dec 1999.

[17] H. Bergeron, N. Rougeau, V. Sidis, M. Sizun, D. Teillet-Billy, and F. Aguillon, "OH Formation from O and H Atoms Physisorbed on a Graphitic Surface through the Langmuir-Hinshelwood Mechanism: A Quasi-Classical Approach," *The Journal of Physical Chemistry A*, vol. 112, pp. 11921–11930, nov 2008.

[18] T. P. M. Goumans and S. T. Bromley, "Hydrogen and oxygen adsorption on a nanosilicate - a quantum chemical study," *Monthly Notices of the Royal Astronomical Society*, vol. 414, pp. 1285–1291, jun 2011.

[19] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, pp. 5648–5652, apr 1993.

[20] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, pp. 785–789, jan 1988.

[21] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical

analysis," *Canadian Journal of Physics*, vol. 58, pp. 1200–1211, aug 1980.

[22] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *The Journal of Physical Chemistry*, vol. 98, pp. 11623–11627, nov 1994.

[23] T. P. M. Goumans and S. T. Bromley, "Efficient nucleation of stardust silicates via heteromolecular homogeneous condensation," *Monthly Notices of the Royal Astronomical Society*, vol. 420, pp. no–no, jan 2012.

[24] S. T. Bromley, J. C. Gómez Martín, and J. M. C. Plane, "Under what conditions does $(SiO)_N$ nucleation occur? A bottom-up kinetic modelling evaluation," *Physical Chemistry Chemical Physics*, vol. 18, pp. 26913–26922, sep 2016.

[25] L. Zamirri, A. Macià Escatllar, J. Mariñoso Guiu, P. Ugliengo, and S. T. Bromley, "What Can Infrared Spectra Tell Us about the Crystallinity of Nanosized Interstellar Silicate Dust Grains?," *ACS Earth and Space Chemistry*, vol. 3, pp. 2323–2338, oct 2019.

[26] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Physical Review Letters*, vol. 77, pp. 3865–3868, oct 1996.

[27] A. Rimola and S. T. Bromley, "Formation of Interstellar Silicate Dust via Nanocluster Aggregation: Insights From Quantum Chemistry Simulations," *Frontiers in Astronomy and Space Sciences*, vol. 8, p. 49, apr 2021.

[28] J. S. Mathis, W. Rumpl, and K. H. Nordsieck, "The size distribution of interstellar grains," *The Astrophysical Journal*, vol. 217, p. 425, oct 1977.

[29] L. Zamirri, M. Corno, A. Rimola, and P. Ugliengo, "Forsterite Surfaces as Models of Interstellar Core Dust Grains: Computational Study of Carbon Monoxide Adsorption," *ACS Earth and Space Chemistry*, vol. 1, no. 7, pp. 384–398, 2017.

[30] L. Zamirri, S. Casassa, A. Rimola, M. Segado-Centellas, C. Ceccarelli, and P. Ugliengo, "IR spectral fingerprint of carbon monoxide in interstellar water-ice models," *Monthly Notices of the Royal Astronomical Society*, vol. 480, no. 2, pp. 1427–1444, 2018.

[31] S. Ferrero, L. Zamirri, C. Ceccarelli, A. Witzel, A. Rimola, and P. Ugliengo, "Binding energies of interstellar molecules on crystalline and amorphous models of water ice by ab-initio calculations," *The Astrophysical Journal*, vol. 904, no. 1, p. 11, 2020.

[32] V. Wakelam, J. C. Loison, R. Mereau, and M. Ruaud, "Binding energies: New values and impact on the efficiency of chemical desorption," *Molecular Astrophysics*, vol. 6, pp. 22–35, 2017.

[33] A. Das, M. Sil, P. Gorai, S. K. Chakrabarti, and J. C. Loison, "An Approach to Estimate the Binding Energy of Interstellar Species," *The Astrophysical Journal Supplement Series*, vol. 237, p. 9, jul 2018.

[34] W. M. C. Sameera, B. Senevirathne, S. Andersson, M. Al-lbadi, H. Hidaka, A. Kouchi, G. Nyman, and N. Watanabe, "$CH_3O$ Radical Binding on Hexagonal Water Ice and Amorphous Solid Water," *The Journal of Physical Chemistry A*, vol. 125, pp. 387–393, jan 2021.

[35] S. Andersson, A. Al-Halabi, G.-J. Kroes, and E. F. van Dishoeck, "Molecular-dynamics study of photodissociation of water in crystalline and amorphous ices," *The Journal of Chemical Physics*, vol. 124, p. 064715, feb 2006.

[36] T. Shimonishi, N. Nakatani, K. Furuya, and T. Hama, "Adsorption Energies of Carbon, Nitrogen, and Oxygen Atoms on the Low-temperature Amorphous Water Ice: A Systematic Estimation from Quantum Chemistry Calculations," *The Astrophysical Journal*, vol. 855, no. 1, p. 27, 2018.

[37] A. Rimola, V. Taquet, P. Ugliengo, N. Balucani, and C. Ceccarelli, "Combined quantum chemical and modeling study of CO hydrogenation on water ice," *Astronomy & Astrophysics*, vol. 572, p. A70, dec 2014.

[38] F. Duvernay, A. Rimola, P. Theule, G. Danger, T. Sanchez, and T. Chiavassa, "Formaldehyde chemistry in cometary ices: the case of

$HOCH_2OH$ formation," *Phys. Chem. Chem. Phys.*, vol. 16, no. 44, pp. 24200–24208, 2014.

[39] J. Enrique-Romero, A. Rimola, C. Ceccarelli, and N. Balucani, "The (impossible?) formation of acetaldehyde on the grain surfaces: Insights from quantum chemical calculations," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 459, pp. L6–L10, jun 2016.

[40] G. Molpeceres, V. Zaverkin, and J. Kästner, "Neural-network assisted study of nitrogen atom dynamics on amorphous solid water – I. adsorption and desorption," *Monthly Notices of the Royal Astronomical Society*, vol. 499, pp. 1373–1384, oct 2020.

[41] L. Song and J. Kästner, "Formation of the prebiotic molecule $NH_2CHO$ on astronomical amorphous solid water surfaces: accurate tunneling rate calculations," *Physical Chemistry Chemical Physics*, vol. 18, pp. 29278–29285, oct 2016.

[42] T. Lamberts and J. Kästner, "Influence of Surface and Bulk Water Ice on the Reactivity of a Water-forming Reaction," *The Astrophysical Journal*, vol. 846, p. 43, aug 2017.

[43] G. Molpeceres and J. Kästner, "Adsorption of $H_2$ on amorphous solid water studied with molecular dynamics simulations," *Physical Chemistry Chemical Physics*, vol. 22, pp. 7552–7563, apr 2020.

[44] S. Grimme, C. Bannwarth, and P. Shushkov, "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86)," *Journal of Chemical Theory and Computation*, vol. 13, no. 5, pp. 1989–2009, 2017.

[45] C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions," *Journal of Chemical Theory and Computation*, vol. 15, no. 3, pp. 1652–1671, 2019.

[46] P. Pracht, E. Caldeweyher, S. Ehlert, and S. Grimme, "A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules," *ChemRxiv*, no. preprint, pp. 1–19, 2019.

[47] B. Temelso, K. A. Archer, and G. C. Shields, "Benchmark structures and binding energies of small water clusters with anharmonicity corrections," *Journal of Physical Chemistry A*, vol. 115, no. 43, pp. 12034–12046, 2011.

[48] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A*, vol. 32, pp. 922–923, sep 1976.

[49] A. Germain and P. Ugliengo, "Modeling Interstellar Amorphous Solid Water Grains by Tight-Binding Based Methods: Comparison Between GFN-XTB and CCSD(T) Results for Water Clusters," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12253 LNCS, pp. 745–753, Springer, Cham, jul 2020.

[50] A. Germain, M. Corno, and P. Ugliengo, "Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12953 LNCS, pp. 632–645, Springer, Cham, sep 2021.

[51] J. Muenter, "Electric dipole moment of carbon monoxide," *Journal of Molecular Spectroscopy*, vol. 55, pp. 490–491, mar 1975.

[52] G. Frenking, C. Loschen, A. Krapp, S. Fau, and S. H. Strauss, "Electronic structure of CO—An exercise in modern chemical bonding theory," *Journal of Computational Chemistry*, vol. 28, pp. 117–126, jan 2007.

[53] L. J. Allamandola, M. P. Bernstein, S. A. Sandford, and R. L. Walker, "Evolution of Interstellar Ices," in *Composition and Origin of Cometary Materials*, vol. 90, pp. 219–232, Dordrecht: Springer Netherlands, 1999.

[54] W. Klemperer, "Astronomical Chemistry," *Annual Review of Physical Chemistry*, vol. 62, pp. 173–184, may 2011.

[55] A. C. Boogert, P. A. Gerakines, and D. C. Whittet, "Observations of the icy universe," *Annual Review of Astronomy and Astrophysics*, vol. 53, no. 1, pp. 541–581, 2015.

# Chapter 4

# ACO FROST: grain modeling and molecular adsorption

# Abstract

This chapter will present the ACO-FROST set of code, developed in python 3 during this thesis and dedicated to the production of cluster grain models and the adsorption of molecules on their surface. For both codes we will explain their functioning, the methodologies developed to form grain models and compute adsorption of molecules on their surface, how to use them, and several example to illustrate their capabilities.


## 4.1 Introduction

As showed in the previous part, quantum mechanical cluster models of interstellar icy grains are mainly obtained by taking a piece of water, often crystalline, undergoing an amorphization process through molecular dynamics simulation at high temperatures followed by a rapid cooling down of the structure[1–7]. This approach, despite being widespread, is somehow inadequate to model interstellar icy grain mantle, as in the dense molecular clouds the heat/cool cycles cannot occur, the temperature being around 10 K. Furthermore, in the ISM, the mantle is build up *via* accretion of water molecules formed *in situ* by reacting oxygen and hydrogen atoms[8–10]. This process starts at the exterior of the core grain (mainly olivine like material); when the core is covered by a water mono-layer, the formation of water continues on the mantle reaching grains of micrometer sizes. Clearly, it is very difficult to computer simulate this process without using huge computational resources. Therefore, in the following, we first simplify the system by ignoring the initial role of the core grain while assuming the water molecules as already formed. The water molecules mutually interact in such a way that the mantle grows in a fractal like manner, also allowing for the formation of some porosity for large enough grain size. This procedure also ensures to obtain the observed amorphous nature of the icy grain. The described procedure has been coded in a Python program named ACO-FROST (available freely on GitHub here) to build up the grain mantle. As the grain is used to study the adsorption energy of selected molecules of astrochemical interest, we also extended the ACO-FROST to deal with that situation (available freely on GitHub here). We are well aware of the limits of the ACO-FROST model. Nevertheless, ACO-FROST allows the interested user to built up grain of a given size in a reproducible fashion avoiding the weakness of the construction of amorphous

icy models proposed in the literature.

The grain models built up by the ACO-FROST program are still dependent on the computational resource available and on the GFN-xTB limitations. The grain models on which adsorption of molecules is carried out can be created from the ACO-FROST program or taken from a geometry file readable by the ASE[11] Python module.

The following part of the thesis will explain the details of the ACO-FROST program, and it is divided into a section devoted to explain the grain building up, and one explaining how the adsorbate molecules is distributed all around the grain surface to compute a physically sound adsorption energy distribution.

## 4.2 Computational requirements

The ACO-FROST program requires a working version of xTB[12–16], available on the GitHub xtb page of the Grimme lab (link available here), that has been configured to be called with the "xtb" alias in the console of the Linux operating system. Besides the stand alone xTB program, ACO-FROST is entirely coded in Python 3, and uses the following extensions:

- Numpy 1.19.2
- Pandas 1.1.3
- ase 3.22.1
- scipy 1.5.2

As of the time of the writing of this thesis, no more requirements are needed. The extensions versions of List 4.2 are the current version in use to develop the ACO-FROST, while older versions of the same extensions could still work without problems.

## 4.3 Grain building

The building of amorphous ice cluster is one of the two goals of ACO-FROST. In the following subsections we will present: (i) the methodology followed by explanations of how to drive the program for the clusters construction; (ii) the

different features available to the user and not presented in the methodology part; (iii) a detailed list of all the keywords present in the code and how to use them; (iv) a list of examples to illustrate extra possibilities of the ACO-FROST program.

### 4.3.1 Main methodology

As stated in the introduction, recombination of O and H atoms to form water molecules is almost impossible to simulate on a cluster of large size. To overcome this problem, we assumed the water molecules already formed and starting aggregating on a seed of one water molecule (*i.e.* we do not consider the silicate grain core). The aggregation is done iteratively, one water molecule added after the other, in sequential steps. The molecules are added with a random orientation around their geometrical center, one random rotation around each of the 3 Cartesian coordinate axes, and a random rotation around the seed which, by default, corresponds to a random point on a sphere surrounding the seed. The different methods to place molecules around the seed are presented in a following section while the default method is presented in the following paragraphs.

Before each water molecule is added, the program computes the geometrical center of the structure via a simple average over every $x$, $y$, and $z$-coordinates of the atoms:

$$x_{center} = \frac{\sum_i^N x_i}{N}; \quad y_{center} = \frac{\sum_i^N y_i}{N}; \quad z_{center} = \frac{\sum_i^N z_i}{N}; \tag{4.1}$$

where $N$ is the number of atom currently present in the cluster being built. The structure is then translated to the geometrical center, ready to accept the next water molecule. The present method of random aggregation of molecules obeys spherical symmetry with some recipes not immediately apparent from the classical spherical scheme:

$$\begin{aligned} x &= r \sin\varphi \cos\theta \\ y &= r \sin\varphi \sin\theta \\ z &= r \cos\varphi \\ \theta \in [0;\pi]\,;\ \varphi &\in [0;2\pi[ \end{aligned} \tag{4.2}$$

where $r$ is the radial length, the length from the origin to the molecule added, $\theta$ is the angle with respect to the polar axis $z$, and $\varphi$ the azimuth angle, the

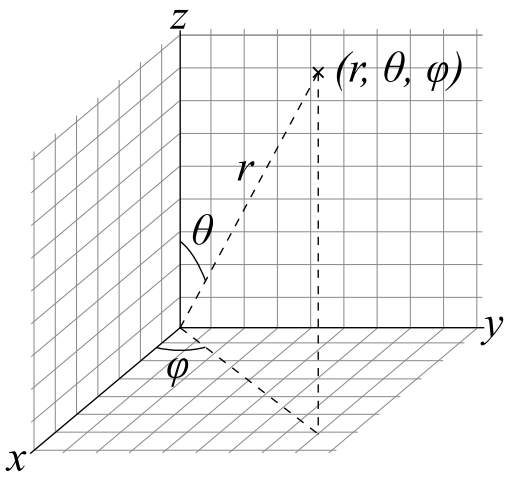



**Figure 4.1:** Spherical coordinates in the physics convention. $r$ is the radial coordinate, $\theta$ the polar angle, and $\varphi$ the azimuth angle. Image taken from Wikipedia.

angle of the $xy$ plane with respect to the $x$ axis. These coordinates are represented graphically in Fig. 4.1.

Generating the coordinates of the added water molecule following the scheme shown in Eq. 4.2 leads to biased results if the coordinates $\theta$ and $\varphi$ are chosen from an uniform distribution, between $[0; \pi]$ and $[0; 2\pi]$, respectively. Indeed, the spherical distribution of molecules is biased towards the poles of the grain, corresponding to $\theta$ values of 0 and $\pi$, as those values bring $x$ and $y$ to 0 and voids the impact of the azimuth $\varphi$ angle for these axes. To better illustrate this point, we can see a sphere as a stacking of circles perpendicular to the $z$ axis. Each circle's circumference goes from 0 for $\theta = 0$, to $2\pi R$ (in which $R$= max $r$) for $\theta = \pi/2$, to 0 again for $\theta = \pi$. For each of these circle, the uniform distribution of $\varphi$ between 0 and $2\pi$, gives to each added molecule an equal probability to belong to any circle, irrespective of its circumference. This leads to an inherent bias of molecule distribution, as each circle of different circumference will contain the same number of water molecule. This is incorrect, as smaller circles should host far less water molecules than larger ones.

To minimize this bias, we can redefine the distribution of $\theta$, and rewrite the definition of Eq. 4.2 so that it is $\cos(\theta)$ that is uniformly distributed

between $-1$ and 1. By posing $\mu = \cos\theta$, we can write:

$$\begin{aligned} x &= r\sqrt{1-\mu^2}\cos\varphi \\ y &= r\sqrt{1-\mu^2}\sin\varphi \\ z &= r\mu \\ \mu \in [-1;1]\,;\ \varphi &\in [0;2\pi[ \end{aligned} \tag{4.3}$$

That way, the spherical distribution of molecules around a seed, or growing cluster, is unbiased.

The radial length $r$ setting the distance of the added molecules, is, at first, set to the value of the farthest molecule in the growing grain plus a distance of 2.5 Å, considered as the average length of O$\cdots$H bonds. Thus, the first molecule added around the one molecule seed is at a distance of 2.5 Å, the second molecule added around the growing grain of the second molecules is added at the same distance as the farthest molecule from the grain's center of atomic positions (CAP) plus 2.5 Å, etc. As the orientation of the molecule around the growing cluster is random and based on the distance of the farthest molecule from the CAP, it may happen the added molecule to fall too far away from the growing cluster and escaping from the cluster during the subsequent geometry optimization. Thus, in case the nearest molecule belonging to the grain is farther than 2.625 Å of the newly added molecule, the value of $r$ is redefined to avoid the escaping process. The ACO-FROST code adjust the distance of the added molecule and its closest neighbor to be equal to 2.5 Åthrough an internal procedure. In essence, a straight line between the newly added molecule and the origin of the growing cluster is defined and setting the newly added molecule to be at exactly 2.5 Å of its closest neighbor. The parametric equation defining if a point of coordinates $(x, y, z)$ belongs to a straight line is defined as:

$$\begin{cases} x &= x_A + at \\ y &= y_A + bt \\ z &= z_A + ct \end{cases} \tag{4.4}$$

where $(x_A, y_A, z_A)$ are the coordinates of the CAP of the newly added molecule, $(a, b, c)$ are the coordinates of a $\vec{u}$ vector parallel to the straight line, and $t$ is a parameter. To find the coordinates we need to solve the simple quadratic equation:

$$t^2 \left\|\vec{u}\right\|^2 - 2t\vec{u}\cdot\vec{M_n} + \left\|\vec{M_n}\right\|^2 = 2.5^2 \tag{4.5}$$

where $M_n$ is the position of the closest neighbor, $\left\|\vec{M}_n\right\|$ its distance from the origin of the growing cluster, and $\|\vec{u}\|$ the norm of the $\vec{u}$ vector. The solutions to Eq. 4.5 are of the form:

$$t = \frac{-\vec{u}\cdot\vec{M}_n \pm \sqrt{\left(\vec{u}\cdot\vec{M}_n\right)^2 - \vec{u}\cdot\left(\vec{M}_n{}^2 - 2.5^2\right)}}{\|\vec{u}\|} \tag{4.6}$$

If the square root inside Eq. 4.6 is equal to 0 then only one solution exists, if it is superior to 0 then two solutions exists and the code selects the one where the newly placed molecule is farthest from the origin of the growing grain. If no adequate solution is found, we perturb a little the system and start again until a valid position, with distances between 2.375 Å and 2.625 Å of the closest molecule, is selected.

Following the addition of a molecule, a geometry optimization of the structure is done at the GFN-xTB level of theory, in which the level of theory (GFN1, GFN2, GFN-FF etc.) can be chosen by the user. Every ten added molecules, we run a NVT molecular dynamic simulation at 10 K temperature lasting 1 ps by default, at the same chosen level of theory. This MD computation is meant to account for the possible energy transfer toward the grain coming from the energy released from each formed O-H bond. As this energy is large (of the order of 20 kJ/mol), its transfer toward the grain will in principle heat it up and causing structure restructuring. Though recent studies from our lab (Pantaleone et al.[17,18]), at *ab-initio* MD level, of the $H_2$ and HCO radical formation on amorphous ice (simulated withing the periodic boundary conditions (PBC) models) did not show signs of local crystallization, they showed the transfer of the covalent bond formation energy in the kinetic energy of the water molecules of the ice grain. Therefore, we use this short MD at 10 K to mimic, in approximate fashion, the missing energy transfer due to the simple GFN-xTB static geometry relaxation.

As a function of external keywords and after a given number of added water molecules, geometry optimizations can be carried out at a different level of theory. As a default, ACO-FROST performs geometry optimization at the GFN-FF level after each added molecule, and at the GFN2-xTB level every 100 added water molecules. This choice ensures a substantial speedup in the grain growth process.

When the desired number of aggregated water molecule is reached, the program perform a "sanity check" of the grain surfaces. This test avoids hav-

ing the most exterior water molecules engaged in only one H-bond with the underneath grain. Examples are illustrated in Figure 4.2. When found, they are removed and the process start again at the previous cycle of accretion. This "sanity check" is performed until no under-connected water molecules resulted.

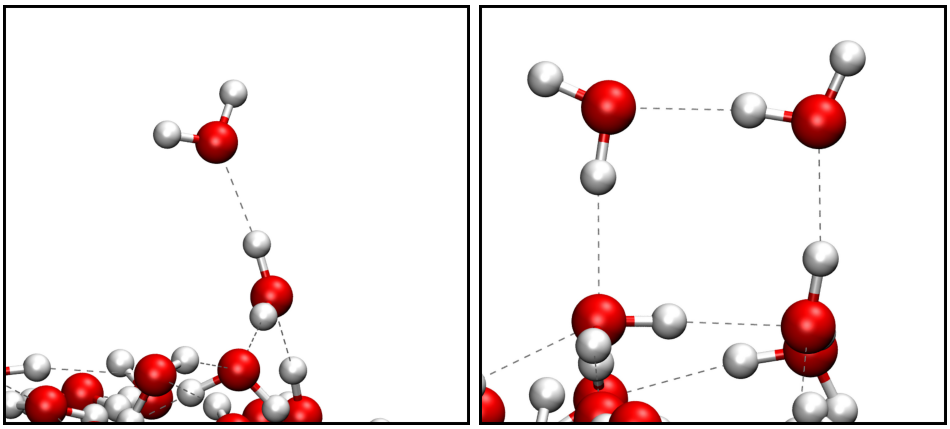


**Figure 4.2:** Two types of surface water configuration deleted by the "sanity check" presented in section 4.3.1. Left: A water molecule partaking in only one Hbond with the surface of the grain. Right: Four water molecules forming a square structure on the surface.

A flowchart of the default grain building process (several hundred molecules with GFN-FF geometry optimization after each added molecule, GFN-FF MD computation every ten added molecules, GFN2-xTB geometry optimization at selected interval, and final surface quality check) is showed in Fig. 4.3.

### 4.3.2 Important features

The ACO-FROST program includes a number of features to fine tune grain models of various sizes, shapes, and compositions using the GFN-xTB set of methods. Several features have already been presented in the previous subsection of this thesis. In the following, we describe what left unexplained, as a list of keywords.

ACO-FROST allows the adsorption of molecules on an existing cluster. In this type of calculation, the preexisting structure is considered as a seed, and its size or composition is not taken into account by the program. For

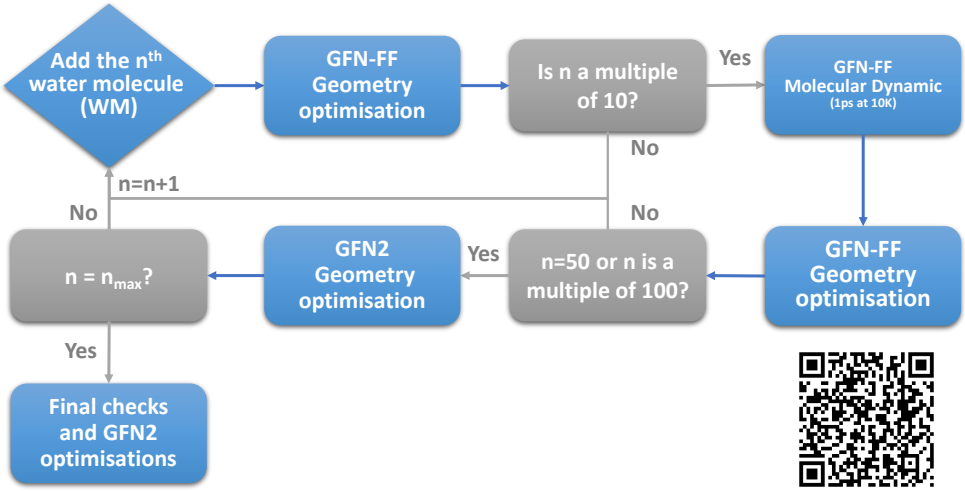



**Figure 4.3:** Flowchart of the default grain building process. Taken from Germain et al. 2022[19] through the Creative Commons 4.0 license. By pointing the phone camera to the QR-code the growing process is shown for a large ice grain.

instance, starting from an existent seed of 40 water molecule and setting to 200 the water molecules to be added, the final size will be 240 molecules. The program allows also to fix the geometry of this seed via the "constrain" keyword of xTB to avoid relaxation of the internal seed provided as an external file.

ACO-FROST allows also to add molecules without any intermediate geometry optimization, or too choose the interval in-between geometry optimization.

The program allows the construction of a clusters made of "dirty" mixed ice, either completely mixed, or built in layers, depending on the user target. The list of molecules available to form the ice is listed in a comma-separated values (CSV) file given with the ACO-FROST code. Each line of the CSV file represent one molecule of interstellar interest linked to a sub-folder of the installed ACO-FROST folder, containing its structure.

For the random distribution of molecules during the grain growing process, ACO-FROST allows the use of several methods. One, the *normal* method was already explained in the previous section. The remaining ones will be explaining in the following:

- **sbiased**: random molecule distribution following the biased spherical distribution ($\theta$ as a uniform random distribution between 0 and $\pi$). See coordinate system as defined by Eq. 4.2.

- **plane**: Same as "sbiased", but with a constant value of $\theta = \pi/2$.

- **hollow**: The "hollow" method uses the "normal" method distribution with the option to choose a minimum distance at which molecules are added from the origin to create a more hollow structure. This method is better used coupled with the **--optimisation_cycle** keyword, (*vide infra*).

- **spherical**: This method coupled with the **--optimisation_cycle** keyword (*vide infra*), allows molecules to be placed in spheres, i.e. the placement of molecules is not random for this method, but deterministic. It places, first, one molecule at the origin of the structure and then adds hollow spheres of molecules around it in layers. Each layer has a radius of 3 Å larger than the previous one (so the first level sphere is at a radius of 3 Å, and the second at a radius of 6 Å). For each layer, the $\theta$ distribution is divided into a list of $level \times 2 + 1$ values. The first level possess 3 values that are $\left[0, \frac{\pi}{2}, \pi\right]$, the second possess 5 values that are $\left[0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi\right]$, the third possess 7 values that are $\left[0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \frac{5\pi}{6}, \pi\right]$, etc. For each $\theta$ value of each level, the $\varphi$ coordinate (set between 0 and $2\pi$, $2\pi$ excluded) is also divided into different values depending on the level of the sphere being produced and the position of the $\theta$ angle ($\theta$ angles at the poles will have a subdivision of $\varphi$ less dense than $\theta$ values around $\frac{\pi}{2}$). If the desired grain size is already achieved in the middle of sphere, the size will be rounded to complete the whole sphere. Fig 4.4 shows the firsts three spheres that can be built.

A complete description of the available keywords is shown in the appendix of the chapter.

The ACO-FROST program allows also the use of input file for more controlled input parameters.

#### 4.3.2.1 Input file

The input file is an important feature of the ACO-FROST grain building code, as specific features of the program are only available through the input file. It is separated in 2 blocks. Each block starts with the string

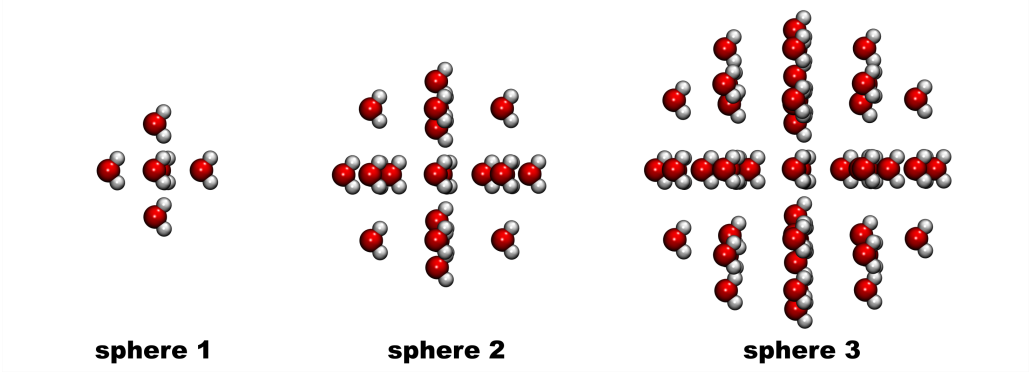



**Figure 4.4:** Examples of input structure (structure before optimization) obtained with the "spherical" keyword of molecule distribution and the optimization cycle keyword. The first level contain 7 water molecules, the second 25, the third 63.

*$name_of_the_block* and ends with the string *$end*. The first block, the "building" block, should contain the list of molecules followed by the total number of that species in the grain. If the sum of each species is equal to 100, the specified number are taken as percentages instead of the absolute number of molecules needed to form the grain. Each number has to be an integer, the code will stop working if floats are used. The "building" block is presented as follows:

```
$building #Block indicating the mixture of ice
MOL1 i #Formula and wanted number of molecules 1
MOL2 j #Formula and wanted number of molecules 2
MOL3 k #Formula and wanted number of molecules 3
... #More formulas can be added
$end
```

For instance, a grain made of 60 water molecules and 20 ammonia molecules, requires a building block like:

```
$building #Block indicating the mixture of ice
H2O 60
NH3 20
$end
```

If a mixed ice grain is not needed, the building block can be ignored while selecting the molecule using the **--molecule** keyword.

The second block is the "parameters" block and includes as keyword those

available for the input line. The main difference is the addition of a "distrib" line taking the values *0* or *1* indicating if the molecules in the "building" block should be added in the order indicated in the block (*0*), or randomly (*1*). The "parameters" block is presented as follows:

```
$parameters #parameters block
distrib i #1 for random, 0 for list
gfn x #level of GFN-xTB method to use
size i #size of grain wanted
mixed x i #See previous subsection
MD x i #See previous subsection
struct x #See --structure of the previous subsection
fixstruct #See --fixed_structure of the previous subsection
restart i #See previous subsection
rand x #See --random_law of previous subsection
opt i #See --optimisation_cycle of the previous subsection
check #See --check_surface of the previous subsection
finalgfn2 #See previous subsection
agermain2022 #See previous subsection
$end
```

Taking the "building" example and adding parameters for making a grain where molecules are added randomly from the list indicated in the input file, following the random law "sbiased", and performing GFN-FF geometry optimization every molecule added and GFN2-xTB geometry optimization every 40 molecule added with have:

```
$building #Block indicating the mixture of ice
H2O 60
NH3 20
$end
$parameters #parameters block
distrib 1
gfn ff
mixed 2 40
$end
```

If these input commands are written in an input file called *input.inp*, the command line needed to use it is simply: `--input input.inp` or `-inp input.inp`.

### 4.3.3 Examples

To better illustrate the different keywords and their action, we show 10 examples of input line and the description of the resulting grain model.

**Example 1** --size 50 equivalent to -size 50 Produces a grain model made of 50 water molecules. Each water molecule is added using the *normal* molecule distribution, and a GFN2-xTB geometry optimization is performed after every added molecule. Figure 4.5 shows a grain model obtained through this example.

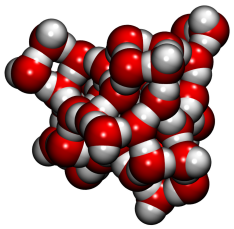


**Figure 4.5:** Grain models obtained by using the command lines of examples 1.

**Example 2** --size 200 --agermain2022 equivalent to -size 200 -agermain2022 Produces a grain model made of 200 water molecules following the methodology presented in Germain et al. 2022[19].

**Example 3** --size 150 --gfn 1 --molecule NH3 --random_law sbiased equivalent to -size 150 -g 1 -mol NH3 -rand sbiased Produces a grain model made of 150 ammonia molecules. Each molecule is added iteratively using the *sbiased* distribution, and a GFN1-xTB geometry optimization is performed after every added molecule. Figure 4.6 shows a grain model obtained through this example.

**Example 4** --size 300 --gfn ff --molecule H2O --restart 143 equivalent to -size 300 -g ff -mol H2O -r 143 Restarts a previous computation of a grain model made of 142 water molecules and optimized at the GFN-FF level after each molecule is added randomly via the *normal* law. The computation continues until the grain model is made of 300 water molecules.

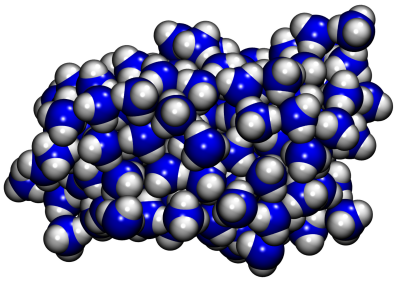
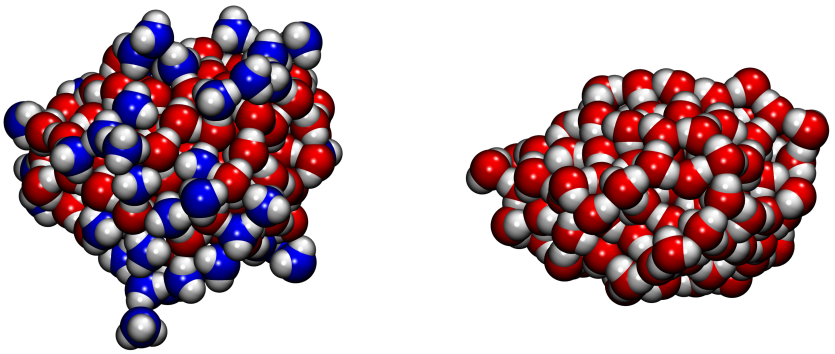


**Figure 4.6:** Grain model obtained by using the command lines of examples 3.

**Example 5** --structure pre-existing-model.xyz --size 50 --gfn 1 –molecule NH3 equivalent to -structure pre-existing-model.xyz -size 50 -g 1 -mol NH3 Takes a pre-existing grain model contained in the file *pre-existing-model.xyz*, adds iteratively 50 molecules of ammonia to it using the *normal* random law, and performs a GFN1-xTB geometry optimization after every water molecule added. Figure 4.7 shows a grain model obtained through this example.



**(a)** Example 5 **(b)** model of example 5

**Figure 4.7:** Grain model obtained by using the command lines of examples 5. Right figure represent the pre-existing grain model used in example 5.

**Example 6** --size 45 --gfn 0 --final_gfn2 equivalent to -size 45 -g 0 -final_gfn2 Produces a grain model made of 45 water molecules added randomly using

the *normal* random law, performs a geometry optimization at the GFN0-xTB level after each molecule added, and performs a final GFN2-xTB relaxation after the last GFN0-xTB geometry optimization.

**Example 7** --size 200 --optimisation_cycle 50 --check_surface equivalent to -size 200 -opt_cycle 50 -check_surface Produces a grain of 200 water molecules added randomly via the *normal* law, performs a GFN2-xTB optimization each 50 molecule added, and performs a "sanity check" of the final structure (see Section 4.3.1).

**Example 8** --size 200 --gfn 2 --mixed ff 20 equivalent to -size 200 -g 2 -mix ff 20 Produces a grain of 200 water molecules added using the *normal* random law. After each molecule added, a geometry optimization using GFN-FF is performed, if the molecule added is a multiple of 20, a GFN2-xTB geometry optimization is also performed.

**Example 9** --size 200 --gfn 2 --mixed ff 20 --MD ff 5 equivalent to -size 200 -g 2 -mix ff 20 -MD ff 5 Same as previous example, but this time, if the molecule added is a multiple of 5, a MD computation of 1 ps at 10 K is also performed. For example, for the 20th molecule added, a GFN-FF geometry optimization will be performed, followed by: (i) a MD computation; (ii) another GFN-FF computation; (iii) a final GFN2-xTB geometry optimization, before the 21st molecule is finally added.

**Example 10** --input inp.inp equivalent to -inp input.inp Produces a grain model following the input information contained in the *input.inp* file. The *input.inp* file needs to follow the convention presented in Section 4.3.2.1. For the following input file

```
$building #Block indicating the mixture of ice
H2O 45
NH3 125
CH3OH 30
$end
$parameters
distrib 1
GFN 0 #level of GFN-xTB method to use
$end
```

A 200 molecules grain model made of a mix of 45 water molecules, 125 ammonia molecules, and 30 methanol molecules, following the *sbiased* random law, will be produced. Each type of molecule will be added in a random order and a GFN0-xTB optimization will be performed after each molecule added.

## 4.4 Molecular adsorption at grains

The simulation of the adsorption process between interstellar molecules of interest ($CH_3OH$, $NH_3$, $H_2O$, $H_2S$, etc.) and the icy mantle of a grain is the next, most important step, in which ACO-FROST plays a relevant role to arrive to the definition of binding energy (BE) distribution of a given adsorbate. Indeed, it is through the BE distribution that comparison with the thermal desorption process (TPD) experiments becomes physically sound. In the following subsections we will present the methodology needed to compute the binding energies, the different features available to the user and not presented in the methodology part, a detailed list of all the keywords present in the code and how to use them. Finally, a list of examples to illustrate using ACO-FROST to compute the BE is also illustrated.

### 4.4.1 Main methodology

In essence, the BE is defined as the energy difference from the complex between the icy grain and the adsorbate with respect to the free constituents. In this work, we only considered the adsorption of *one molecule* per icy grain. In other words, *no multiple adsorption* has been considered, even if ACO-FROST can be easily tweaked to handle this situation. Therefore, in the following, all described steps are meant to properly position *one adsorbate* molecule around *all possible adsorption sites* of the previously generated icy grains.

As with the grain building process, ACO-FROST uses GFN-xTB to perform geometry optimizations as well as to compute the BEs. The first step is to select a specific already formed ice grain and the adsorbate molecule of interest. At that point, ACO-FROST generates all possible adsorption starting positions to be subsequently optimized to arrive to relatively accurate BE values. The algorithm to generate the adsorbate starting positions differs from that described in Section 4.3.1 to add water molecules during the

**Table 4.1:** 12 original vertices for the first level grid. The parameter $\phi$ is equal to $2\cos(\pi/5)$. Taken from Teanby.[20]

| Vertex | x | y | z |
|---|---|---|---|
| 1 | 0 | $\phi$ | 1 |
| 2 | 0 | - $\phi$ | 1 |
| 3 | 0 | $\phi$ | - 1 |
| 4 | 0 | - $\phi$ | - 1 |
| 5 | 1 | 0 | $\phi$ |
| 6 | 1 | 0 | - $\phi$ |
| 7 | - 1 | 0 | $\phi$ |
| 8 | - 1 | 0 | - $\phi$ |
| 9 | $\phi$ | 1 | 0 |
| 10 | - $\phi$ | 1 | 0 |
| 11 | $\phi$ | - 1 | 0 |
| 12 | - $\phi$ | - 1 | 0 |

Note – To project these points into a sphere, each vertex needs first to be normalized by dividing by $\sqrt{1+\phi^2}$ and then multiplied by the desired sphere radius.

grain growing process (*i.e.* using a uniform distribution of spherical coordinates). That choice would generate an exceedingly large number of starting positions so that we opted instead for a grid of discrete points taken from Teanby[20]. This grid is superimposed around the grain model, provides a set of equally separated starting positions, and gives a complete and uniform coverage of the grain ensuring an unbiased sampling of surface sites.

The grid is made initially of the 12 vertices of an icosahedron (the coordinates are shown in Table 4.1). These vertices serve as starting positions for the molecule to be adsorbed on the grain. This initial grid can then be subdivided into additional points, increasing the density of adsorption starting points. The subdivision is performed by producing new points via the simple addition of the coordinates of every coupled grid points (points that are first neighbors). An illustration of the first, second and third grid levels is available on Fig. 4.8.

After selecting the grid desired level, each vertex is projected on the grain's surface at a distance of 2.5 Å as explained in Section 4.3.1 (This pro-

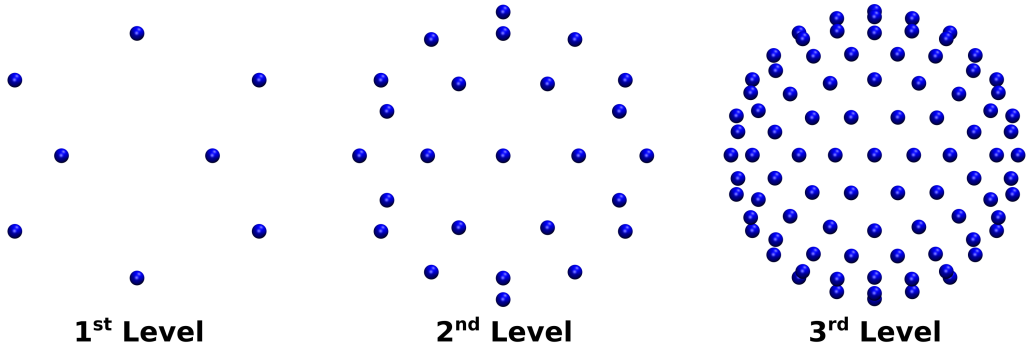



**Figure 4.8:** 3 different levels of grids. Level 1: 12 vertices. Level 2: 42 vertices. Level 3: 162 vertices. Taken from Germain et al. 2022[19] through the Creative Commons 4.0 license.

cess is illustrated on Fig. 4.9). Each of the vertices represent then the average center of adsorbate nuclei where a single adsorbate molecule is positioned. We adopted two different methods to compute the BE values. The first one is the straightforward application of the BE definition, in which the structures of the complex and the separate components (adsorbate and grain) are fully optimized at the selected GFN-xTB level. While there is nothing wrong with this approach, the interaction with the adsorbate can cause subtle changes in the grain geometry which, once re-optimized in absence of the adsorbate, will give an energy lower than that of the pristine ice cluster. Therefore, the supposed reference of the original ice cluster is no longer the absolute minimum, biasing in an unpredictable way the BE values. This is clearly due to two factors: the somehow limited size of the simulated grain, as this effect will disappear for very large cluster sizes; the relatively weak interactions (H-bonding) holding the grain in place, which causes some fluxionality to the grain. For instance, this effect will never appear when considering covalent or ionic bounded clusters (for instance $SiO_2$, MgO, $TiO_2$ clusters). A second alternative, bypassing this problem, computes the adsorption of the molecules on the grain surface in several steps. In the first step, the geometry of the entire grain is fixed at the free grain optimized structure, while only the adsorbate is allowed to freely move and relax. During this step, the adsorbate get closer to the grain surface. The fixing of the grain geometry is attained using the *constrain* keyword of GFN-xTB, which apply a force constant of 0.05 hartree.bohr$^{-2}$ to the distance among the selected atoms (in

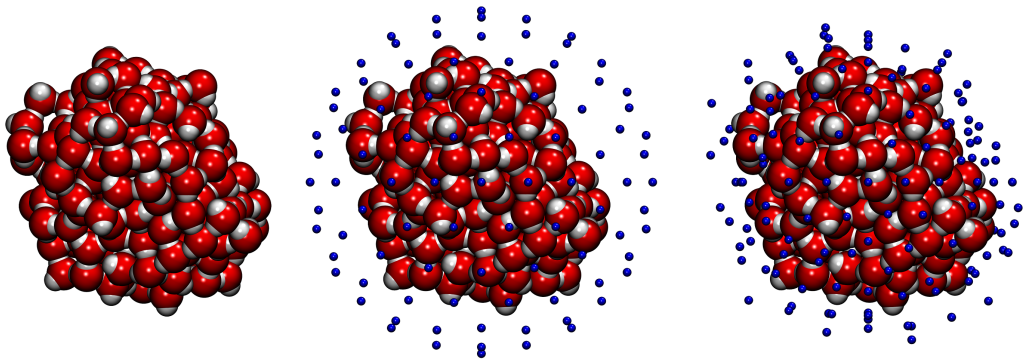


**Figure 4.9:** From left to right: Grain model studied; third level grid of 162 vertices placed around the grain model; same grid projected onto the grain. Taken from Germain et al. 2022[19] through the Creative Commons 4.0 license.

this case, the whole grain model). In the second step, a selection of grain atoms within a 5 Å radius apart from the center of the adsorbate nuclei is selected and the positions of the selected atoms is no longer constrained. A second geometry optimization only relaxing the adsorbate and the selected atoms of the grain is then carried out. After this second step, a further check is performed to ensure that no new grain atom has penetrated the 5 Å radius as a consequence of the optimization. If this is the case, ACO-FROST sets the adsorbate in the next grid point, ready for a new BE calculation. Otherwise, the list of atoms is updated and a new optimization is carried out up to the point where the list of moving atoms does not change. Our experience showed that rather few cases required more than three iterations to settle the optimization. For each fully converged case, the full set of harmonic vibrational frequencies is computed to ensure the final structure to be a true minimum on the potential energy surface. As a sub-product, the zero point energy (ZPE) correction can be used to correct the electronic BE. An illustration of these three steps is shown on Fig. 4.10 in the form of a flowchart. It is worth noting that the final set of the optimized structures does contain redundant structures with the same geometry and BE value. This is because adsorbates starting at different grid point positions may collapse during the geometry optimization to the same end point of the potential energy surface. Therefore, an algorithm will clean up the redundant structures to arrive to a final irreducible set. This process may purge a significant amount of the

total structures.

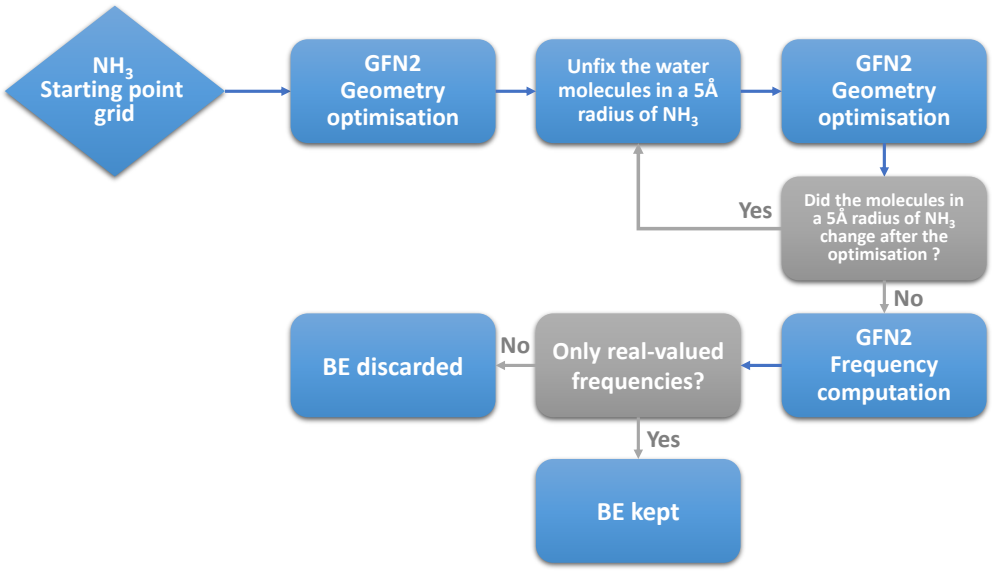



**Figure 4.10:** Flowchart of the default binding energy computation process for the adsorption of $NH_3$ on a grain model. This process is performed for each of the $NH_3$ molecule generated by the grid level selected by the user. Adapted from Germain et al. 2022[19] through the Creative Commons 4.0 license.

### 4.4.2 Important features

The ACO-FROST program includes a number of features to fine tune the calculation of the BEs. The specific keywords needed to that purpose will be described in the next subsection.

The main ACO-FROST feature of the BE energy calculation is the space sampling around the icy grain cluster using a grid of discrete points separated by almost equal distances. The initial grid is made of 12 points and can be divided to obtain tighter grid levels. The division is done by bisecting the segments formed by any two points next to each other. Thus, a lower grid level is always contained into a higher one, meaning that when a grid level is increased, points are only added, no points are deleted. It is therefore possible to increase the coverage of a grain from one grid level to a tighter

one without throwing away the results of the previous grid level. Grid levels can also have their segments trisected to offer a better control over the grain coverage. Figure 4.11 shows a first level grid being trisected to obtain a grid of 92 points. The downside of trisecting a grid is the breaking of the continuation between grid levels, hence, a second level grid of 42 points is not contained in a first level trisected grid of 92 points.

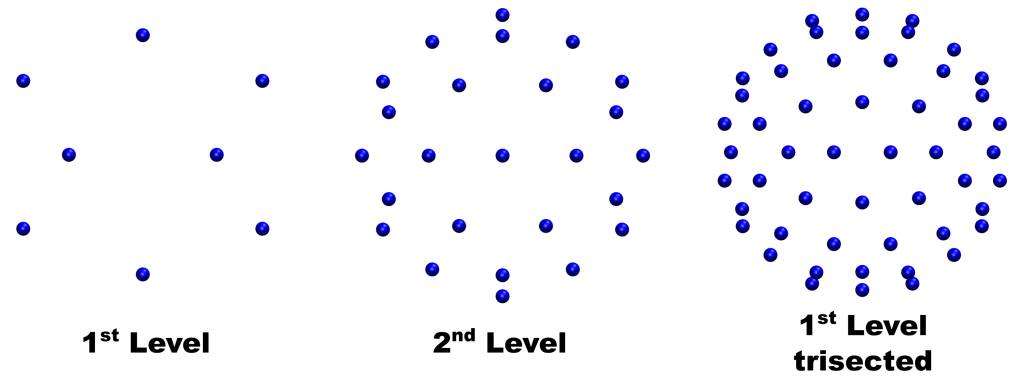



**Figure 4.11:** Example of grid level trisection. Left and center images illustrating the first (12 vertices) and second (42 vertices) level grid are taken from Fig 4.8, right image represent the trisection of the first level containing 92 vertices.

When adsorbates are placed onto the grain, they are randomly rotated around the three Cartesian axes relative to their geometrical center. This random rotation can influence their final conformation. The ACO-FROST BE code allows the use of multiple random rotation per grid points. Thus, selecting a second level grid containing 42 grid points, and requesting 3 rotations per grid points, creates 126 structures where each structure shares the same grid point with 2 others.

The list of molecules available is identical between the grain building code and the molecular adsorption codes.

A complete description of the available keywords is shown in the appendix of the chapter.

### 4.4.3 Examples

**Example 1** grain.xyz –gfn2 –level 2 –molecule NH3 Computes the BEs of ammonia around the grain model contained in the file *grain.xyz* using a third

level grid made of 162 vertices. Fixed optimization are performed first, then unfixed (with a default radius of 5 Å), and finally the vibrational frequencies are computed. This example is illustrated with Fig. 4.12.

**Example 2** grain_other.xyz --gfn 2 --level 1 --molecule CH4 --no_freq equivalent to grain_other.xyz -g 2 -level 1 -mol CH4 -nfreq Takes a grain model from the *grain_other.xyz* file. Computes the fixed and unfixed geometry optimization of Methane with a second level grid made of 42 vertices and at the GFN2-xTB level. Does not compute the vibrational frequencies.

**Example 3** grain_other.xyz --gfn 2 --level 1 --molecule CH4 --only_freq equivalent to grain_other.xyz -g 2 -level 1 -mol CH4 -ofreq Takes the previous' example computation and continues it by only computing the vibrational frequencies at the GFN2-xTB level.

**Example 4** model.xyz --gfn 0 --level 1 --molecule CH3OH --radius 6 equivalent to model.xyz -g 0 -level 1 -mol CH3OH -r 6 Takes a grain model from the *model.xyz* file. Performs fixed, unfixed, and vibrational frequencies computation at the GFN1-xTB level with methanol while using a grid level of 42 vertices and a radius of 6 Å for the unfixed part.

**Example 5** model.xyz --gfn 0 --level 1 --molecule CH3OH --radius 6 --no_fixed --restart 20 equivalent to model.xyz -g 0 -level 1 -mol CH3OH -r 6 -nf -restart 20 Same as the previous example, if the previous example computation had failed at the 20th vertex during the unfixed computation. This way, the fixed part is not computed again, and the program continues by starting at the 20th vertex of the unfixed computation. When the unfixed part is finished, the vibrational frequencies are computed.

**Example 6** grain_other.xyz --level 1 --only_fixed equivalent to grain_other.xyz -level 1 -of: Takes the grain model from the *grain_other.xyz* file, and computes only the fixed geometry optimization at the GFN1-xTB level with a grid of 42 vertices and water molecules.

**Example 7** grain_other.xyz --level 1 --only_unfixed equivalent to grain_other.xyz -level 1 -ouf: Continues from the previous example by taking the

fixed computation and performing the unfixed geometry optimization from it.

**Example 8** grain_other.xyz --level 1 --only_freq --range 0 21 equivalent to grain_other.xyz -level 1 -ofreq -range 0 21: Continues the previous example by taking the fixed and unfixed geometry optimization, and performing the vibrational frequencies computation. But, instead of computing the vibrational frequencies of the 42 BE, it only computes them for half of the BE (from 1st to the 22nd). This allows to divide the computation into different parts if someone wants to compute them in parallel on different CPUs. For computing the rest of the vibrational frequencies, the input line would be grain_other.xyz --level 1 --only_freq --range 22 41 equivalent to grain_other.xyz -level 1 -ofreq -range 22 41.

**Example 9** model.xyz --gfn 1 --level 2 --molecule NH3 --radius 10.5 --range 0 20 --rotation 2 equivalent to model.xyz -g 1 -level 2 -mol NH3 -r 10.5 -range 0 20 -rotation 2: Takes a grain model from a *model.xyz* file, computes the fixed and unfixed geometry optimizations, followed by the vibrational frequencies of a 162 vertices grid (third level). The molecules adsorbed are ammonia molecules, and the radius for the unfixed computations is set at 10.5 Å. For each vertex, two different random orientation of ammonia are used, meaning that only computations linked to the first 10 vertices will be done due to the *range* keyword.

**Example 10** model.xyz --gfn 0 --level 1 --molecule CH3OH --radius 6 --add_rotation 1 3 equivalent to model.xyz -g 0 -level 1 -mol CH3OH -r 6 -add_rotation 1 3: Takes an already done fixed, unfixed, and vibrational frequency computation using a grain model contained in a *model.xyz* file and a second level grid of 42 vertices. The computations are done for methanol with a radius of 6 Å for the unfixed part. This previous computation is continued by adding 2 random rotation to the previous sole rotation, meaning that 84 additional BE will be computed.

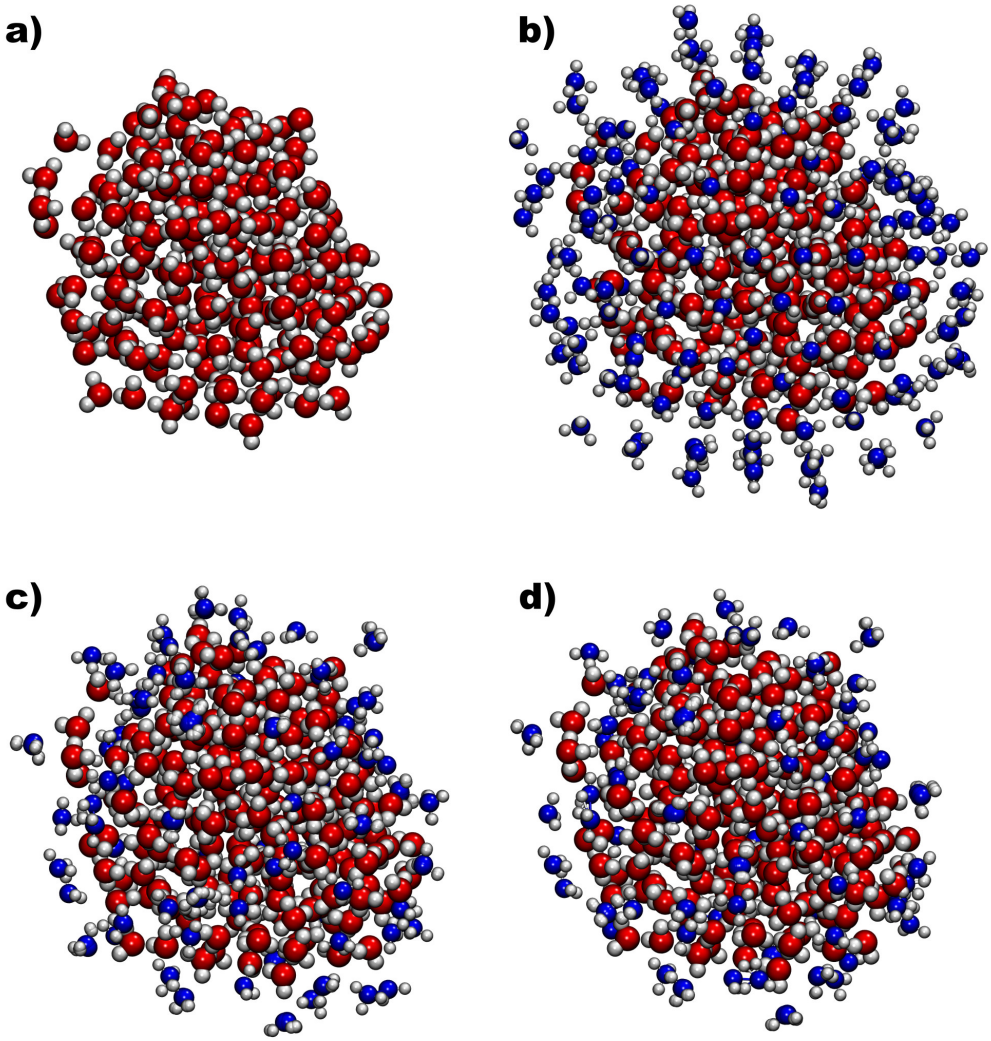



**Figure 4.12:** Illustration of example 1: a) The bare grain contained in *model.xyz*; b) Every input positions corresponding to the 162 grid points. For this example the $NH_3$ molecules are all showed at the same time, but in reality the program process the computation one $NH_3$ at a time; c) Every $NH_3$ positions after the *fixed* computation was performed; d) Every $NH_3$ positions after the *unfixed* computation was performed. As every of the 162 $NH_3$ positions are showed at the same time on this figure, we can notice that some input positions fall on the same final adsorption point.

# Appendix

## Grain building: List of keywords

Keywords for an input line are indicated in the terminal after the name of the ACO-FROST code file python3 ACO-FROST.py --keyword1 value1 --keyword2 value2 --keyword3 value3. They can either be called through their complete name like follows --keyword, or through a contracted name like this -contracted_keyword. Some keywords are mandatory, some are not, and some are incompatible between each other. This part will present all the keyword available in the ACO-FROST grain building code, the values they take, and the possible incompatibilities.

**--gfn *x* or -g *x*** Selects the GFN-xTB method to be used for geometry optimization during the growing of the grain. The parameter $x$ can take the values: *0* for GFN0-xTB, *1* for GFN1-xTB, *2* for GFN2-xTB, and *ff* for GFN-FF. If the keyword is not indicated in the input line, the default value *2* is selected.

**--size *i* or -size *i*** Indicates the final size of the desired grain. The value $i$ is necessarily an integer. If the keyword is not indicated, the default value for $i$ selected by the program is *0*, terminating immediately the program.

**--structure *x* or -structure *x*** Takes the path and name, with extension, of an existing grain structure via the $x$ value if the user wants to continue a grain growing process from an existing structure. If the keyword is not indicated, the grain growing process starts from the beginning with a one molecule seed determined by the **--molecule** keyword or the input file selected via **--input**, *vide infra* for both keywords. This keyword is useful to restart previous run.

**--fixed_structure or -fixed_structure** If a pre-existing structure is indicated via the **--structure** keyword, its geometry can be fixed using the *constrain* keyword of GFN-xTB, which applies a force constant of exactly 0.05 hartree.bohr$^{-2}$ to the distance among the selected atoms. All molecules subsequently added will not be fixed. If **--fixed_structure** is not indicated,

the pre-existing structure will also be relaxed during the geometry optimization of the grain growing process. If **--fixed_structure** is indicated without **--structure** it will proceed as a normal run.

**--mixed *x i* or -mix *x i*** If the user wants to produce a grain model using a mixed high-level/low-level method. The *x* parameter indicates the selected high-level method among the available GFNn-xTB ones (see keyword **--gfn** for the values authorized). The low-level method is taken from the **--gfn** keyword. The parameter *i* indicates the number of steps needed for performing the high-level geometry optimization. For example, if *100* is selected for *i* and *2* is selected for *x*, then a GFN2-xTB geometry optimization will be performed every 100 molecules added. If the low-level method indicated via the **--gfn** keyword is GFN-FF, an automatic high-level geometry optimization will be performed, for convention, at the 50th added molecule. This keyword has no default value.

**--input *x* or -inp *x*** Indicates if an input file, as showed in Section 4.3.2, needs to be read by the program. The path and name of the input file are indicated via the *x* parameter. This keyword has no default value.

**--molecule *x* or -mol *x*** If the growing grain needs to be made of pure ice, this keyword, through the *x* parameter, indicates the formula of the molecule from which the grain model will be made of. The molecules available are the same as presented in Section 4.3.2. If no input file is indicated via the **--input** keyword, and the **--molecule** is also not indicated, the default molecule chosen will be a water molecule (equivalent to *x* equal *H2O*).

**--restart *i* or -r *i*** Restarts the program from the *i* step; useful when the program did not finish correctly or stopped for any reason. This keyword has no default value.

**--MD *x i* or -md *x i*** To introduce a molecular dynamics simulation using one of the GFNn-xTB method via the *x* parameter (*0*, *1*, *2*, or *ff*), and at a certain number of growing steps via the *i* parameter. The MD computation will be done at 10 K and for 1 ps, to change these values see Section 4.3.2.1 below and the **--input** keyword. Example: **--MD ff 20** perform a MD

computation at the GFN-FF level every 20 added molecules. This keyword has no default value.

**--random_law *x* or -rand *x*** Indicates which random rule to follow for random positioning of molecules. The $x$ parameter can take the values: *normal*, *sbiased*, *plane*, *hollow*, or *spherical*. Each rule is explained in Section 4.3.2. If the keyword is not indicated, $x$ takes by default the value *normal*.

**--optimisation_cycle *i* or -opt_cycle *i*** To selected after how many added water molecules the geometry optimization is carried out. For example, by indicating --optimization_cycle 20, a geometry optimization will only be performed every 20 added molecules. If the keyword is not indicated, a geometry optimization is performed every added molecule, corresponding to a default value $i$ of *1*.

**--check_surface or -check_surface** It performs a "sanity check", as explained in Section 4.3.1, after the geometry optimization of the last added molecule. This keyword takes no value.

**--final_gfn2 or -final_gfn2** It performs a GFN2-xTB optimization as a very last step of the grain growing process. This keyword takes no value.

**--agermain2022 or -agermain2022** Produce a grain model using the method describe in Germain et al. 2022.[19] Equivalent to --gfn 2 --mixed 2 100 --molecule H2O --MD ff 10 --final_gfn2 --check_surface. Useful to reproduce exactly the published grains. This keyword takes no value.

## Molecular adsorption: List of keywords

**grain_model_file** Mandatory parameter to put before any keywords. Indicates the path and name, with extension, of the grain model to be used. It is not preceded by a keyword and should be written after the python script name in the terminal. While the file can be in any file format accepted by the ASE python package[11], the structures produced by ACO-FROST are always in the "xyz" format.

**--level *i j* or -l *i j*** Select the grid level to be used through $i$, starting from *0* for the first level grid. The parameter $j$ is optional and indicates if the grid should be also trisected (*1*) or not (*0*).

**--molecule *x* or -mol *x*** The formula $x$ of the molecule to be adsorbed on the grain model. If not indicated, $x$ takes the default value of *H2O* for water.

**--gfn *x* or -g *x*** Selects the GFN-xTB method to be used for geometry optimization during the binding energy computation. The parameter $x$ can take the values: *0* for GFN0-xTB, *1* for GFN1-xTB, *2* for GFN2-xTB, and *ff* for GFN-FF. If this keyword is not indicated in the input line, the default value *2* is selected.

**--radius *f* or -r *f*** Length $f$ of the radius in which the molecules to be unfixed will be selected. Takes the default value *5.0* if not indicated.

**--othermethod or -om** Replaces the default method of BE computation by a straightforward application of the BE definition. The grain is not fixed, and is fully optimized with the adsorbates.

**--range *i j* or -range *i j*** Indicates, through the integers $i$ and $j$, the range of structure to be computed. For example, for a third level grid containing 162 different grid points, --range 43 57 will only compute the structures of grid points 43 to 57.

**--rotation *i* or -rotation *i*** The number of random rotation $i$ per grid points wanted. Takes the default value *1* if not indicated.

**--grid_continue *i j* or -gc *i j*** After a successful BE computation, increase the original grid level selected $i$ to the newly wanted one $j$ without computing again the structures of grid level $i$.

**--add_rotation *i j* or -add_rotation *i j*** After a successful BE computation, increase the number of random rotations per grid points from the original number $i$ to the newly wanted one $j$.

**--restart *i* or -restart *i*** Restarts the program from a certain grid point $i$.

**--only_fixed or -of** Only computes the *fixed* part of the BE computation scheme. Incompatible with keywords **--only_unfixed**, **--only_freq**, and **--no_fixed**.

**--only_unfixed or -ouf** Only computes the *unfixed* part of the BE computation scheme. Incompatible with keywords **--only_fixed**, **--only_freq**, and **--no_unfixed**.

**--only_freq or -ofreq** Only computes the *vibrational frequency* part of the BE scheme. Incompatible with keywords **--only_fixed**, **--only_unfixed**, and **--no_freq**.

**--no_fixed or -nf** Do not compute the *fixed* part of the BE computation scheme. To be used as a continuation of a previous BE computation where the *fixed* part was already computed. Incompatible with keyword **--only_fixed**.

**--no_unfixed or -nu** Do not compute the *unfixed* part of the BE computation scheme. Incompatible with keyword **--only_unfixed**.

**--no_freq or -nfreq** Do not compute the *vibrational frequency* part of the BE computation scheme. Incompatible with keyword **--only_freq**.

# References


[1] L. Song and J. Kästner, "Formation of the prebiotic molecule $NH_2CHO$ on astronomical amorphous solid water surfaces: accurate tunneling rate calculations," *Physical Chemistry Chemical Physics*, vol. 18, pp. 29278–29285, oct 2016.

[2] T. Lamberts and J. Kästner, "Influence of Surface and Bulk Water Ice on the Reactivity of a Water-forming Reaction," *The Astrophysical Journal*, vol. 846, p. 43, aug 2017.

[3] G. Molpeceres, V. Zaverkin, and J. Kästner, "Neural-network assisted study of nitrogen atom dynamics on amorphous solid water – I. adsorption and desorption," *Monthly Notices of the Royal Astronomical Society*, vol. 499, pp. 1373–1384, oct 2020.

[4] G. Molpeceres and J. Kästner, "Adsorption of $H_2$ on amorphous solid water studied with molecular dynamics simulations," *Physical Chemistry Chemical Physics*, vol. 22, pp. 7552–7563, apr 2020.

[5] A. Al-Halabi, E. F. van Dishoeck, and G. J. Kroes, "Sticking of CO to crystalline and amorphous ice surfaces," *The Journal of Chemical Physics*, vol. 120, pp. 3358–3367, feb 2004.

[6] S. Andersson, A. Al-Halabi, G.-J. Kroes, and E. F. van Dishoeck, "Molecular-dynamics study of photodissociation of water in crystalline and amorphous ices," *The Journal of Chemical Physics*, vol. 124, p. 064715, feb 2006.

[7] W. M. C. Sameera, B. Senevirathne, S. Andersson, M. Al-lbadi, H. Hidaka, A. Kouchi, G. Nyman, and N. Watanabe, "$CH_3$ O Radical Binding on Hexagonal Water Ice and Amorphous Solid Water," *The Journal of Physical Chemistry A*, vol. 125, pp. 387–393, jan 2021.

[8] K. Hiraoka, T. Miyagoshi, T. Takayama, K. Yamamoto, and Y. Kihara, " Gas-Grain Processes for the Formation of $CH_4$ and $H_2O$: Reactions of H Atoms with C, O, and CO in the Solid Phase at 12 K ," *The Astrophysical Journal*, vol. 498, no. 2, pp. 710–715, 1998.

[9] F. Dulieu, L. Amiaud, E. Congiu, J. H. Fillion, E. Matar, A. Momeni, V. Pirronello, and J. L. Lemaire, "Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms," *Astronomy and Astrophysics*, vol. 512, no. 5, pp. 1–5, 2010.

[10] Y. Oba, N. Watanabe, T. Hama, K. Kuwahata, H. Hidaka, and A. Kouchi, "Water formation through a quantum tunneling surface reaction, OH + $H_2$, at 10K," *Astrophysical Journal*, vol. 749, no. 1, 2012.

[11] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, "The atomic simulation environment—a Python library for working with atoms," *Journal of Physics: Condensed Matter*, vol. 29, p. 273002, jul 2017.

[12] P. Pracht, E. Caldeweyher, S. Ehlert, and S. Grimme, "A Robust Non-Self-Consistent Tight-Binding Quantum Chemistry Method for large Molecules," *ChemRxiv*, no. preprint, pp. 1–19, 2019.

[13] S. Grimme, C. Bannwarth, and P. Shushkov, "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1-86)," *Journal of Chemical Theory and Computation*, vol. 13, no. 5, pp. 1989–2009, 2017.

[14] C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions," *Journal of Chemical Theory and Computation*, vol. 15, no. 3, pp. 1652–1671, 2019.

[15] S. Spicher and S. Grimme, "Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems," *Angewandte Chemie - International Edition*, vol. 59, no. 36, pp. 15665–15673, 2020.

[16] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme, "Extended tight-binding quantum chemistry methods," *WIREs Computational Molecular Science*, vol. 11, p. e1493, mar 2021.

[17] S. Pantaleone, J. Enrique-Romero, C. Ceccarelli, P. Ugliengo, N. Balucani, and A. Rimola, "Chemical Desorption versus Energy Dissipation: Insights from Ab Initio Molecular Dynamics of HCO· Formation," *The Astrophysical Journal*, vol. 897, p. 56, jul 2020.

[18] S. Pantaleone, J. Enrique-Romero, C. Ceccarelli, S. Ferrero, N. Balucani, A. Rimola, and P. Ugliengo, "$H_2$ Formation on Interstellar Grains and the Fate of Reaction Energy," *The Astrophysical Journal*, vol. 917, p. 49, aug 2021.

[19] A. Germain, L. Tinacci, S. Pantaleone, C. Ceccarelli, and P. Ugliengo, "Computer Generated Realistic Interstellar Icy Grain Models: Physicochemical Properties and Interaction with $NH_3$," *ACS Earth and Space Chemistry*, vol. 6, pp. 1286–1298, may 2022.

[20] N. A. Teanby, "An icosahedron-based method for even binning of globally distributed remote sensing data," *Computers and Geosciences*, vol. 32, no. 9, pp. 1442–1450, 2006.

# Chapter 5

# Applications

# Abstract

In this chapter we will present several results obtained with the ACO-FROST set of codes. First we will introduce a large cluster model of 1000 water molecules (LC) formed using the methodology presented in the previous chapter. This grain will be extensively characterized for several properties such as H-bond length and presence of dangling species. Then we will study and characterized small cluster models of 200 water molecules (SC). After that we will present the adsorption of several molecules on a SC, namely $NH_3$, water, and $CH_3OH$, and their resulting zero-point energy corrected binding energy distribution. For ammonia and water we will also present a refinement to the GFN2 BE distribution using a ONIOM scheme, and use the refined water BEs to illustrate the importance of BE distribution by introducing them in a simplified astronomical model of gaseous water in protoplanetary disks. For methanol, we will compute the GFN2 infrared (IR) spectrum of the adsorbed molecules using ORCA, and compare our results with observed IR features.


## 5.1 Introduction

Using the methodology presented in the previous chapter through the ACO-FROST program, two amorphous water ice grain models of different sizes have been build. The largest one (1000 water molecules, hereafters named LC), has been extensively characterized for structural features and for molecular adsorption studies. The smallest one (200 water molecules, hereafter named SC), has been used to compute the binding energies distributions of three interstellar relevant molecules ($NH_3$, $H_2O$, $CH_3OH$) sampling the whole set of adsorption sites.

In this chapter we will present the main results for LC and SC models, either free and when interacting with the reported adsorbates.

## 5.2 Grain models

### 5.2.1 LC model

#### 5.2.1.1 Methodology

Using the ACO-FROST grain building code we built up the LC optimized at the GFN2-xTB. The adopted methodology has already been described in Chapter 4 of this thesis and only quickly summarized here.

The grain is built by aggregating 1000 $H_2O$ molecules randomly, in a spherical symmetry, around an initial seed made of a single water molecule. After each water addition, a geometry optimization at the GFN-FF level of theory is performed using the default settings of the xTB code. From the first molecule added to the last, several intermediate actions are performed at given steps of the process. Every 10 added molecules, a NVT MD run is performed at 10 K for 1 ps, with a time step of 1 fs and using GFN-FF, to simulate the ice grain relaxation due to the formation of water which is not explicitly simulated in ACO-FROST (see Chapter 4 for details). We counted, on average, an increase of 6 H-bonds after the MD steps. Additional intermediate steps are taken during the grain building process in the form of GFN2-xTB geometry optimizations each 100th added water molecules. GFN2-xTB, the most accurate method of the xTB program (described in Chapter 2), is a semi-empirical quantum mechanical method, therefore more accurate than the GFN-FF force field. The adopted GFN-FF/GFN2 strategy, dramatically reduces the computing time when compared to a full GFN2-xTB treatment. As showed in Table 5.1, for the same number of molecules, GFN-FF is at least two orders of magnitude faster than GFN2-xTB, with a global speedup of almost 50 for a geometry optimization of a 800 water grain. Unfortunately, GFN-FF tends to form exceedingly open structures, with a significant number of external water molecules showing incomplete H-bond coordination, imparting some fluffiness to the grain (see Fig. 5.1). The GFN2-xTB optimization step indeed recovers the correct compactness of the H-bond network. Interestingly, a next GFN-FF optimization on the GFN2 optimized structure, does not break the correct H-bonding network, but for a small global expansion of the grain size due to the slightly longer H-bond distance at GFN-FF compared to GFN2. When building up the LC cluster, a check of the finally surface is carried out to remove possible dangling water. A flowchart of the process is showed in Section 4.3.

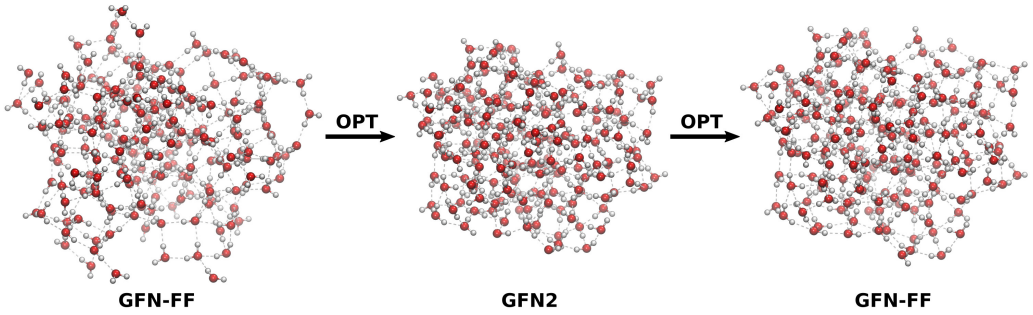



**Figure 5.1:** Successive GFN-FF and GFN2 geometry optimization of an icy grain cluster. Left: a GFN-FF optimized grain; center: the GFN2 optimization shrinks the GFN-FF structure; right: GFN-FF re-optimized GFN2 structure only brings a small global increase of the grain radius while keeping the GFN2 H-bond network. Each image have the same scale. Taken from Germain et al. 2022[1].

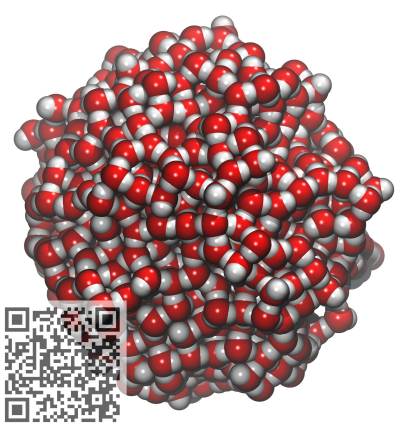
**Table 5.1:** Comparison between the GFN-FF and GFN2 geometry optimization times for different grain sizes. Taken from Germain et al. 2022[1].

| Number of water molecules | GFN2 | GFN-FF | ratio GFN2/GFN-FF |
|---|---|---|---|
| 200 | 34 min | 1 min 26 s | 24 |
| 400 | 5 h 15 min | 7 min | 44 |
| 800 | 1 day 17 h | 53 min | 46 |

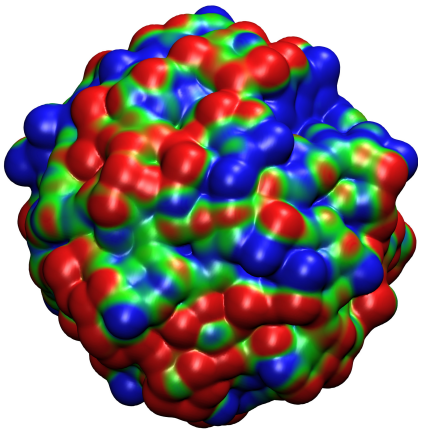
Note – The xTB program was run on a 20 core Intel Xeon E5-2630 v4 @ 2.20 GHz.

#### 5.2.1.2 Hydrogen-bonds features

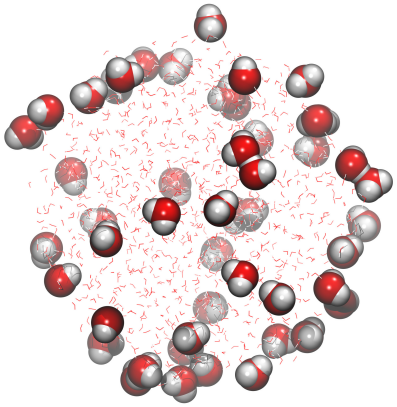
The characterization of the LC grain (shown in Fig. 5.2-a) is exclusively performed for the GFN2-xTB structures, considering intermediate stages of the grain building up (100, 400, 700 and 1000 water icy grains). For each structure, we extracted different properties and plotted them as a function of the grain size, namely: (i) the H-bond length distribution (Fig. 5.3); (ii) the number of H-bonds per water molecule (Fig. 5.4-Left); (iii) the ratio of the number of surface dangling oxygens (dO) over the number of surface dangling hydrogens (dH) (Fig. 5.4-Right). The Hydrogen bonds were sampled by computing the distances between oxygens and hydrogens atoms (O$\cdots$H) on the GFN2-xTB optimized structures. A threshold of 1.2/2.2 Å was added to filter covalent bonds and O$\cdots$H distances that do not represent H-bonds.



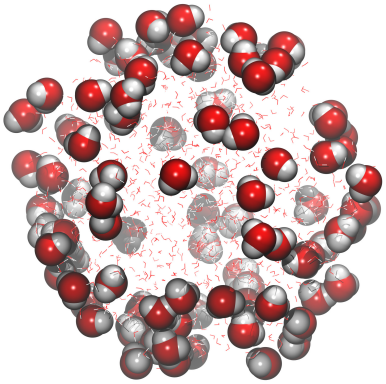
**(a)** LC icy grain.



**(b)** PBEsol electostatic potential surface (ESP) map of the LC grain.



**(c)** LC grain highlighting the surface dH molecules.



**(d)** LC grain highlighting the surface dO molecules.

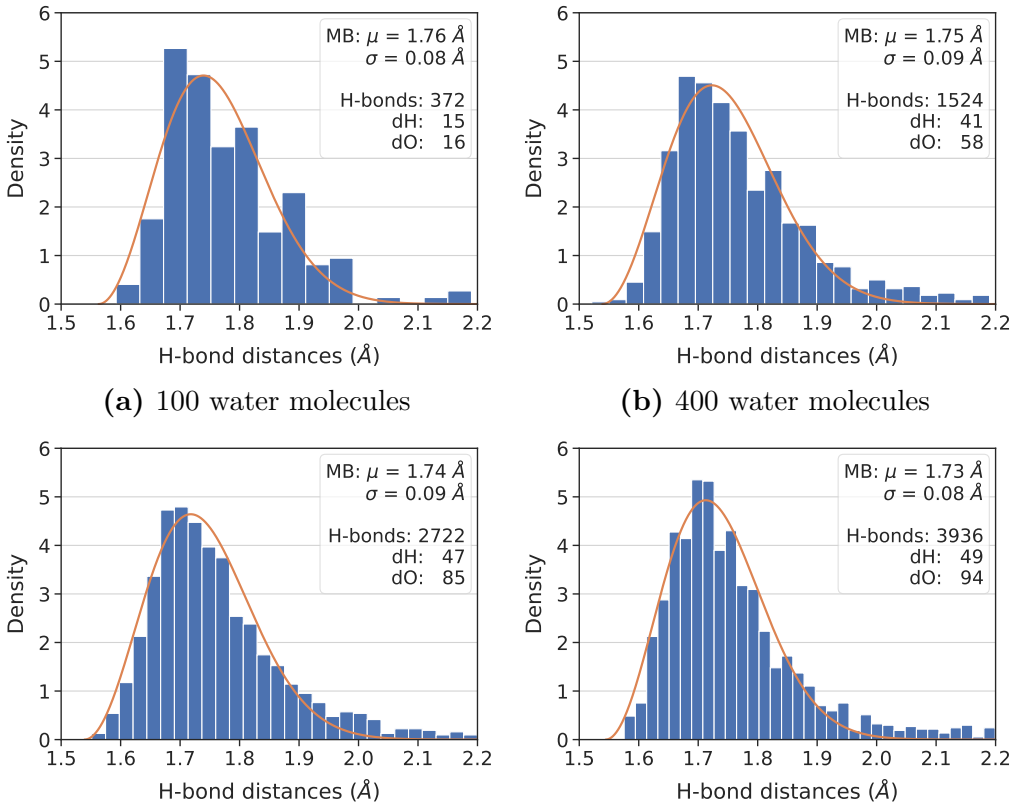
**Figure 5.2:** Different representation of the LC grain highlighting: a) the vdW surface sites (the JSmol interactive grain is accessible by scanning the QR code); b) the electrostatic potential map (blue(positive)/red(negative zones); c) and d) the dH and dO surface sites. Taken and adapted from Germain et al. 2022[1].

The dO/dH water molecules were defined as those in which oxygen/hydrogen atoms do not participate in any H-bonds. ACO-FROST implements a filter reducing the dO/dH counting errors due to possible minor cavities inside the grain structure exhibiting water molecules with incomplete H-bond coordination. It filters out dO and dH present at an arbitrary-selected minimum distance from the geometrical center of the grain. The dangling species of the LC grain are shown in Fig. 5.2-c for dH and Fig. 5.2-d for dO.

The H-bond length distribution showed in Fig. 5.3 is fitted quite nicely by a Maxwell-Boltzmann (MB) probability density function. For the four grain models, the features indicated (mean and standard deviation of the MB distribution) are very similar, with the average H-bond lengths between 1.73 Å for the LC grain, and 1.76 Å for the smallest 100 water cluster. Fig. 5.4-Left shows the average number of H-bonds per water molecules for the considered icy grains. The average values varies from 3.72 (for the 100 water cluster grain) to 3.94 (for the LC grain). Considering that a water bulk involves 4 H-bonds, the slightly reduced number for the LC grain (Fig. 5.4-Left) derives from the dO/dH surface water molecules, relevant for the adsorption process. Molecules behaving as H-bond acceptors, like $NH_3$, will be adsorbed at the grain via H-bond with the dH grain sites, while H-bond donor, like $CH_3OH$, will be adsorbed at the dO sites. Obviously, for the latter cases, also dH sites play a role, engaging H-bond with the oxygen atom of the adsorbate.

The EPS map, computed using the parallel version of CRYSTAL17[2] at the PBEsol level[3], shows in blue/red colors the dH/dO zones. Inspection of the EPS map immediately reveals a non uniform distribution of red/blue zones, with the latter being in a smaller number than the former, in agreement with Fig. 5.2-c/d). shows the number of dangling species per steps size, and exhibit an increase of dO and dH with the sizes, an expected result as the number of dangling species depends on the size of the grain's surface and grows with the number of molecules. The ratio of dO/dH reaches 1.9 at infinite cluster size (see Fig. 5.4-Right), meaning that the number of dO is almost twice the number of dH, reflected in the EPS map where we estimated roughly the surface to be occupied by negative zones (red-dO) at 46%, by positive (blue-dH) by 38%, and by almost neutral zone (green) by the remaining 13%. We can conclude from this unbalanced dO/dH ratio that the grain preferentially adsorb H-bond donors molecules than H-bond receiver molecules, impacting its adsorbing properties consequently.

We computed the density $\rho$ of each intermediate structure using a Monte Carlo integration scheme, with 30000 sampling points, and by assigning van




**(a)** 100 water molecules **(b)** 400 water molecules

**(c)** 700 water molecules **(d)** 1000 water molecules

**Figure 5.3:** H-bond length (O$\cdots$H) distribution of four icy grains of different size. Top right inset to each plot: average H-bond length $\mu$ and its standard deviation $\sigma$ of the Maxwell-Boltzmann distribution fit; sum of the number of H-bonds given and received for each $H_2O$ molecule; Number of dH (dangling hydrogen) and dO (dangling oxygen) sites. Taken from Germain et al. 2022[1].

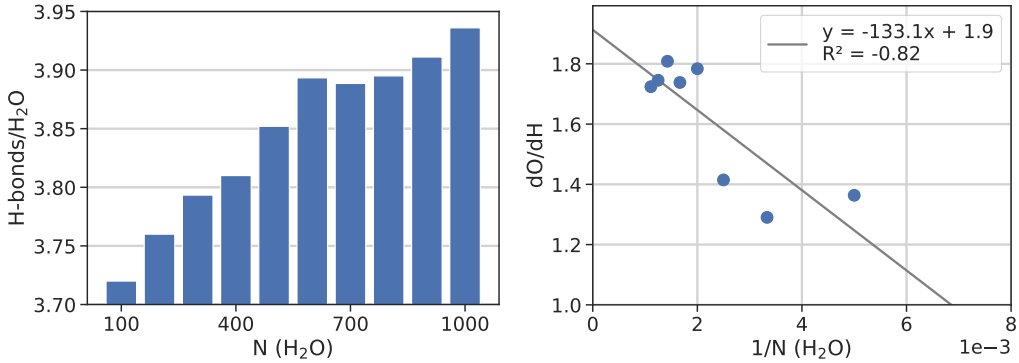



**Figure 5.4:** Number of H-bonds received and given per water molecule for different grain sizes (left). Ratio of dangling oxygens (dO) and hydrogens (dH) for the different grain sizes (right). Taken from Germain et al. 2022[1].

der Waals radii of 1.30 Å and 2.05 Å to H and O atoms respectively to compute the grain volume as encoded in the MOLDRAW program[4]. The radii were chosen so that the density computed for a structure of crystalline ice $I_h$ resulted in the same value as the experimental[5] one of $\rho = 0.92\ gr/cm^3$. The density $\rho$, in $gr/cm^3$ and with an error of $\pm 0.03\ gr/cm^3$, was computed to be 1.12 for 200 water molecules, 1.21 for 500, and 1.25 for 900. The density computed for the different size grain is superior to the experimental one of crystalline ice, in agreement with the amorphous, liquidlike, character of the simulated grains. However, as GFN2-xTB does not support periodic boundaries conditions, the absence of a GFN2-xTB periodic computation leaves some uncertainty in the absolute density values, due to expected systemic errors of GFN2.

#### 5.2.1.3 Shape characterization

To characterize the shape of the generated icy grains we computed the gyration tensor as described in the paper from Theodorou and Suter[6]. For a three dimensions Cartesian coordinate system, it reads:

$$\mathbf{S} = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \tag{5.1}$$

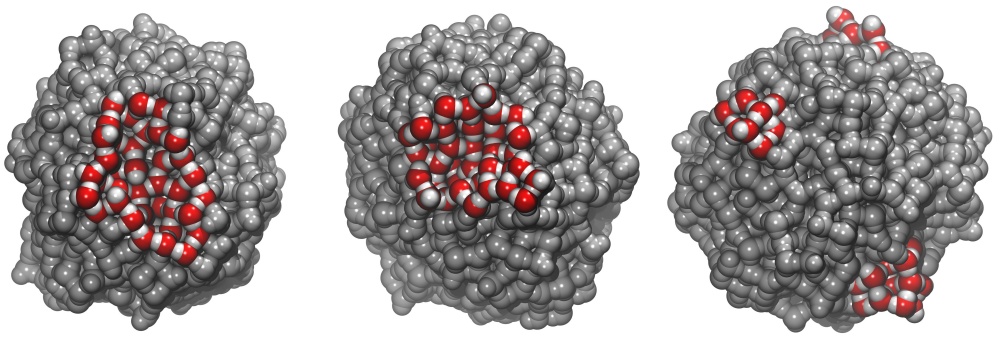


**Figure 5.5:** Left and center: Two cavities. Right: several protruding spots. Taken from Germain et al. 2022[1].

where each component $S_{mn}$ of the $S$ matrix are defined as:

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} m^{(i)} n^{(i)}, \qquad m, n = x, y, z \tag{5.2}$$

with $N$, the number of atom in the system, and $m^{(i)}$ the Cartesian coordinate $m = x, y, z$ of the $i^{th}$ atom.

By diagonalizing the **S** matrix we obtain the eingenvalues $\lambda_x^2$, $\lambda_y^2$, and $\lambda_z^2$, ordered as $\lambda_x^2 \leq \lambda_y^2 \leq \lambda_z^2$. From them, we derived the gyration radius $R$, the asphericity $b$, and the relative shape anisotropy $\kappa^2$ defined as:

$$R = \sqrt{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}\,; \quad b = \lambda_z^2 - \frac{1}{2}\left(\lambda_x^2 + \lambda_y^2\right)\,; \quad \kappa^2 = \frac{3}{2}\frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{\left(\lambda_x^2 + \lambda_y^2 + \lambda_z^2\right)^2} - \frac{1}{2} \tag{5.3}$$

The asphericity $b$ is greater than zero, becoming exactly 0 only with a spherically symmetrical system, which is equivalent to $\lambda_x = \lambda_y = \lambda_z$. The relative shape anisotropy $\kappa^2$ is bounded between 0, all points are spherically symmetric, and 1, all points lie on a line. Thus, for a linear molecule such as $CO_2$, $\kappa^2 = 1$, for the vertices of a polygon $\kappa^2 = 1/4$, and for a tetrahedral, spherical, or higher symmetry structure $\kappa^2 = 0$.

Fig. 5.6 shows the computed quantities related to the gyration radius. The b plot shows a good linear correlation between the natural logarithm of the gyration radius R and the size of the GFN2-xTB steps of the grain growing process. The slope is close to 1/3 and can be justified by considering the volume of the grain associated with the gyration radius as

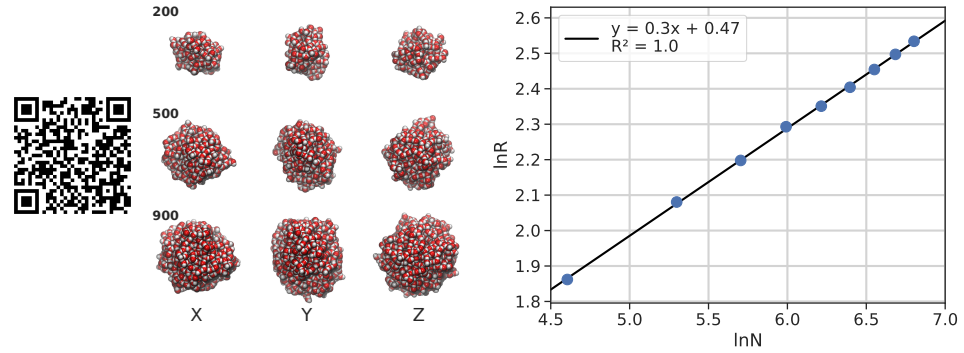
$V = (4/3)\,\pi R^3 \approx \nu_w N$, with $\nu_w$ being the gyration radius of a single water molecule. By taking the natural logarithm of both sides we obtain the dependence of $(1/3)\ln(N)$ which is close to the best fit value. From the shortening of the distance between each points, we can gather that the increase of the radius slows down, *i.e*, a certain number of molecules added at a size $N_1$ will not have the same impact as the same number of molecule added at a size $N_2$ if $N_2 > N_1$. For doubling the radius of our 1000 water molecule model from 13 to 26 Å an increase of 1 order of magnitude in the number of molecule would be necessary, implying a steep increase of computational resources needed to simulate a grain of realistic size. The asphericity b and relative shape anisotropy $\kappa^2$ showed in plots c and d of Fig. 5.6 exhibit a rapid convergence toward small values close to 0 due to the increased spherical shape of the grain shown in Fig. 5.6-a when increasing the grain's size.

Figure 5.5 highlights several amorphous structures of the grain's surface. The variety of these amorphous structures is an important feature for molecular adsorption if we aim to sample a vast area for binding energy.

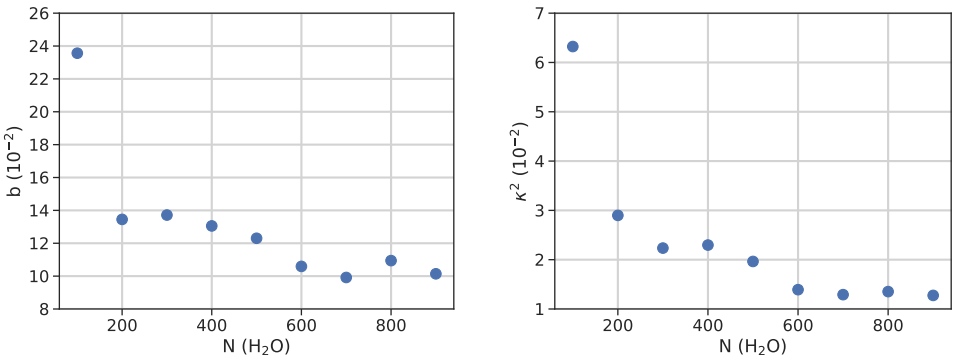
## 5.2.2 SC grain model

### 5.2.2.1 H-bonds characterization

The binding energy distribution of $NH_3$, $H_2O$, and $CH_3OH$ (*vide infra*) were computed on the SC model built up with the same methodology as the LC model. The size envisaging 200 water molecules was selected as it offered a good trade-off between the computational power demanded and the number of adsorption sites suggested by its properties (number of dO and dH, or number of H-bonds), making it large enough to have a significant statistical distribution of BE. Using the LC grain model would be a better option and Fig. 5.4 shows, for the SC, an underestimated dO/dH ratio and H-bond coordination compared to sizes superior to 500 water molecules. Nevertheless, we will show that, at least for the dO/dH ratio, different values do not impact significantly the BE distribution. Furthermore, the computational cost of sampling the LC grain model would be too high for too little rewards compared to the SC grain model. A van der Waals representation of the SC model and its associated ESP map are showed in Fig. 5.7. SC exhibits 752 H-bonds (3.76 per water molecule), and 30 dO, 24 dH with a dO/dH ratio of 1.25, considered a relatively balanced value for the sampling of H-bond




**(a)** Grain steps of different sizes **(b)** Logarithm of the gyration radius




**(c)** Asphericity b **(d)** Relative shape anisotropy $\kappa^2$

**Figure 5.6:** Different shape descriptors from the gyration tensor for different grain sizes (JSmol views can be seen by scanning the QR code). In (a) X, Y, and Z represent the view axis facing the screen. Taken from Germain et al. 2022.[1]

donor and acceptor species.

#### 5.2.2.2 Statistical study of H-bonds properties

As molecules are added randomly during the grain building process, different starting random seeds used to initiate the pseudo-random numbers will lead to different morphology of grains. To explore such a dependence we produced 20 models of the SC size with the same methodology as presented previously, and with different random seeds. Averaging over the 20 grain model for the resulting structural features we obtain 379 ± 4.4 H-bonds, 30.6 ± 2.8 dO, and

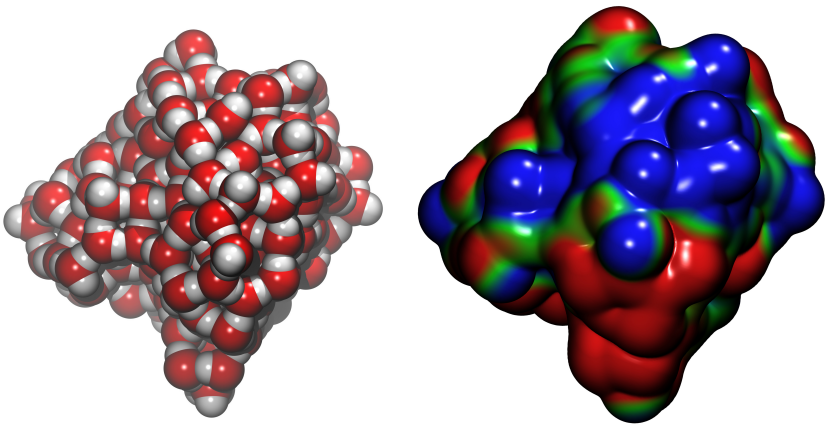


**Figure 5.7:** Left: Image of the SC grain model (200 water molecules) Right: Electrostatic potential surface map of the same grain. In blue H-bond donor (dangling Hydrogen) and red H-bond acceptor (dangling Oxygen). Taken from Germain et al. 2022[1].

21.4 ± 3.1 dH. Additionally, we computed the H-bond length distribution for the 20 grains and plotted the average in Fig. 5.8-Left. All these properties possess small standard deviation and very similar H-bond length distributions, indicating similar characteristic for grain models exhibiting the same size. These similarities indicate that BE distributions obtained on grains with the same size but different random seeds are similar.

We also built up 20 SC grains in which all geometry optimizations were performed with GFN2-xTB instead of the standard mixed GFN-FF/GFN2 relaxations. For all considered grains we obtained, on average, 381 ± 3.1 H-bonds, 29.2 ± 2.7 dO, and 19.4 ± 2.1 dO. These results are almost indistinguishable from the standard mixed GFN-FF/GFN2 method. Figure 5.8-Right shows the H-bond length distribution averaged over the 20 SC GFN2 grains. Comparing Fig. 5.8 Left and Right panels we can see that there is almost no difference between the two building methods (*i.e.* adopting a mixed strategy or a complete GFN2 one), giving credit to the adopted building strategy. Furthermore, adopting the mixed strategy allows a huge saving of computer resources as illustrated on Table 5.1.

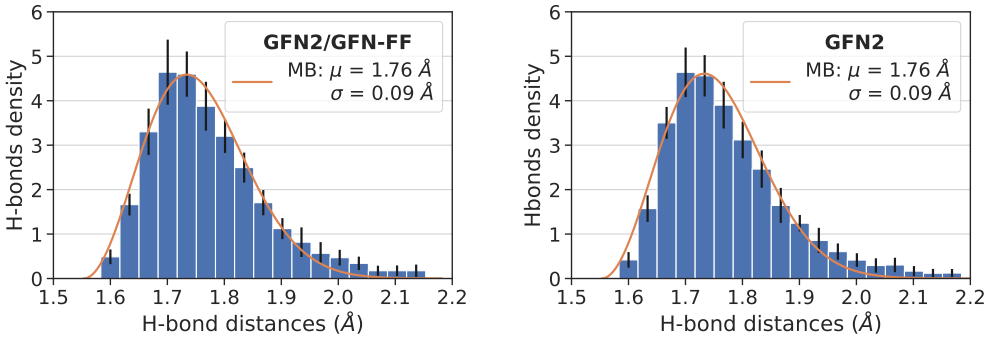



**Figure 5.8:** Average H-bond length for 20 water clusters of the SC size. The building process was carried out at both GFN2/GFN-FF (left) and at full GFN2 (right) levels. Black lines represent the standard deviation and orange line the Maxwell-Boltzmann fit. Taken from Germain et al. 2022[1].

## 5.3 Molecular adsorption of Ammonia

Ammonia, in its gaseous and attached to grains, is an omnipresent molecule in the ISM. It can be formed in the gas-phase of the ISM via molecular nitrogen,[7] and on the grain surface by hydrogenation of atomic nitrogen[8]. Via its binding energy it is possible to know if $NH_3$ is in the gaseous or adsorbed at the grain in the different part of the ISM. In very cold object ($\leq$ 10 K) like the L1544 prestellar core[9] (harboring a dust temperature of only 7 K[10] at the center of the condensation), ammonia should be completely frozen onto the icy grains mantles[11,12], however, it is still observed in its gaseous form[13]. Several reasons have been suggested, like the desorption of ammonia from the grain due to the chemical energy released during its formation. As the BE governs the energy needed for ammonia to be desorbed from the grain, a distribution of BE containing values spread over a range of both low and high BEs could elucidate the topic.

The binding energy distribution of $NH_3$ was computed using ACO-FROST methodology explained in Section 4.4.1 coupled with the SC model grain showed in Fig. 5.7 and described in section 5.2.2. Nevertheless, we only provide a short summary of the BE computation methodology.

### 5.3.1 Methodology: Short summary

A grid of 162 points was placed around the SC grain model (see Fig. 5.7) and each projected grid point hosts a molecule of ammonia randomly oriented around its geometrical center. For each of the 162 starting structures, three types of calculation are done: (i) a geometry optimization at the GFN2-xTB level of theory with the structure of the SC grain model kept fixed, while $NH_3$ is free to relax its position; (ii) a 5 Å sphere around the geometrical center of $NH_3$ defining a set of unfixed water molecules followed by a GFN2-xTB optimization while keeping fixed all the rest of the SC model; (iii) the water molecules contained in the sphere are recounted and if a variation occurred due to the optimization of the previous step we go back to step (ii) up to convergence; (iv) the harmonic vibrational frequencies of each of the 162 optimized structures was carried out to ensure all are minima. The last (iv) step also provides, as a by product, the ZPE to correct all the BEs.

#### 5.3.1.1 Refinement of the GFN2-xTB BE distribution using the ONIOM method

In collaboration with Lorenzo Tinacci, PhD student at the University of Grenoble and the University of Turin and involved in a different work package of the same ACO project, we refined the GFN2-xTB structures obtained for ammonia using a ONIOM[14] (DFT:GFN2-xTB) model chemistry available as an option in the Gaussian program[15] (v.16, Revision B.01). The model zone of each structure, treated with high-level methods, was chosen as the "unfixed" water molecules of the SC model inclusive of the adsorbed ammonia at its final GFN2-xTB optimized geometry; the real system, treated with the low-level GFN2-xTB, was the whole SC model. The xTB program (v.6.3.3) was used as an external program for the real system in the ONIOM scheme, as GFN2 is not yet implemented in Gaussian. The model system was optimized with the B97D3[16,17] functional, using the aug-cc-pVTZ basis set[18,19]. The real system was kept fixed during geometry optimization. After geometry optimization, the energy of the model were refined at the DLPNO-CCSD(T) level of theory[20], with aug-cc-pVTZ as the primary basis set, and aug-cc-pVTZ/C as the auxiliary basis set[21] for the resolution of the identity (RI) approximation in electron repulsion integrals. ORCA[22,23] (v.4.2.1) program was adopted for this part of the calculation. Further details of the ONIOM method were given in section 2.5.4 of this thesis.

### 5.3.2 Results and discussion

#### 5.3.2.1 Cleaning of the structures and BE(0) distribution

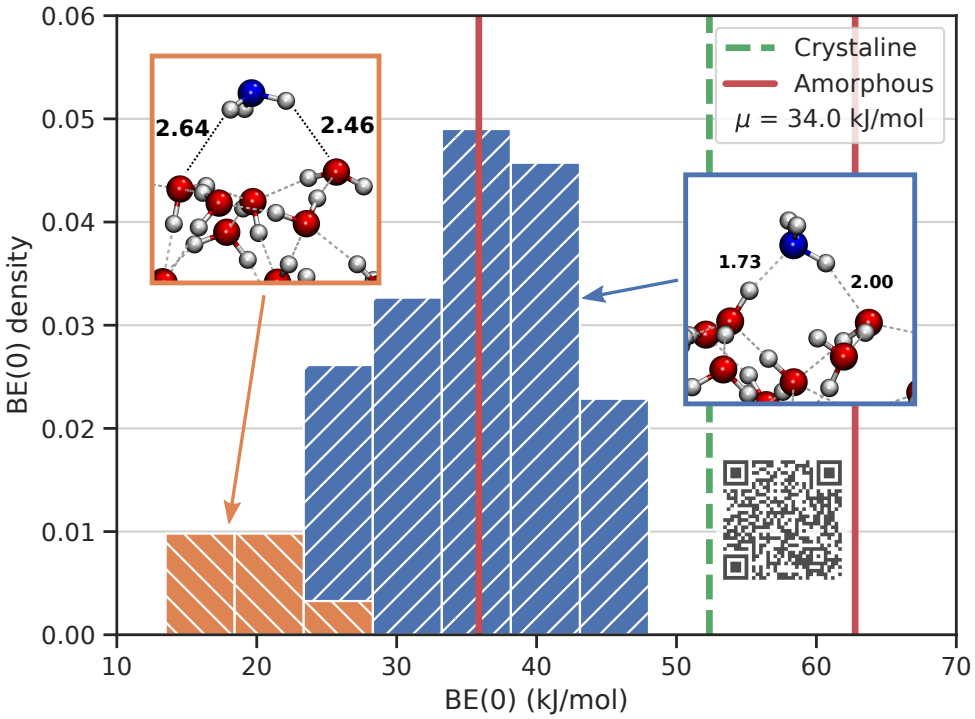
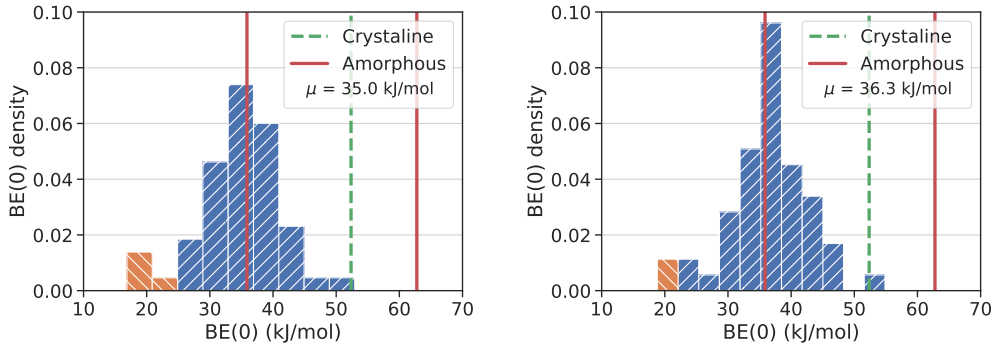



**Figure 5.9:** BE(0) distribution for $NH_3$ adsorption at the SC grain mode. Dotted green and continuous red lines are from Ferrero et al.[24] (JSmol representative structures can be seen by scanning the overlapping QR code). Taken from Germain et al. 2022[1].

On the 162 optimized structures three further quality checks were performed: (i) discarding any structure having at least one imaginary frequency; (ii) discarding any structure where the second step of the BE computation (see previous (iii) point) was needed more than once; (iii) purging final structures for possible duplicates. As for the (iii) case, it can happen that $NH_3$ starting at different grid points fall on the same adsorbing sites, giving rise to the same BE value and biasing the final BE distribution. After (i) and (ii) checks only 115 structures over the original 162 remained. Of these 115

structures, 31 land to a unique minimum, while the 84 left collapse to a common minimum. At the end of the cleaning process, only 62 unique structures (corresponding to 62 binding sites) remained to contribute to the BE distribution showed in Figure 5.9. In Fig. 5.9 we separated the ZPE corrected BE energies into two distributions collecting structures in which $NH_3$ behaves either as an H-bond acceptor (blue striped distribution) or as an H-bond donor (orange striped distribution). It is well known that $NH_3$ is a much stronger H-bond acceptor than H-bond donor. The most stable interactions are those maximizing the number of H-bonds. Indeed this is the case, and the higher BE are exhibited by $NH_3$ acting simultaneously as H-bond donor and acceptor. As expected, the least stable ones show $NH_3$ exclusively behaving as a H-bond donor. These structures are particularly interesting for potential chemical reactions occurring in the ISM, as the probability to diffuse and, therefore, to react with other species adsorbed on the grain's surface increases when the BE decreases.

Figure 5.9 also includes the BE values for $NH_3$ reported by Ferrero et al.[24] for both crystalline (green dashed line) and amorphous (red full lines) slab models of water ice. In the case of the amorphous slab model we report the minimum and maximum values of BE computed. The dO and dH distribution of crystalline ice is controlled by the imposed crystal symmetry and thus only one adsorption site, giving rise to one value only of BE, is found for the $NH_3$ case. Crystalline ice is found in significant abundance in protostellar molecular shocks[25] and protoplanetary disks[26] but the dominant form of ice found in the ISM is amorphous. Therefore, a crystalline model is not a realistic representation of an interstellar icy grain and cannot possess the variety of binding sites and adsorption properties found on amorphous ices. The periodic amorphous model developed by Ferrero et al.[24] is large enough to identify different sites, but still too small to allow the definition of a sensible BE distribution (we refer to chapter 3 for a more in depth description of the two water ice models used by Ferrero et al.[24]). BE reported by Ferrero et al.[24] on the amorphous model are highers than that coming from the SC model. We argue this to be due to two main reasons: first our cluster is more rigid due to the fixing of parts of the grain, leading to less deformation of the cluster during geometry optimization compared to the full relaxation of the amorphous slab; second, the adsorption sites for the amorphous slab where hand-selected and thus biased toward strong H-bond interactions, at variance with the unbiased procedure of the ACO-FROST scheme.




**(a)** GR17 grain **(b)** GR26 grain

**Figure 5.10:** $NH_3$ BE(0) distributions for GR17 and GR26 clusters grains. Taken from Germain et al. 2022.[1]

#### 5.3.2.2 Impact of the dangling species on the BE distribution

To study the impact of the number of dO and dH on the BE distribution we also computed the $NH_3$ distribution on two more grains among those generated with different random seeds in section 5.2.2.2. From these 20 grain models we selected two with the largest difference in the number of dH but a relatively close number of dO. The chosen grains are called GR17 and GR26 and possess 30 and 32 dO respectively, and 17 and 26 dH respectively. The distribution of $NH_3$ BEs for these two grains is shown in Figure 5.10. For both grains, we operated the same cleaning of BEs as explained before and found an identical number of 54 unique adsorption sites on their surface, allowing for a fair comparison. The grain GR26 possesses more dH than GR17 and thus we expected it to give higher BEs due to $NH_3$ being a better H-bond acceptor than H-bond donor. However this is not really the case, and the BE distributions for both grains are very similar. The average value of BE are equal to 35.0 kJ/mol for GR 17 and 36.3 kJ/mol for GR 26, a very slight increase compared to the difference in dH of almost 53%. Minima and maxima present also a slight increase for GR26, with a minimum of 16.9 kJ/mol and maximum of 52.9 kJ/mol for GR17, and a minimum of 18.8 kJ/mol and maximum of 54.9 kJ/mol for GR26. All values are almost independent of the grain model despite the modest increase in the BE for GR26, showing convergence of data for the final shape of the grain at a fixed size. A more careful analysis of the BE distribution reveals a difference in

the shape for high BE values with an over-representation of values around 35 kJ/mol for GR26 compared to GR17, while the tail at low BE values is rather similar, in agreement with the excess of dH of GR26 over GR17.

#### 5.3.2.3 Correlation between BE(0) and BE: circumventing the expensive computation of the Hessian matrix

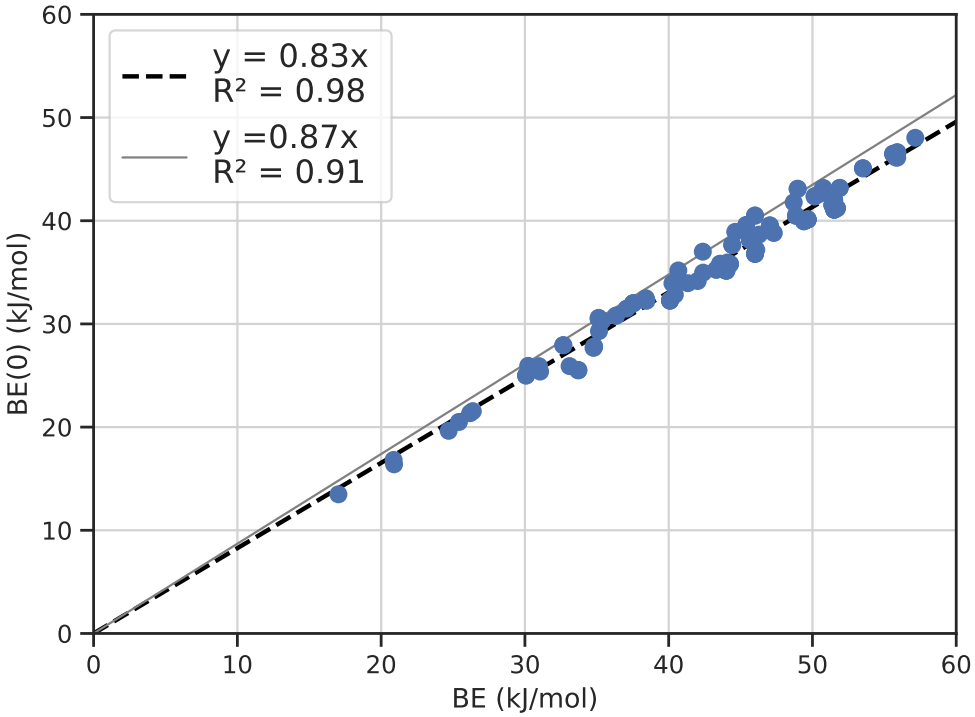
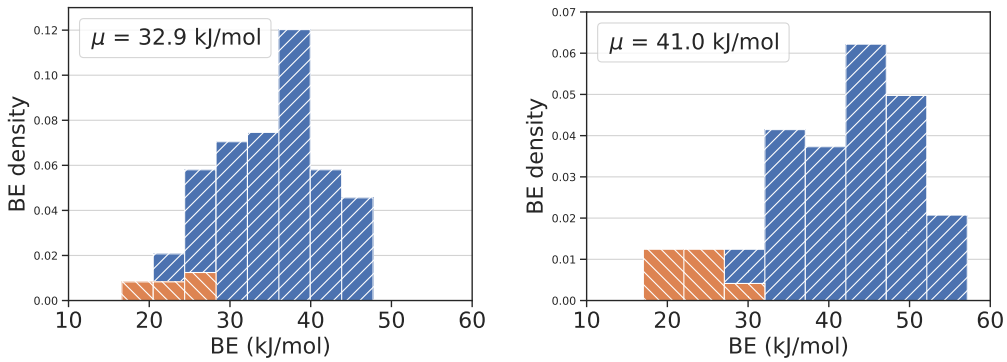



**Figure 5.11:** Linear correlation between BE and zero point energy corrected BE(0) values (kJ/mol) at GFN2 level from a previous work[1] (continuous line) and for all considered case of $NH_3$ adsorption on the SC grain model (best fitted dashed line on filled circle points). Taken from Germain et al. 2022[1].

As said in Chapter 3, the full set of harmonic frequencies is important to ensure that all structures are minima and to compute the corresponding ZPE. ZPE corrected BEs, BE(0), are very important as they can be compared with experimental data. However, calculation of the complete Hessian matrix, corresponding to the second derivative of the GFN2-xTB energy with respect to

the nuclear Cartesian coordinates, is very expensive in computational power and time. For example, using the same hardware reported on Table 5.1, the Hessian for the SC grain model (200 water molecules) takes less than 40 min with GFN2-xTB, while the time needed to compute the Hessian for 4×SC grain takes almost 6 days. Therefore, finding a way to circumvent this step, just to obtain accurate ZPE correction to the electronic BE would be a very useful endeavor. In our recent work[1,27], also presented in Chapter 3 of this thesis, we showed that the GFN2-xTB BE nicely correlated with the ZPE corrected ones BE(0) for the set of molecules adopted by Ferrero et al.[24] on a crystalline slab. Figure 5.11 shows the best linear correlation (continuous line) using our previous data, contrasted with the present GFN2-xTB BE and ZPE corrected BE (dashed line) for all considered $NH_3$ adsorption cases on the grain. Both linear fits are extremely similar and the old scaling factor could be used to obtain ZPE corrected BE without the need of the Hessian matrix. We will see that also for $H_2O$ this simple procedure holds true. In conclusion, it seems that one can avoid the full calculation of the Hessian matrix for the purpose of getting the ZPE correction to the BE. Instead, one can use the scaling factor derived in Ref.[1] to get a good estimate of the ZPE.

#### 5.3.2.4 Effect of partially unfixing the SC




**(a)** Fixed results **(b)** Partially unfixed results

**Figure 5.12:** $NH_3$ BE distributions for: (a) the SC grain when entirely fixed; (b) partially unfixed.

Fig. 5.12 plots the BE distribution computed when the SC cluster is kept

at fixed geometry leaving only $NH_3$ as mobile during the geometry optimization as well as when also the water molecules included in a 5 Å radius sphere around $NH_3$ are allowed to relax. The case in which water molecules closer to $NH_3$ are unfixed, shows a higher BE due to better accommodation of $NH_3$. BEs in which $NH_3$ acts as H-bond donors (weak binding) increased only slightly: the minimum value of the BE went from 16.6 kJ/mol to 17.0 kJ/-mol, the maximum went from 27.8 kJ/mol to 30.2 kJ/mol, and the average moves from 22.8 kJ/mol to 23.6 kJ/mol. As a whole, the distribution queue towards low BE values does not change significantly (a mere 0.8 kJ/mol increase). However, we see a broadening of the distribution in the high BE region: the minimum BE value moves from 23.4 kJ/mol to 26.2 kJ/mol, the maximum moves from 47.7 kJ/mol to 57.2 kJ/mol, and the average moves from 34.1 kJ/mol to 43.2 kJ/mol. While the lowest BE value increases by 2.8 kJ/mol when $NH_3$ behaves as H-bond acceptor (blue bars of Fig. 5.12), both the highest and the average BE values increase by 9.5 kJ/mol and 9.1 kJ/mol. Fig. 5.13 shows the subtle structural changes when passing from fixed to unfixed models, in which $NH_3$ undergoes an extra weak H-bond with a movable water molecule. When the water molecules included in a 5 Å radius sphere around $NH_3$ are allowed to relax, extra weak H-bonds can be formed, justifying the broadening of the BE distribution (see Fig. 5.12).

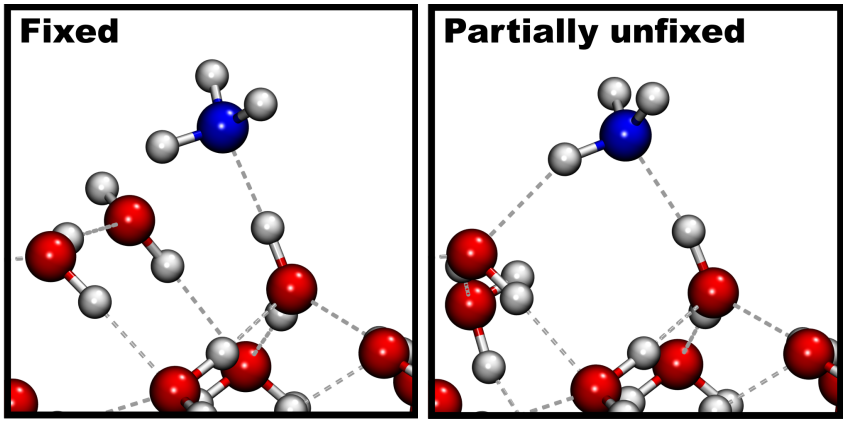



**Figure 5.13:** Left: Optimized structure keeping the underneath SC geometry fixed. Right: Optimized structure including relaxation of the water molecules contained in a 5 Å radius sphere around $NH_3$. Taken from Germain et al. 2022[1].

#### 5.3.2.5 ONIOM refinement: BE(0) distribution

To refine the GFN2-xTB ZPE corrected results showed in Fig. 5.9-Left we performed (as already stated in collaboration with Lorenzo Tinacci, PhD of the ACO project) a ONIOM calculation at the ONIOM(DLPNO-CCSD(T) //B97D3:GFN2-xTB) level. The refined BE(0) ONIOM values, published in Tinacci et al. 2022[28] are shown in Fig. 5.14. The data are more spread out toward the high BE(0) values, with binding energies ranging from 12.7 to 50.6 kJ/mol, probably due to the increase of unique structures (from 62 for GFN2-xTB to 77 for the ONIOM scheme). The average ONIOM BE(0) value is 31.1 kJ/mol, $\sim$ 3 kJ/mol less than the GFN2-xTB one, in the same good agreement with the BEs by Ferrero et al.[24]. When comparing the GFN2-xTB results to the ONIOM ones, we see a systematic underestimation of GFN2-xTB results; nevertheless, as can be seen from Fig. 5.9-Right, the correlation between the GFN2-xTB and DLPNO-CCSD(T) BE(0) distributions is quite good, especially considering the negligible computational cost of GFN2-xTB compared to DLPNO-CCSD(T) one. In section 5.4.2.5 of this thesis we will report an example of the relevance of the BE distribution in the astrochemical models using the ONIOM refined BE(0) distribution for the case of $H_2O$ adsorption.

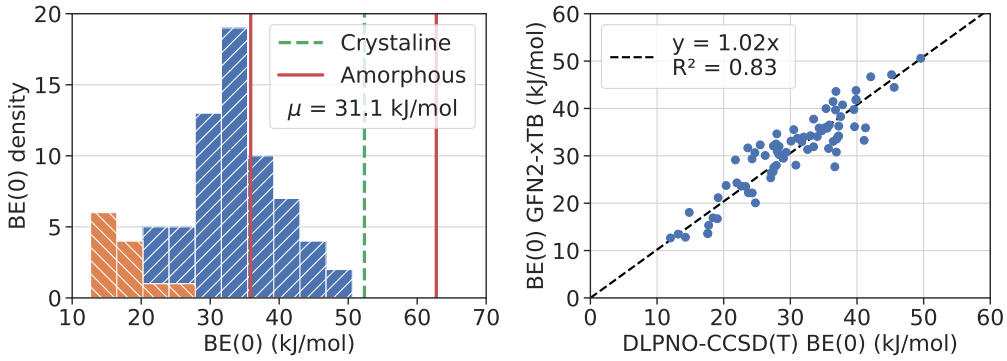



**Figure 5.14:** Left: BSSE corrected $NH_3$ BE(0) distribution at DLPNO-CCSD(T)/aug-cc-pVTZ level. ZPE calculated at ONIOM(B97D3/aug-cc-pVTZ:GFN2-xTB) level. Dotted green and continuous red lines are from Ferrero et al.[24] Right: GFN2-xTB BE(0) distribution against the DLPNO-CCSD(T) one. Best fit represented by the dashed line.

## 5.4 Molecular adsorption of Water

After $H_2$, water is the second most abundant molecule in the ISM. It is formed in the gas phase[29–32] and, with better efficiency, on the surface of interstellar icy grains[8,33–36]. Water vapor in large quantities (abundance up to about 2 $\times$ $10^{-4}$ with respect to $H_2$) is observed in a multitude of warm objects[37–43] (T$\geq$50 K) such as hot corinos and galactic hot cores, protoplanetary disks, protostellar shocks, nearby and far galaxies, etc. In colder (T$\leq$50) regions such as molecular clouds, prestellar cores, and the outer protoplanetary disks, water vapor is also observed but in lower abundance ($\leq 10^{-8}$). Furthermore, frozen water is believed to constitute the amorphous water surface (AWS) mantles of interstellar dust grains[44–46]. Rocky planets, as well as asteroids and comets, which are believed to bring water to newly formed planets, are formed by the aggregation of interstellar dust grains. Leading to the question, how much of the water stays frozen of the grain mantles during the accretion process? A key parameter to answer his question is of course the binding energy of water.

### 5.4.1 Methodology: Short summary

The methodology used to compute the BE distribution of water is almost the same as the one used previously for ammonia. The grain model used is the same, as are the "fixed", "partially unfixed", and "vibrational frequencies" computations steps. The main difference are, first, the version of xTB going from 6.3.3 for ammonia to 6.4.1 for water, and second, the number of random rotation for each adsorbed molecules going from 1 for ammonia to 3 for water. Three rotations for 162 grid points brings the total number of starting structures to 486, increasing significantly our sample and possibly, the final coverage of the grain.

#### 5.4.1.1 Refinement of the GFN2-xTB BE distribution using the ONIOM method

The ONIOM refinement of the water BE(0) distribution is equivalent to the one presented in section 5.3.1.1. Each GFN2-xTB final structures are refined via a multilevel ONIOM(QM:QM2) approach as implemented in ORCA (v.5.0.2).[22,23] The model zone was optimized at B97-3c method[47] instead of B97D3[16,17] for the $NH_3$ case. After the geometry optimizations, the ener-

gies of the structures were computed using DLPNO-CCSD(T)[20] and aug-cc-pVTZ basis as the primary basis set[18,19], and aug-cc-pVTZ/C as the auxiliary one[21] (for the same reasons as seen before for the $NH_3$ distribution).

## 5.4.2 Results and discussion

### 5.4.2.1 Cleaning of the structures and BE(0) distribution

After obtaining the 486 final structures through our BE computation scheme we performed several quality checks: 66 structures were discarded for re-starting the "unfixed" computation more than once (no structures with imaginary frequencies), leading to 420 structures. After removing the duplicated structures, we end up with 133 unique adsorption cases.

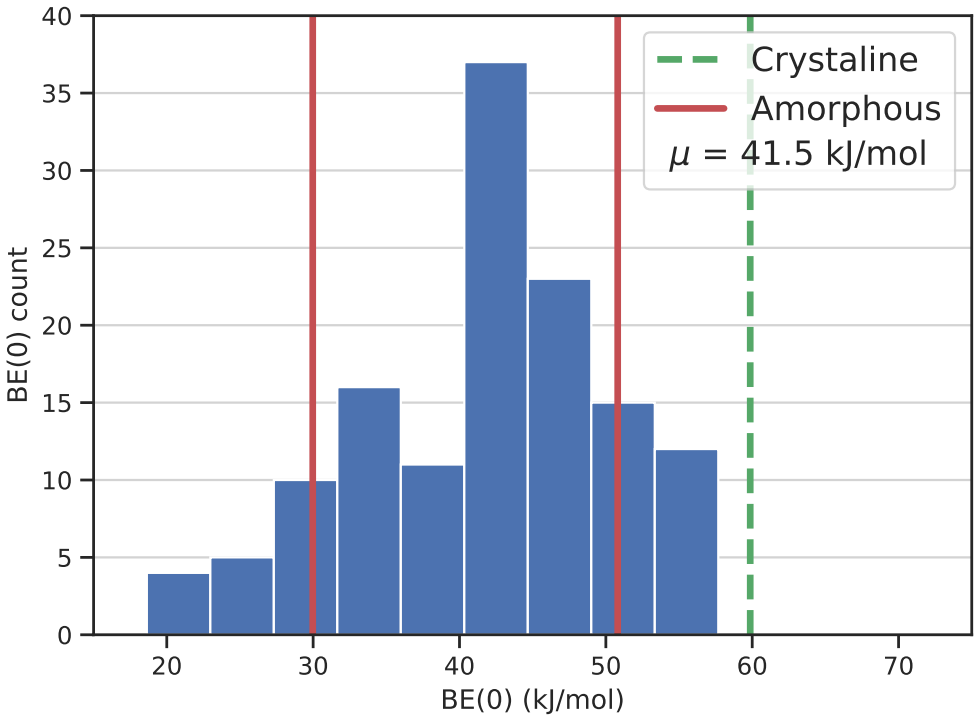



**Figure 5.15:** BE(0) distribution for $H_2O$ adsorption at the SC grain model. Dotted green and continuos red lines are from Ferrero et al.[24].

The final BE(0) distribution is shown in Figure 5.15. As for $NH_3$, the

crystalline slab BE(0) value, and the amorphous slab minimum and maximum BE(0) values from Ferrero et al.[24] are highlighted on the plot. The Ferrero et al. BE(0) DFT values are well within the present distribution, particularly for the periodic amorphous model. BE(0) for the crystalline case are on the higher side of the distribution, as expected (see also the discussion about $NH_3$ adsorption).

#### 5.4.2.2 Number of H-bonds after adsorption: impact on the BE(0) values

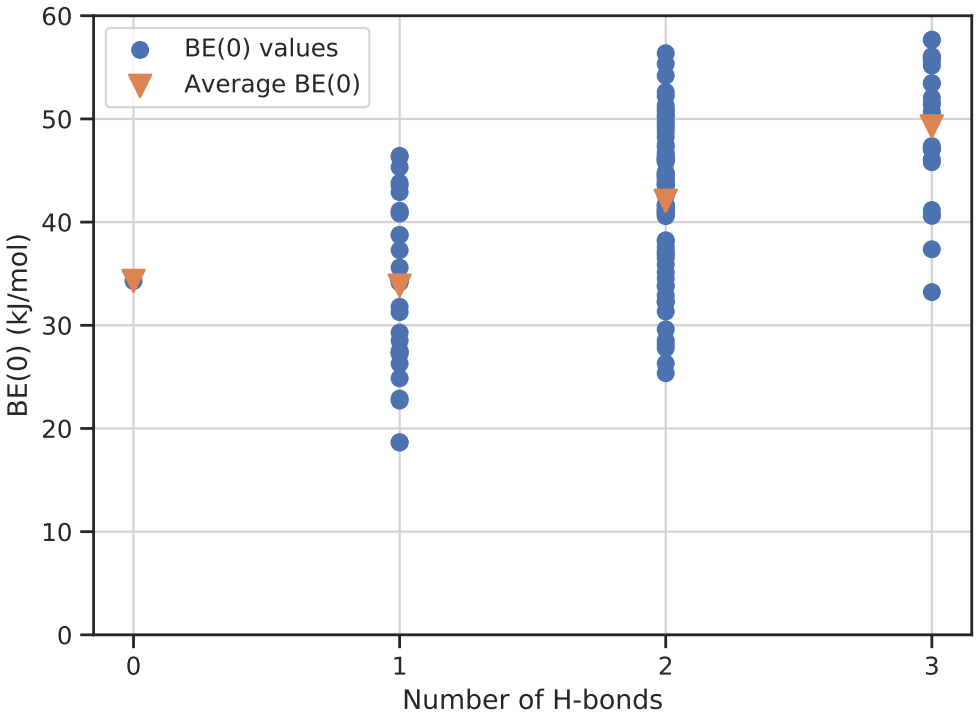
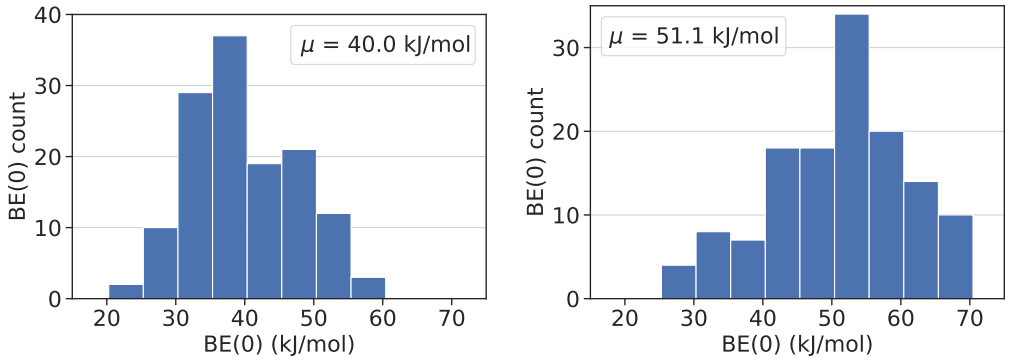



**Figure 5.16:** BE(0) for $H_2O$ as a function of the number of H-bonds with the SC cluster. Orange inverse-triangles represent the average BE(0) value for each different number of H-bonds.

Figure 5.16 shows the BE of the 133 unique structures as a function of the number of H-bonds between $H_2O$ and the SC cluster. The plot shows one structure H-bond free ($\langle BE \rangle$=34.3 kJ/mol), 29 exhibiting one H-bond




**(a)** Fixed results **(b)** Partially unfixed results

**Figure 5.17:** $H_2O$ BE distributions for: (a) the SC grain when entirely fixed; (b) partially unfixed.

($\langle$BE$\rangle$=33.8 kJ/mol), 78 involved in two H-bonds ($\langle$BE$\rangle$=42.0 kJ/mol) and 25 forming up to three H-bonds ($\langle$BE$\rangle$=49.2 kJ/mol). Thus, the lowest BE values of Ferrero et al.[24] are closer to the average values for a singly H-bonded structures, while the highest ones are closer to structures involving three H-bonds.

As already discussed in Chapter 3, Wakelam et al.[48] and Das et al.[49] also computed the BE of several molecules relevant to the interstellar medium. For water, Wakelam et al.[48] computed a BE of 38.2 kJ/mol, close to our average BE value of 41.6 kJ/mol, while Das et al.[49] reported a computed value of 34.6 kJ/mol, close to the average for structures involving one H-bond.

#### 5.4.2.3 Effect of partially unfixing the SC

Figure 5.17 plots the BE distribution computed when the SC cluster is kept at fixed geometry leaving only the adsorbed $H_2O$ as mobile during the geometry optimization as well as when also the water molecules included in a 5 Å radius sphere around the adsorbed $H_2O$ are allowed to relax. Figure 5.17 shows a rather dramatic change compared to the $NH_3$ case (*vide supra*): while the lowest BE value remains similar, the whole distribution is re-shuffled towards higher BE values, with the average BE value moving from 40.0 kJ/mol for the fixed grain to 51.1 kJ/mol for the unfixed one.

#### 5.4.2.4 Correlation between BE(0) and BE: circumventing the expensive computation of the Hessian matrix

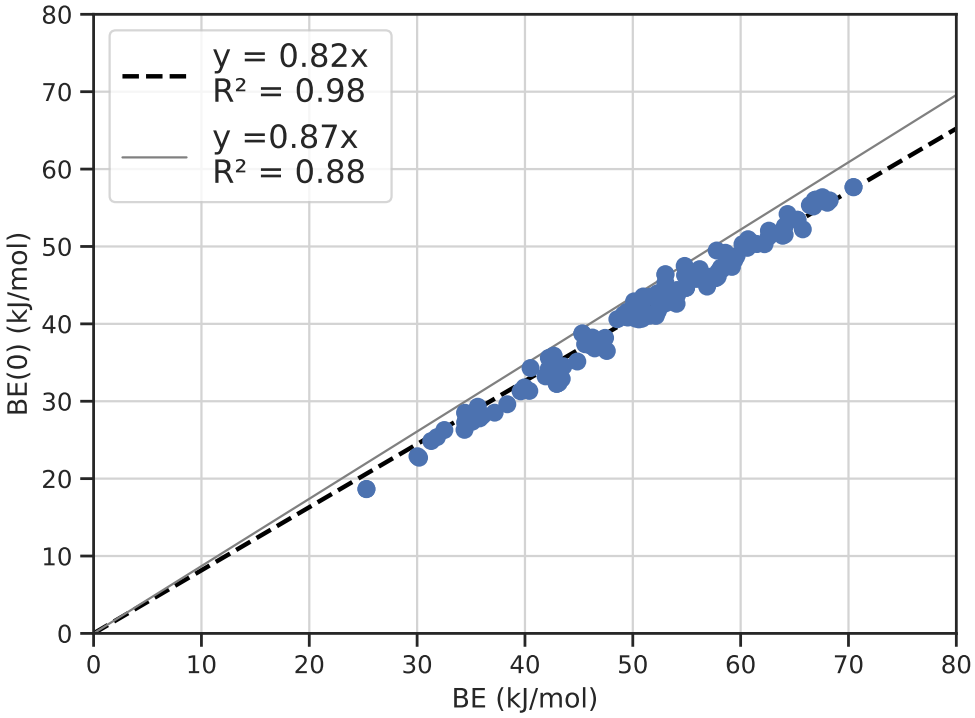



**Figure 5.18:** Linear correlation between BE and zero point energy corrected BE(0) values (kJ/mol) at GFN2 level from a previous work[1] (continuous line) and for all considered case of $H_2O$ adsorption on the SC grain of the present work (best fitted dashed line on filled circle points).

The ZPE corrected BE were plotted against the ZPE uncorrected ones (Figure 5.18), alongside their linear "best fit" line (dashed line) and the computed linear best fit line of a previous work[1] (gray line). Compared to the $NH_3$ results (see Fig. 5.11), the correlation between the $H_2O$ results and the Germain et al.[1] results is less good.

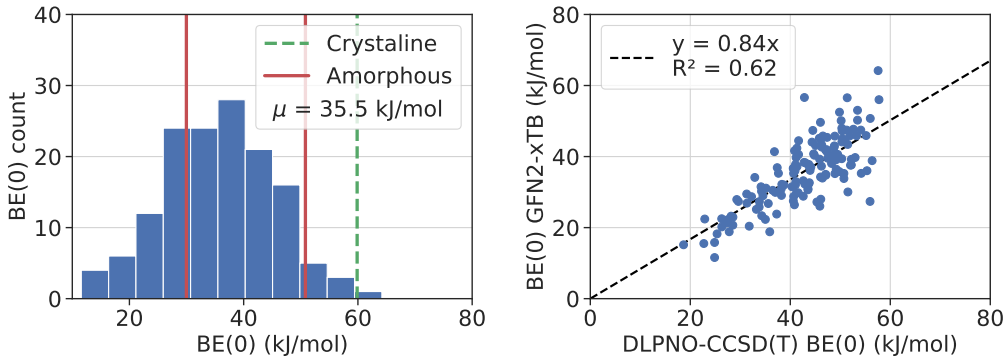



**Figure 5.19:** Left: BSSE corrected $H_2O$ BE(0) distribution at DLPNO-CCSD(T)/aug-cc-pVTZ level. ZPE calculated at ONIOM(B97-3c:GFN2-xTB) level. Dotted green and continuous red lines are from Ferrero et al.[24] Right: GFN2-xTB BE(0) distribution against the DLPNO-CCSD(T) one. Best fit represented by the dashed line.

#### 5.4.2.5 ONIOM refinement: BE(0) distribution

To refine the GFN2-xTB ZPE corrected results shown in Fig. 5.9-Left we performed, in collaboration with Lorenzo Tinacci, a ONIOM calculation at the ONIOM(DLPNO-CCSD(T)//B97-3c:GFN2-xTB) level. These refined BE(0) ONIOM values are shown in Fig. 5.19-Left. The data is, like for ammonia, more spread out in the high BE(0) values but also in the low BE(0) values, with binding energies ranging from 18.7 to 57.7 kJ/mol. The average BE value is 35.5 kJ/mol, $\sim$ 6 kJ/mol less than the GFN2-xTB results, two times more than for ammonia. When comparing the GFN2-xTB results to the refined ONIOM ones we observe a systematic underevaluation of GFN2-xTB. Nevertheless, and unlike the refined $NH_3$ results, the correlation between the GFN2-xTB and DLPNO-CCSD(T) distributions is quite weak, as can be seen on Fig. 5.19-Right, validating the need for a refinement of the GFN2-xTB geometries with a better level of theory.

#### 5.4.2.6 Application of the BE(0) ONIOM distribution: gaseous water in protoplanetary disks

In this part we present a simplified model of gaseous water in protoplanetary disks with a single BE value and with the previously refined ONIOM dis-

tribution to highlight the importance of a BE distribution for astronomical models.

In the interstellar gas, both cold and warm, the abundance of gaseous water is governed by (i) the photo-desorption of water ice, (ii) its photo-dissociation into atomic oxygen, that reforms in part gaseous water, (iii) the freezing of gaseous water onto the grain surfaces, and (iv) its thermal desorption[50]. As thermal desorption is the main process of gaseous water abundance in these parts ($< 200$ K) of the gas, the BE is a crucial parameter for studying this quantity.

**5.4.2.6.1 Brief description of the model** The astrochemical model used is a modified and updated version of the model presented by Dominik et al. in 2005[50]. It computes the gaseous and frozen abundance of water by assuming that atomic oxygen freezing on the grains instantaneously form frozen water, the $H_3^+$ density is equal to the electron density, and that a fraction of the ice is sublimated into the gas-phase.

At steady state, the water gaseous abundance is computed analytically solving the following equations:

$$\begin{aligned} n_O \; k_{accO} &= k_{phd} \; n_{H_2O} - k_{formH_2O} \; n_O \; n_{H_3^+} \\ n_{H_2O} \; (k_{accH_2O} + k_{phd}) &= k_{formH_2O} \; n_O \; n_{H_3^+} + k_{phdes} \; n_{gr} \; k_{thdes} n_{ice} \\ n_{H_2O} + n_O + n_{ice} &= A_{ox} \; f_{ice} \; n \end{aligned} \tag{5.4}$$

where $n_O$, $n_{H_2O}$, $n_{H_3^+}$, $n_{gr}$, $n_{ice}$ and $n$ are, respectively, the number density of: gaseous atomic oxygen, water, $H_3^+$, dust grains, frozen water, and the total H-nuclei density; $k_{accO}$ and $k_{accH_2O}$ are the accretion rates of O and $H_2O$ onto the grain surface; $k_{phd}$, $k_{phdes}$ and $k_{thdes}$ are the gaseous water photo-dissociation rate, the frozen water photo-desorption rate, and the frozen water thermal desorption rates; $k_{formH_2O}$ is the formation rate of gaseous water from atomic oxygen; $A_{ox}$ is the total volatile oxygen elemental abundance (*i.e.* of the oxygen no trapped in the dust grains); $f_{ice}$ is the fraction of sublimated ice.

**5.4.2.6.2 Results** Figure 5.20 shows the results of the modeling for one BE value of 46.6 kJ/mol (5600 K) taken from Wakelam et al. 2017[48]. Figure 5.21 shows the results of the modeling using our distribution of water BE. With the results of Wakelam et al. water ice is entirely sublimated at a temperature of 150 K (light green region of Fig. 5.20), the water abundance

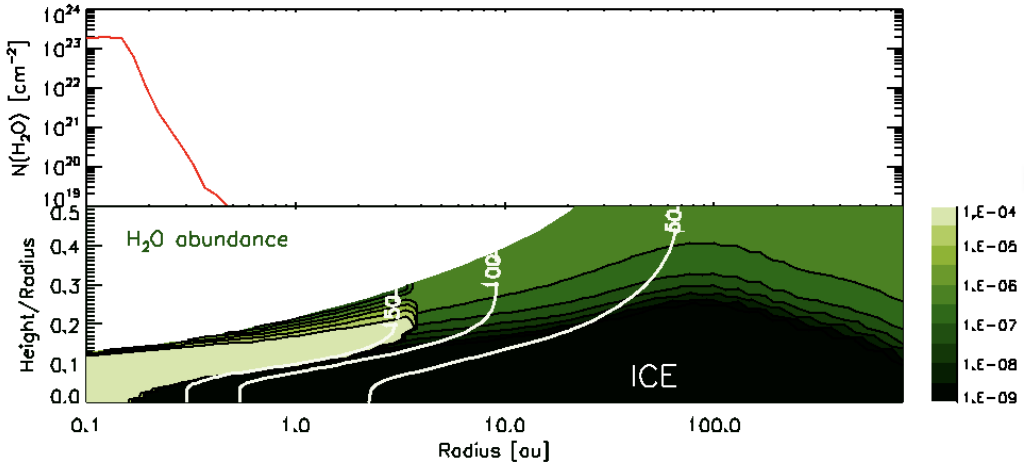



**Figure 5.20:** Astrochemical modeling results obtained assuming the water BE of 46.6 kJ/mol (5600 K) from[48]. *Upper panel:* Gaseous water column density as a function of the radius. *Lower panel:* Water abundance as a function of the radius and the height/radius. The white solid lines indicate where the dust temperature (assumed to be equal to that of the gas) is equal to 150, 100 and 50 K, respectively. In the black region, water is frozen onto the grain ice.

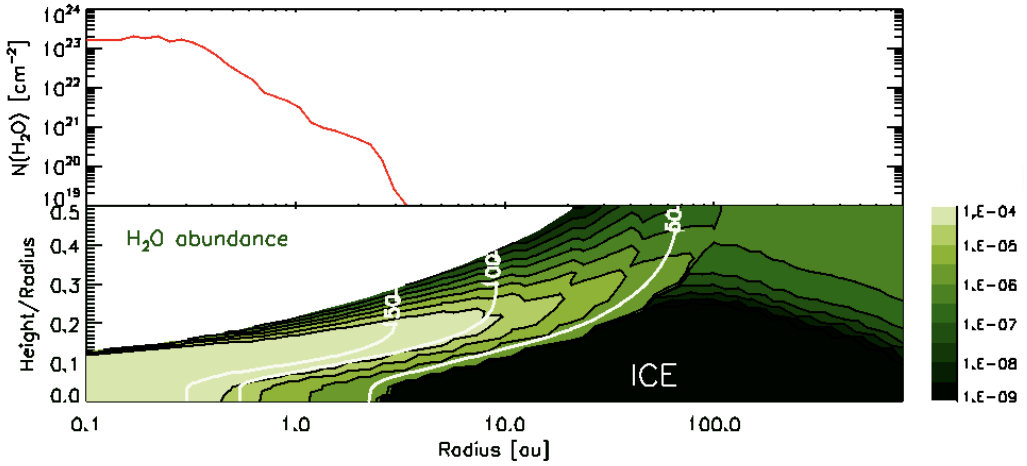



**Figure 5.21:** As Fig. 5.20 but assuming the distribution of water BE computed in this work (CSDP).

reaches a value of $10^{-4}$ in the $\leq 3$ au zone, $\geq 0.2$ au in the midplane. In regions where water ice does not sublimate, the abundance of gaseous water is governed by the photo-desorption of the ices, around $10^{-7} - 10^{-6}$ (dark green), and the freezing of water onto the grain surfaces, $10^{-8}$ (black).

With our BE distribution of water, gaseous water is present in a larger region, $\geq 3 \times 10^{-4}$ (lightest green region), where the dust temperature is $\geq 100$ K. The region with gaseous water abundance $\geq 10^{-8}$ is much more extended on the disk plan and on the layers above it, while the ice region is more reduced and starts at 2 au instead of 0.2.

Using the Wakelam et al.[48] value, the column density of gaseous water, $N(H_2O)$, is, in the $0.1 - 0.16$ au region, $2 \times 10^{23}$ cm$^{-2}$ and then drops by more than a factor of 100 in the $\geq 0.3$ au region. With our BE, the $N(H_2O)$ peak levels up to 0.3 au and then decreases gently to a factor 100 lower than the peak value at $\geq 1$ au. This is due to the fact that 80% of the BEs of the distribution are inferior to the wakelam et al.[48] value. The 20% of the BEs higher than the wakelam et al.[48] value indicate that 20% of the ice will need a higher dust temperature to sublimate. In fact, with our distribution, $\sim 10\%$ of the water stays frozen. If this ice stays trapped, it could remain in planetesimals, eventually forming rocky planets and asteroids inside the standard snowline, the distance from which water stays frozen, and impact their water content. Whether this has an impact on the origin of terrestrial water would require a more sophisticated model, beyond the scope of this work.

## 5.5 Molecular adsorption of Methanol

Methanol is a molecule of biochemical relevance,[51] and can be considered as a potential precursor for more complex species such as sugars or amino acids[52]. It has been observed in both the gas-phase and the solid phase of the ISM from sources such as dark clouds, hot cores, protoplanetary disks, protostars, etc[53–60]. In the gas phase, and more particularly in cold (10 K) dark clouds, it has been observed at an abundance relative to $H_2$ of approximately $10^{-9}$.[61,62] For low mass protostars and hot cores ($\geq 90$ K) it has been observed at relative abundance of $10^{-9} - 10^{-7}$ and $10^{-7} - 10^{-6}$ respectively[63–66]. In the gas-phase, methanol is believed to be formed via the additive association of $H_2O$ and $CH_3OH_2^+$, creating protonated methanol that will then produce methanol via dissociative recombination. However, the formation

of methanol via gas-phase reactions does not account for its abundance in the ISM[67,68]. Methanol is one the the main constituent of the icy grains mantle (with water and CO), and on the grain surface it is believed to be formed via subsequent hydrogenation of CO[69–72]. The presence of gas-phase methanol in warm sources can be explained via the removal of the grain mantles in these environment through thermal desorption, but the case of colder sources is more complex even though it could be partly explained by photodesorption[61]. Co-desorption of methanol through thermal desorption of CO has been proposed as a possible indirect pathway explaining methanol abundance in low temperature environments such as dark clouds[73], but many details are still to be clarified. Predicting the methanol BE distribution is, therefore, important to assess its distribution in both gas-phase and within the icy mantle of the ISM grains.

### 5.5.1 Methodology: Short summary

The methodology used to compute the BE distribution of methanol is identical to the one adopted for water and ammonia. For methanol, however, we propose a more efficient way to compute the Hessian matrix for the frequencies and ZPE data, by limiting its calculation to the zone allowed to relax during the geometry optimization (*vide supra* for further details). As the plain *xTB* code does not allow to compute the Hessian matrix for a subset of atoms, we switch to the *ORCA*[22] program which is capable of running QM/QM calculation within the ONIOM model chemistry (see Computational section for ONIOM further details). The trick was to compute both, the high level zone (the movable zone of the SC model including methanol) and the real system (the whole cluster) at the same GFN2-xTB level. In this way we restricted the vibrational frequency calculation to the high level zone, dramatically reducing the computation time from 30-50 min (full GFN2-xTB Hessian) to 5-10 min (high level zone only GFN2-xTB Hessian). To check for the accuracy of such method, we compared the ZPE corrected BE computed in section 5.4 for the water adsorption with the corresponding ones computed with the QM/QM method. Figure 5.23 shows an excellent agreements, proving the equivalency of both methods for vibrational frequencies computation. In Figure 5.24 we plotted side by side the GFN2 BE(0) distribution with vibrational frequency calculation computed on the full grain (same as Fig. 5.15), and the GFN2 BE(0) distribution with vibrational frequency calculation computed using only the unfixed fragment via ORCA.

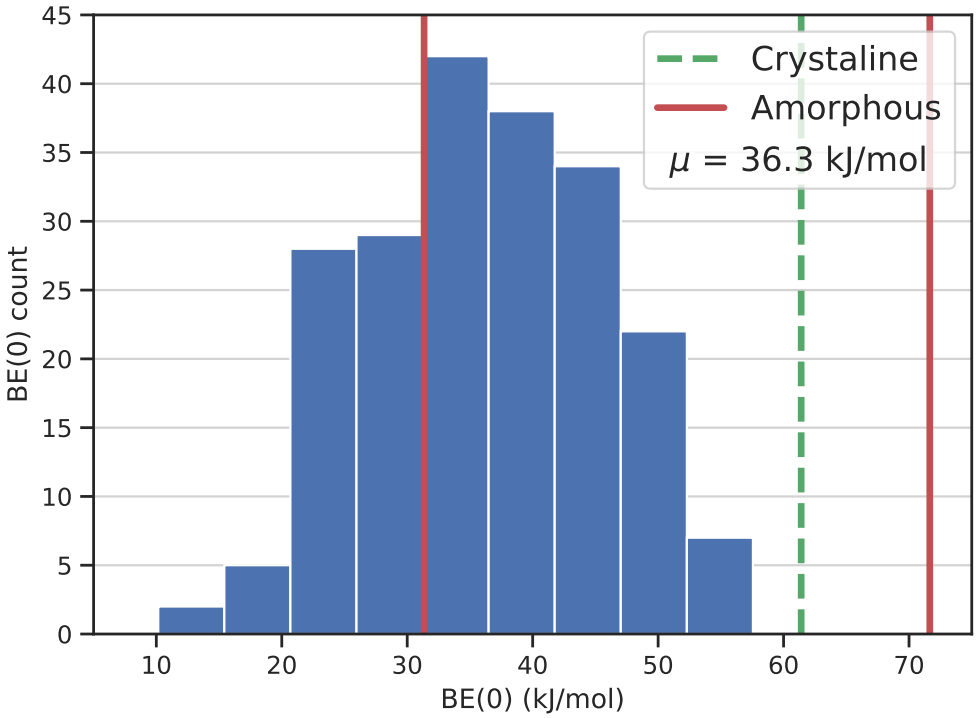



**Figure 5.22:** BE(0) distribution for $CH_3OH$ adsorption at the SC grain model. Dotted green and continuous red lines are from Ferrero et al.[24].

The two distributions show a mean BE(0) value with only 0.2 kJ/mol of difference, and bins populations almost identical with only a few variations.

## 5.5.2 Results and discussion

### 5.5.2.1 Cleaning of the structures and BE(0) distribution

The ZPE corrected BE are showed in Figure 5.22. The same quality checks carried out as for $NH_3$ and $H_2O$ were adopted for $CH_3OH$ to discard several duplicated structures. Of the 486 final computed structures, 82 were discarded following too many iteration of the "unfixed" step or presence of imaginary frequencies during the vibrational computation. The total number of unique adsorption was finally found to be 207. The number of unique ad-

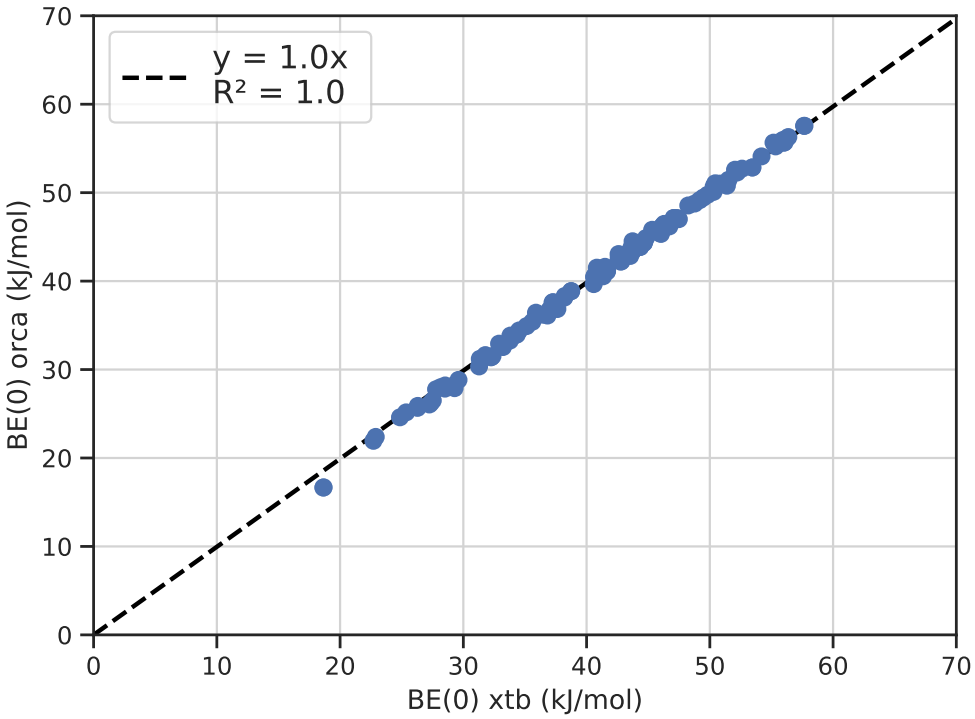



**Figure 5.23:** For the adsorption of water: Linear correlation between GFN2-xtb BE(0) computed at QM/QM level with ORCA code for the model zone only and BE(0) values for the real system. Best fit line as dashed line.

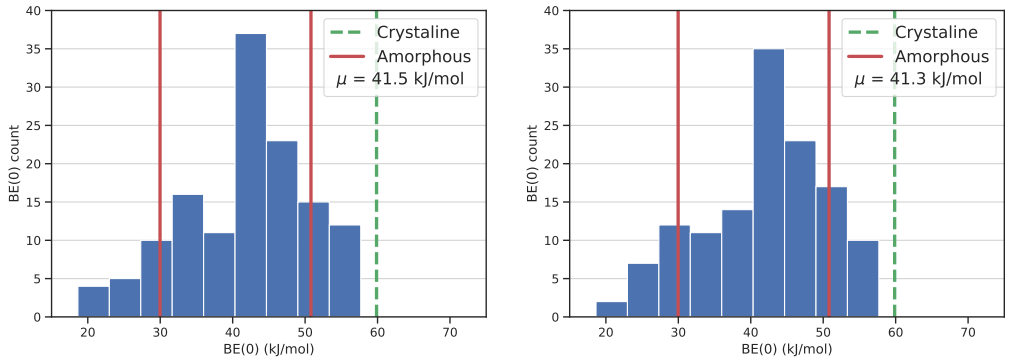
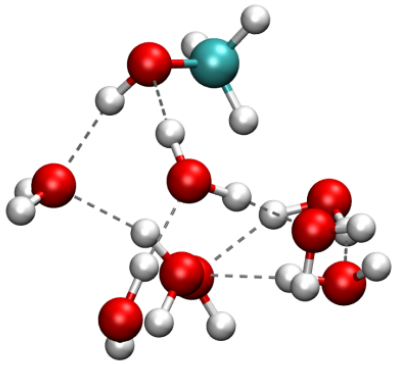


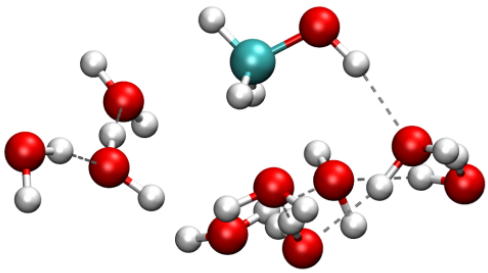

**Figure 5.24:** Left: ZPE corrected $H_2O$ BE(0) distribution at the GFN2 level with vibrational frequency calculation computed on the full grain. Right: ZPE corrected $H_2O$ BE(0) distribution at the GFN2 level with vibrational frequency calculation computed only on the unfixed fragment via ORCA. Dotted green and continuous red lines are from Ferrero et al.[24].

sorption cases is significantly higher for methanol than for water (207 against 133 cases). This difference is probably due to the presence of the methyl $CH_3$ group which renders the potential energy surface for adsorption more rich than for water. For methanol, we found 4 main ways in which it bonds to the grain (showed in Figure 5.25). For two of them, it behaves as a very weak C-H bond donor to the oxygen atoms of the icy cluster (section d of Figure 5.25) or as a regular H-bond donor through the OH group (section b of Figure 5.25); it can also behave as a H-bond acceptor, through the lone pairs of its oxygen atom (section c of Figure 5.25), and lastly, can behave simultaneously as (b) and (c) (section a of Figure 5.25). The (d) approach, only occurs 4 times out of 207 cases, and is also the weakest one, with $\langle BE \rangle$ =14.0 kJ/mol; the (c) approach is present 24 times with $\langle BE \rangle$=30.5 kJ/mol; the (b) case contributes with 67 structures and $\langle BE \rangle$=29.2 kJ/mol; finally, the (a) model is the dominant one, for both presence (102 structures) and BE strength ($\langle BE \rangle$=43.3 kJ/mol). Two structures not included in the previous list are standing apart (shown in Fig. 5.26); the first one, is bounded by pure dispersive interactions (no H-bonds accepted/donated) with a sizeable BE of 28.0 kJ/mol; the second case is a combination of the previous (b) and (d) cases, accounting for a BE of 37.7 kJ/mol.

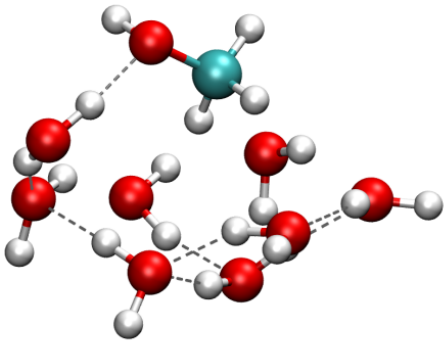
Figure 5.22 shows the BE(0) distribution with $\langle BE(0) \rangle$ value of 36.3 kJ/-mol. Alike the case of ammonia and water, we reported also the BE(0) values



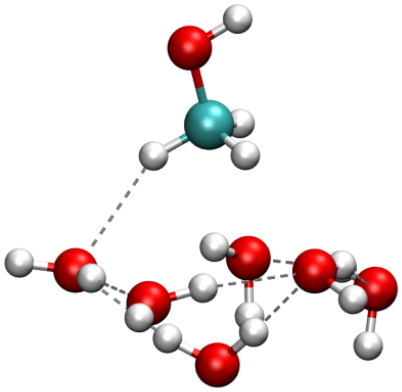
**(a)** $CH_3OH$ acting as both H-bond donor/acceptor



**(b)** $CH_3OH$ acting as H-bond donor

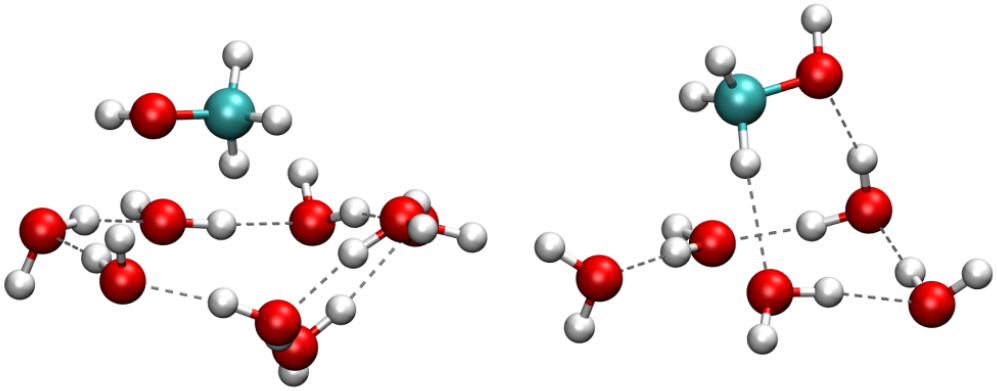


**(c)** $CH_3OH$ acting as H-bond acceptor



**(d)** $CH_3OH$ acting as C-H donor

**Figure 5.25:** Different modes of $CH_3OH$ adsorption at the SC grain. Only the active part (molecules inside the 5 Å radius) is showed.

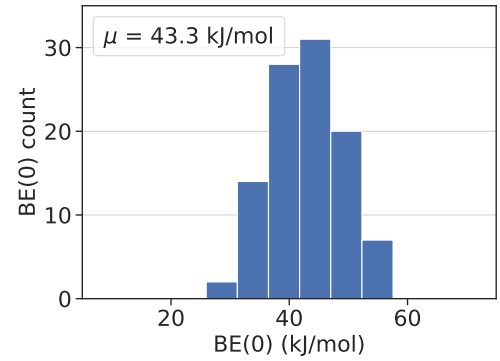


**(a)** $CH_3OH$ bounded through dispersive interactions

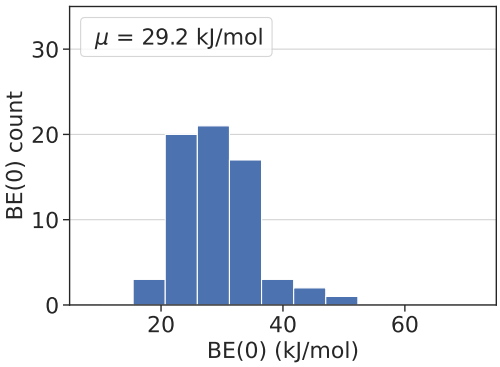
**(b)** $CH_3OH$ acting as both H-bond acceptor (c) and C-H donor (d)

**Figure 5.26:** Two structures standing apart from the manifold showed in Fig. 5.25. Only the active part (molecules inside the 5 Å radius) is showed.

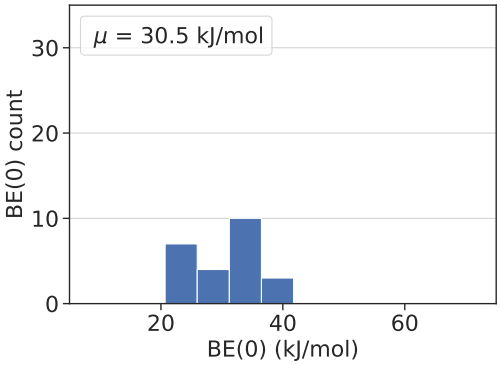
computed for methanol on a crystalline and amorphous periodic (minimum and maximum) slab by Ferrero et al.[24] The discrepancy in the values are similar to the ones already discussed for ammonia and will not be repeated here. Among other theoretical values, Wakelam et al.[48] report two different BE values: one similar to our case (c), with a BE of 42.4 kJ/mol and one resembling case (b), with a BE of 37.4 kJ/mol. The comparison with our data is not straightforward, as ZPE is missing in the Wakelam BEs. Our BE(0) distribution was also split to highlight the four main binding approaches described above, and plotted on Figure 5.27. The comparison with the Wakelam's data is with the average BE(0) for the (b) and (c) modes. In our case, the $\langle BE(0) \rangle$ of 29.5 kJ/mol and 30.5 kJ/mol are very close, albeit the (b) distribution is much more populated, as discussed before. The distribution referring to the (a) case exhibits the higher $\langle BE(0) \rangle$ of 40.7 kJ/mol, as expected. Das et al.[49] reported a BE value of 37.5 kJ/mol, also very close to our average value.



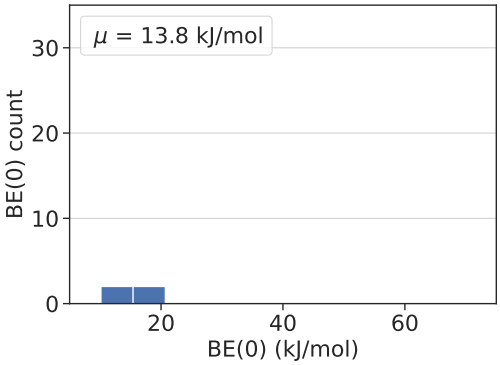

**(a)** $CH_3OH$ acting as both H-bond donor/acceptor




**(b)** $CH_3OH$ acting as H-bond donor




**(c)** $CH_3OH$ acting as H-bond acceptor

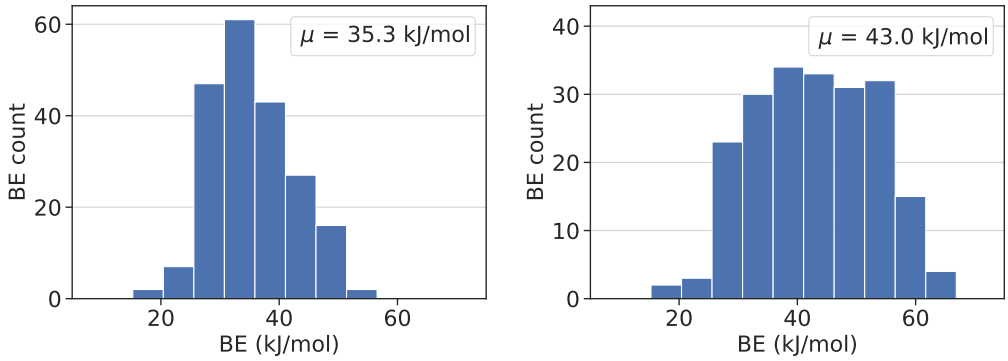



**(d)** $CH_3OH$ acting as C-H donor

**Figure 5.27:** Distribution of the BE(0) for the four different binding methods presented in Figure 5.25.

#### 5.5.2.2 Effect of partially unfixing the SC

As a further check of the role played by the geometry relaxation, we plot in Figure 5.28 the BE(0) distributions coming from the fixed SC geometry and from the optimization of the 5 Å sphere around the adsorbed $CH_3OH$. As shown, the two distributions are different in shape and position, the $\langle BE(0) \rangle$ moving from 35.3 kJ/mol (fixed SC) up to 43 kJ/mol for the partly optimized case. This is expected, as $CH_3OH$ will force the local environment of the adsorption site to relax in order to maximize the H-bond interactions.




**(a)** Fixed SC geometry **(b)** Partly optimized SC geometry

**Figure 5.28:** $CH_3OH$ BE(0) distributions for: (a) the case where the SC grain geometry was fixed. (b) only the water molecules included in a 5 Å radius sphere around the adsorbed $CH_3OH$ are optimized.

#### 5.5.2.3 Correlation between BE(0) and BE: circumventing the expensive computation of the Hessian matrix

As already carried out for $NH_3$ and $H_2O$ cases, we compared the BE(0)s against the BE ones (Figure 5.29) to see the role of ZPE on the electronic BE. We also included the "best fit" line (dashed line) and the computed linear best fit line from a previous work[1] (gray line). The correlation between the previous fit and the new one is excellent. Therefore, we compare the BE(0) distribution as previously described with that obtained by multiplying the BE value with the scaling factor 0.87 reported in Figure 5.29. The result is showed on Figure 5.30.

The distribution (b) of Fig. 5.30 was chosen as to have the exact same bins as Fig. 5.22 (also shown in Fig. 5.30-a) to better observe the population of each bins of the distribution. We can observe a small change in average BE value of 1.1 kJ/mol between the two distributions. Population of bins representing BE inferior to 20 kJ/mol and superior to 50 kJ/mol seems to be the closest in BE count, but were also the ones with the less number of BE, while the bins between this range, possessing the highest number of BE count, suffered the most visible changes. In terms of individual and average BE value this correction can be relevant, but for a BE distribution obtained results seems to far from the ones obtained via vibrational frequency computations to be conclusive. Probably this kind of correction is superfluous if a QM/QM

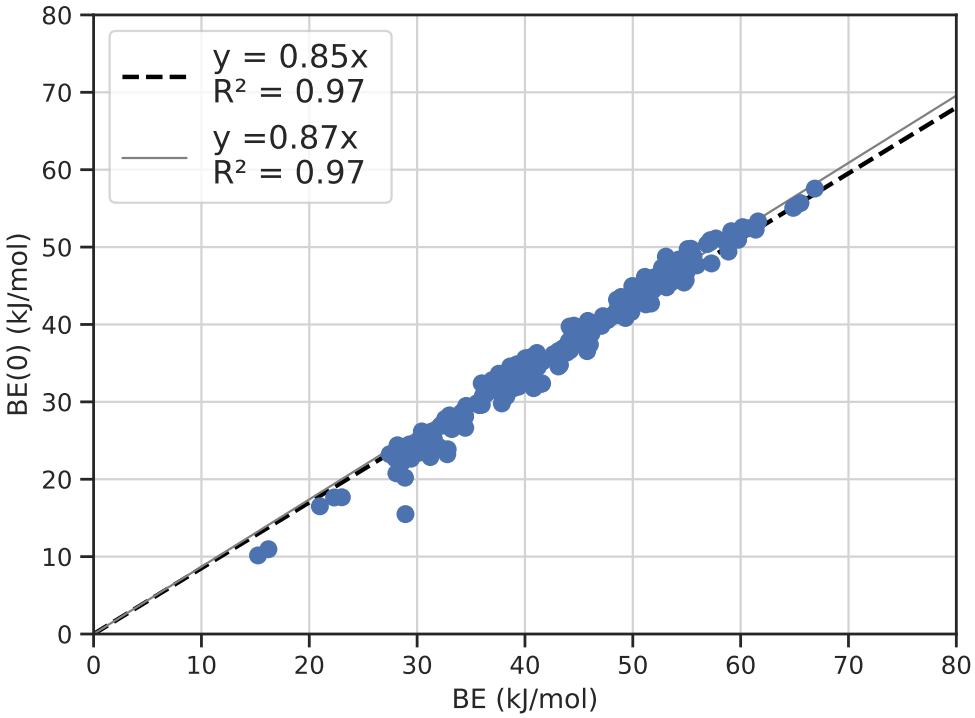
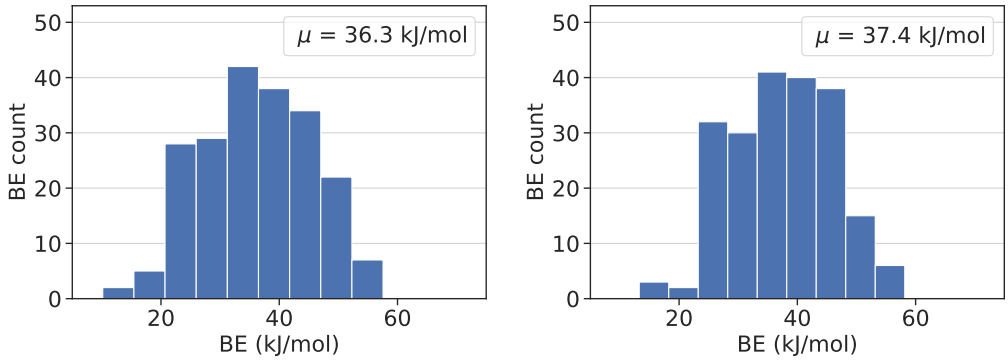



**Figure 5.29:** Linear correlation between GFN2 BE and BE(0) values (kJ/-mol) from a previous work[1] (continuous line) and for all considered cases of $CH_3OH$ adsorption on the SC model grain here considered (best fitted dashed line on filled circle points)

computation of vibrational frequencies, a lot less computationally demanding than a direct GFN2-xTB one, is possible.

## 5.6 IR spectra for methanol on the icy grain

As a proof of concept study, we computed the IR spectrum for each case of the final set of unique structures, as a sum of the GFN2-xTB IR spectra computed with ORCA for methanol interacting with the subset of the icy water molecules (*Fragment*), as discussed in the previous sections. The same approach was followed to compute the IR spectrum for the bare icy cluster (*Grain*), where the vibrational features were only relative to the water




**(a)** BE ZPE corrected via vibrational frequency computation.

**(b)** BE ZPE corrected via fitting using previous data.

**Figure 5.30:** $CH_3OH$ BE(0) distributions (already showed in Fig. 5.22) ZPE corrected via vibrational frequency computation (a) and $CH_3OH$ BE(0) distributions ZPE corrected using the distribution of Fig. 5.29 taken from a previous publication[1].

molecules belonging to the fragments defined for each adsorption case. The result is shown in Figure. 5.31. The spectra for *Fragment* and *Grain* were normalized with respect to the number of considered structures contributing to the whole spectrum.

Methanol literature data, since its first observation back in 1970, was through the radio line emission at 36 $cm^{-1}$, due to a K-type doubling transition splitting of the ground internal-rotation state[53]. Astrochemical evidence for methanol infrared spectral features are still missing or uncertain while laboratory spectra of mixture of $H_2O/CH_3OH$ are available[74].

Figure 5.31 shows a small general bathochromic shift between the bare icy fragment spectrum and the one involving methanol. Some spectral feature of methanol are evident and were reported in Table 5.2 with their respective counterpart in the fragment's spectrum. The same work was done for the spectral feature of water, of which only the bending mode at 1539 $cm^{-1}$ was reported to be shifted by $-4$ $cm^{-1}$ in the *fragment* spectrum. The manifold of bands involving the OH/CH stretching modes are too broad to allow the identification of specific features of methanol.

Methanol can be identified via its 1031 $cm^{-1}$ CO stretching mode, 1124 $cm^{-1}$ $CH_3$ rock, 1340 $cm^{-1}$ OH bending mode, 1450 $cm^{-1}$ $CH_3$ deformation mode,

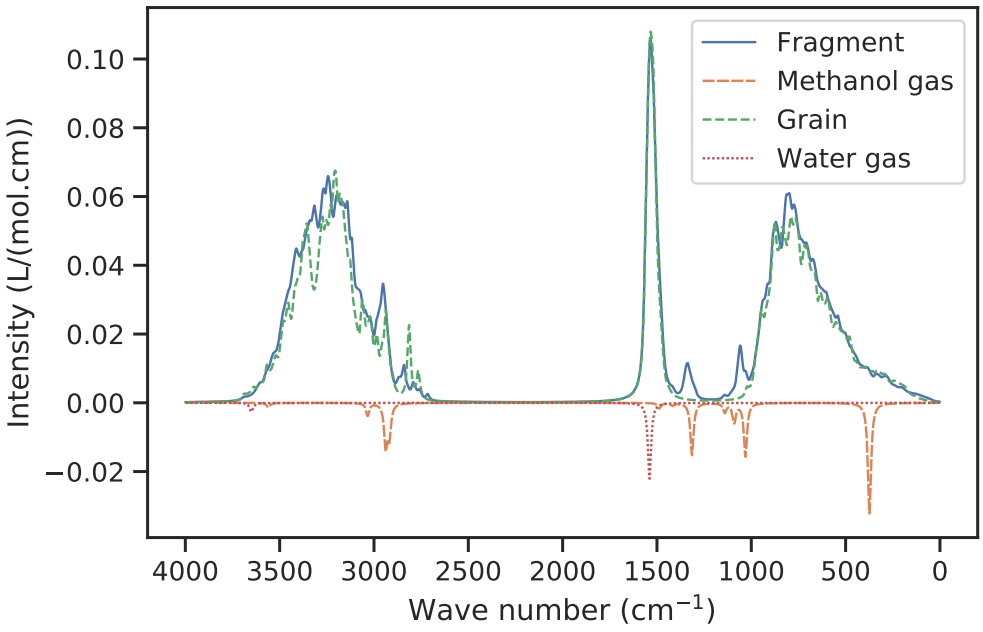



**Figure 5.31:** IR spectrum of: methanol interacting with the icy grain (Fragment: continuous blue line); the bare icy grain (Grain: dashed green line); the gas-phase methanol molecule (densely dashed orange line, inverted for better visibility); the gas-phase water molecule (dotted red line, inverted for better visibility). The intensities (y-axis) are in L/(mol.cm).

$2538-2604$ cm$^{-1}$ combination modes, and 2825 cm$^{-1}$ asymmetric stretching mode[75,76].

While we cannot cope with combination modes, our computed spectroscopic data reported in Table 5.2 and shown in Fig. 5.31 allow us to assign: (i) a relatively intense CO stretching spectral band at 1031 cm$^{-1}$ for the gas phase methanol, hypsochromic shifted at 1058 cm$^{-1}$ (+27 cm$^{-1}$) in the interacting fragment; (ii) a second intense band at 1314 cm$^{-1}$, +24 cm$^{-1}$ in the fragment, identified as the C-O-H bending mode. Two intense bands are present at 2929 and 2937 cm$^{-1}$, the asymmetric and symmetric C-H stretching mode respectively, falling in the manifold of the H-bonded OH stretching frequency (between 3700 and 2700 cm$^{-1}$), rendering *de facto* impossible any specific assignment of these band for the interacting methanol. The same

**Table 5.2:** Selected features of the GFN2-xTB harmonic vibrational features of the gas phase molecules (methanol and water) compared with the corresponding bands in the *fragment*. IR intensities in L/(mol.cm). Labels s, b, ss, as for stretching, bending, symmetric stretching and asymmetric stretching.

| Free molecule | $\nu$ ($cm^{-1}$) | Intensity ($\times 10^{-2}$) | Fragment $\nu$ ($cm^{-1}$) | $\Delta\nu$ ($cm^{-1}$) | Intensity ($\times 10^{-2}$) |
|---|---|---|---|---|---|
| $CH_3OH$ | 1031 s(CO) | 1.6 | 1058 | +27 | 1.6 |
| - | 1314 b(COH) | 1.5 | 1338 | +24 | 1.1 |
| - | 2920 ss(CH) | 1.2 | ?? | ?? | 1.6 |
| - | 2937 as(CH) | 1.4 | ?? | ?? | 1.6 |
| $H_2O$ | 1539 b(HOH) | 2.2 | 1535 | −4 | 9.9 |

reasoning holds true for the 3320 $cm^{-1}$ feature, only present in *fragment*.

While the present approach to compute the harmonic IR spectra (a fragment treated with ONIOM(ORCA:GFN2-xTB) scheme) is promising, much work has to be done for the assignment of the vibrational bands. One possibility is to use isotopic substitution (specifically transforming the regular icy grain in a fully deuterated $D_2O$ one) avoiding overlap with the methanol features. Progress along this line is being carried out in our laboratory.

# References


[1] A. Germain, L. Tinacci, S. Pantaleone, C. Ceccarelli, and P. Ugliengo, "Computer Generated Realistic Interstellar Icy Grain Models: Physicochemical Properties and Interaction with $NH_3$," *ACS Earth and Space Chemistry*, vol. 6, pp. 1286–1298, may 2022.

[2] 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," *WIREs Computational Molecular Science*, vol. 8, p. e1360, jul 2018.

[3] 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," *Physical Review Letters*, vol. 100, p. 136406, apr 2008.

[4] P. Ugliengo, D. Viterbo, and G. Chiari, "MOLDRAW: Molecular graphics on a personal computer," *Zeitschrift für Kristallographie - Crystalline Materials*, vol. 207, pp. 9–24, nov 1993.

[5] A. H. Harvey, "Properties of ice and supercooled water," in *CRC Handbook of Chemistry and Physics (97th ed.)* (W. M. Haynes, D. R. Lide, and T. J. Bruno, eds.), Boca Raton, FL: CRC Press, 2017.

[6] D. N. Theodorou and U. W. Suter, "Shape of unperturbed linear polymers: polypropylene," *Macromolecules*, vol. 18, pp. 1206–1214, jun 1985.

[7] R. Le Gal, P. Hily-Blant, A. Faure, G. Pineau des Forêts, C. Rist, and S. Maret, "Interstellar chemistry of nitrogen hydrides in dark clouds," *Astronomy & Astrophysics*, vol. 562, p. A83, Feb. 2014.

[8] Y. Oba, N. Miyauchi, H. Hidaka, T. Chigai, N. Watanabe, and A. Kouchi, "Formation of compact amorphous $h_2o$ ice by codeposition of hydrogen atoms with oxygen molecules on grain surfaces," *The Astrophysical Journal*, vol. 701, pp. 464–470, 2009.

[9] P. Caselli, C. Walmsley, M. Tafalla, L. Dore, and P. Myers, "Co depletion in the starless cloud core l1544," *The Astrophysical Journal Letters*, vol. 523, no. 2, p. L165, 1999.

[10] E. Keto and P. Caselli, "Dynamics and depletion in thermally supercritical starless cores," *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 3, pp. 1625–1634, 2010.

[11] Y. Aikawa, V. Wakelam, F. Hersant, R. T. Garrod, and E. Herbst, "From prestellar to protostellar cores. ii. time dependence and deuterium fractionation," *The Astrophysical Journal*, vol. 760, no. 1, p. 40, 2012.

[12] O. Sipilä, P. Caselli, E. Redaelli, M. Juvela, and L. Bizzocchi, "Why does ammonia not freeze out in the centre of pre-stellar cores?," *Monthly Notices of the Royal Astronomical Society*, vol. 487, no. 1, pp. 1269–1282, 2019.

[13] A. Crapsi, P. Caselli, M. C. Walmsley, and M. Tafalla, "Observing the gas temperature drop in the high-density nucleus of L 1544," *Astronomy & Astrophysics*, vol. 470, no. 1, pp. 221–230, 2007.

[14] F. Maseras and K. Morokuma, "IMOMM: A new integratedab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states," *Journal of Computational Chemistry*, vol. 16, pp. 1170–1179, sep 1995.

[15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, "Gaussian16 Revision B.01," 2016. Gaussian Inc. Wallingford CT.

[16] S. Grimme, "Semiempirical GGA-type density functional constructed

with a long-range dispersion correction," *Journal of Computational Chemistry*, vol. 27, pp. 1787–1799, nov 2006.

[17] S. Grimme, S. Ehrlich, and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory," *Journal of Computational Chemistry*, vol. 32, pp. 1456–1465, may 2011.

[18] T. Kendall Jr, "Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions," *J. Chem. Phys*, vol. 96, p. 6796, 1992.

[19] R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *The Journal of Chemical Physics*, vol. 96, p. 6796, jun 1998.

[20] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, "Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]," *The Journal of chemical physics*, vol. 148, no. 1, p. 011101, 2018.

[21] F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," *The Journal of Chemical Physics*, vol. 116, pp. 3175–3183, feb 2002.

[22] F. Neese, "The ORCA program system," *WIREs Computational Molecular Science*, vol. 2, pp. 73–78, jan 2012.

[23] F. Neese, "Software update: the orca program system, version 4.0," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 1, p. e1327, 2018.

[24] S. Ferrero, L. Zamirri, C. Ceccarelli, A. Witzel, A. Rimola, and P. Ugliengo, "Binding energies of interstellar molecules on crystalline and amorphous models of water ice by ab-initio calculations," *The Astrophysical Journal*, vol. 904, no. 1, p. 11, 2020.

[25] S. Molinari, C. Ceccarelli, G. J. White, P. Saraceno, B. Nisini, T. Giannini, and E. Caux, "Detection of the 62 Micron Crystalline $H_2O$ Ice Feature in Emission toward HH 7 with the *Infrared Space Observatory*

Long-Wavelength Spectrometer," *The Astrophysical Journal*, vol. 521, pp. L71–L74, aug 1999.

[26] H. Terada and A. T. Tokunaga, "DISCOVERY OF CRYSTALLIZED WATER ICE IN A SILHOUETTE DISK IN THE M43 REGION," *The Astrophysical Journal*, vol. 753, p. 19, jul 2012.

[27] A. Germain, M. Corno, and P. Ugliengo, "Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12953 LNCS, pp. 632–645, Springer, Cham, sep 2021.

[28] L. Tinacci, A. Germain, S. Pantaleone, S. Ferrero, C. Ceccarelli, and P. Ugliengo, "Theoretical Distribution of the Ammonia Binding Energy at Interstellar Icy Grains: A New Computational Framework," *ACS Earth and Space Chemistry*, vol. 6, pp. 1514–1526, jun 2022.

[29] M. Elitzur and T. de Jong, "A model for the maser sources associated with H II regions.," *Astronomy and Astrophysics*, vol. 67, pp. 323–332, July 1978.

[30] W. D. Langer and T. E. Graedel, "Ion-Molecule Chemistry of Dense Interstellar Clouds: Nitrogen-, Oxygen-, and Carbon-bearing Molecule Abundances and Isotopic Ratios," *The Astrophysical Journal Supplement*, vol. 69, p. 241, Feb. 1989.

[31] M. J. Kaufman and D. A. Neufeld, "Far-Infrared Water Emission from Magnetohydrodynamic Shock Waves," *The Astrophysical Journal*, vol. 456, p. 611, Jan. 1996.

[32] D. Hollenbach, M. J. Kaufman, E. A. Bergin, and G. J. Melnick, "Water, $O_2$, and Ice in Molecular Clouds," *The Astrophysical Journal*, vol. 690, pp. 1497–1521, Jan. 2009.

[33] A. G. G. M. Tielens and W. Hagen, "Model calculations of the molecular composition of interstellar grain mantles," *Astronomy and Astrophysics*, vol. 114, pp. 245–260, Oct. 1982.

[34] F. Dulieu, L. Amiaud, E. Congiu, J. H. Fillion, E. Matar, A. Momeni, V. Pirronello, and J. L. Lemaire, "Experimental evidence for water formation on interstellar dust grains by hydrogen and oxygen atoms," *Astronomy and Astrophysics*, vol. 512, p. A30, Mar. 2010.

[35] T. Lamberts, P. K. Samanta, A. Köhn, and J. Kästner, "Quantum tunneling during interstellar surface-catalyzed formation of water: the reaction H + $H_2O_2$ → $H_2O$ + OH," *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)*, vol. 18, pp. 33021–33030, Jan. 2016.

[36] G. Molpeceres, A. Rimola, C. Ceccarelli, J. Kästner, P. Ugliengo, and B. Maté, "Silicate-mediated interstellar water formation: a theoretical study," *Monthly Notices of the Royal Astronomical Society*, vol. 482, pp. 5389–5400, feb 2019.

[37] F. Combes and T. Wiklind, "Detection of Water at z = 0.685 toward B0218+357," *The Astrophysical Journal Letters*, vol. 486, pp. L79–L82, Sept. 1997.

[38] J. Fischer, M. L. Luhman, S. Satyapal, M. A. Greenhouse, G. J. Stacey, C. M. Bradford, S. D. Lord, J. R. Brauher, S. J. Unger, P. E. Clegg, H. A. Smith, G. Melnick, J. W. Colbert, M. A. Malkan, L. Spinoglio, P. Cox, V. Harvey, J. P. Suter, and V. Strelnitski, "ISO Far-IR spectroscopy of IR-bright galaxies and ULIRGs," *Astrophysics and Space Science*, vol. 266, pp. 91–98, Jan. 1999.

[39] E. González-Alfonso, H. A. Smith, J. Fischer, and J. Cernicharo, "The Far-Infrared Spectrum of Arp 220," *The Astrophysical Journal*, vol. 613, pp. 247–261, Sept. 2004.

[40] A. Omont, C. Yang, P. Cox, R. Neri, A. Beelen, R. S. Bussmann, R. Gavazzi, P. van der Werf, D. Riechers, D. Downes, M. Krips, S. Dye, R. Ivison, J. D. Vieira, A. Weiß, J. E. Aguirre, M. Baes, A. J. Baker, F. Bertoldi, A. Cooray, H. Dannerbauer, G. De Zotti, S. A. Eales, H. Fu, Y. Gao, M. Guélin, A. I. Harris, M. Jarvis, M. Lehnert, L. Leeuw, R. Lupu, K. Menten, M. J. Michałowski, M. Negrello, S. Serjeant, P. Temi, R. Auld, A. Dariush, L. Dunne, J. Fritz, R. Hopwood, C. Hoyos, E. Ibar, S. Maddox, M. W. L. Smith, E. Valiante, J. Bock, C. M. Bradford, J. Glenn, and K. S. Scott, "$H_2O$ emission in high-z ultra-luminous

infrared galaxies," *Astronomy and Astrophysics*, vol. 551, p. A115, Mar. 2013.

[41] C. Yang, A. Omont, A. Beelen, E. González-Alfonso, R. Neri, Y. Gao, P. van der Werf, A. Weiß, R. Gavazzi, N. Falstad, A. J. Baker, R. S. Bussmann, A. Cooray, P. Cox, H. Dannerbauer, S. Dye, M. Guélin, R. Ivison, M. Krips, M. Lehnert, M. J. Michałowski, D. A. Riechers, M. Spaans, and E. Valiante, "Submillimeter $H_2O$ and $H_2O^+$emission in lensed ultra- and hyper-luminous infrared galaxies at z 2-4," *Astronomy and Astrophysics*, vol. 595, p. A80, Nov. 2016.

[42] M. Imanishi, K. Nakanishi, T. Izumi, and S. Baba, "ALMA Sub-arcsecond-resolution 183 GHz $H_2O$ and Dense Molecular Line Observations of Nearby Ultraluminous Infrared Galaxies," *The Astrophysical Journal*, vol. 926, p. 159, Feb. 2022.

[43] A. Pensabene, P. van der Werf, R. Decarli, E. Bañados, R. A. Meyer, D. Riechers, B. Venemans, F. Walter, A. Weiß, M. Brusa, X. Fan, F. Wang, and J. Yang, "Unveiling the warm and dense ISM in z $>$ 6 quasar host galaxies via water vapor emission," *Astronomy & Astrophysics*, vol. 667, p. A9, nov 2022.

[44] A. Leger, J. Klein, S. de Cheveigne, C. Guinet, D. Defourneau, and M. Belin, "The 3.1 µm absorption in molecular clouds is probably due to amorphous $H_2O$ ice.," *Astronomy and Astrophysics, vol. 79, no. 1-2, Oct. 1979, p. 256-259.*, vol. 79, p. 256, oct 1979.

[45] E. L. Gibb, D. C. B. Whittet, A. C. A. Boogert, and A. G. G. M. Tielens, "Interstellar ice: The infrared space observatory legacy," *The Astrophysical Journal Supplement Series*, vol. 151, pp. 35–73, 3 2004. $H_2O$ predominent ice constituent in Dense molecular clouds.

[46] A. C. Boogert, P. A. Gerakines, and D. C. Whittet, "Observations of the icy universe," *Annual Review of Astronomy and Astrophysics*, vol. 53, pp. 541–581, 2015.

[47] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme, "B97-3c: A revised low-cost variant of the b97-d density functional method," *The Journal of chemical physics*, vol. 148, no. 6, p. 064104, 2018.

[48] V. Wakelam, J. C. Loison, R. Mereau, and M. Ruaud, "Binding energies: New values and impact on the efficiency of chemical desorption," *Molecular Astrophysics*, vol. 6, pp. 22–35, 2017.

[49] A. Das, M. Sil, P. Gorai, S. K. Chakrabarti, and J. C. Loison, "An Approach to Estimate the Binding Energy of Interstellar Species," *The Astrophysical Journal Supplement Series*, vol. 237, p. 9, jul 2018.

[50] C. Dominik, C. Ceccarelli, D. Hollenbach, and M. Kaufman, "Gas-Phase Water in the Surface Layer of Protoplanetary Disks," *The Astrophysical Journal*, vol. 635, pp. L85–L88, dec 2005.

[51] Y.-J. Kuan, S. B. Charnley, H.-C. Huang, Z. Kisiel, P. Ehrenfreund, W.-L. Tseng, and C.-H. Yan, "Searches for interstellar molecules of potential prebiotic importance," *Advances in Space Research*, vol. 33, pp. 31–39, jan 2004.

[52] N. Balucani, "Elementary Reactions and Their Role in Gas-Phase Prebiotic Chemistry," *International Journal of Molecular Sciences*, vol. 10, pp. 2304–2335, may 2009.

[53] J. A. Ball, C. A. Gottlieb, A. E. Lilley, and H. E. Radford, "Detection of Methyl Alcohol in Sagittarius," *The Astrophysical Journal*, vol. 162, p. L203, dec 1970.

[54] R. J. A. Grim, F. Baas, T. R. Geballe, J. M. Greenberg, and W. A. Schutte, "Detection of solid methanol toward W 33A.," *Astronomy & Astrophysics*, vol. 243, p. 473, Mar. 1991.

[55] D. Bockelée-Morvan, P. Colom, J. Crovisier, D. Despois, and G. Paubert, "Microwave detection of hydrogen sulphide and methanol in comet Austin (1989c1)," *Nature*, vol. 350, pp. 318–320, mar 1991.

[56] S. E. Bisschop, J. K. Jørgensen, E. F. van Dishoeck, and E. B. M. de Wachter, "Testing grain-surface chemistry in massive hot-core regions," *Astronomy & Astrophysics*, vol. 465, pp. 913–929, apr 2007.

[57] S. Bottinelli, A. C. A. Boogert, J. Bouwman, M. Beckwith, E. F. van Dishoeck, K. I. Öberg, K. M. Pontoppidan, H. Linnartz, G. A. Blake, N. J. Evans, and F. Lahuis, "THE c2d SPITZER SPECTROSCOPIC

SURVEY OF ICES AROUND LOW-MASS YOUNG STELLAR OBJECTS. IV. $NH_3$ AND $CH_3OH$," *The Astrophysical Journal*, vol. 718, pp. 1100–1117, aug 2010.

[58] V. Taquet, A. López-Sepulcre, C. Ceccarelli, R. Neri, C. Kahane, and S. B. Charnley, "CONSTRAINING THE ABUNDANCES OF COMPLEX ORGANICS IN THE INNER REGIONS OF SOLAR-TYPE PROTOSTARS," *The Astrophysical Journal*, vol. 804, p. 81, may 2015.

[59] P. Boduch, R. Brunetto, J. Ding, A. Domaracka, Z. Kaňuchová, M. Palumbo, H. Rothard, and G. Strazzulla, "Ion processing of ices and the origin of $SO_2$ and $O_3$ on the icy surfaces of the icy jovian satellites," *Icarus*, vol. 277, pp. 424–432, oct 2016.

[60] C. Walsh, R. A. Loomis, K. I. Öberg, M. Kama, M. L. R. van 't Hoff, T. J. Millar, Y. Aikawa, E. Herbst, S. L. Widicus Weaver, and H. Nomura, "FIRST DETECTION OF GAS-PHASE METHANOL IN A PROTOPLANETARY DISK," *The Astrophysical Journal*, vol. 823, p. L10, may 2016.

[61] P. Friberg, S. C. Madden, A. Hjalmarson, and W. M. Irvine, "Methanol in dark clouds.," *Astronomy & Astrophysics*, vol. 195, pp. 281–289, Apr. 1988.

[62] I. W. M. Smith, E. Herbst, and Q. Chang, "Rapid neutral-neutral reactions at low temperatures: a new network and first results for TMC-1," *Monthly Notices of the Royal Astronomical Society*, vol. 350, pp. 323–330, may 2004.

[63] S. Maret, C. Ceccarelli, A. G. G. M. Tielens, E. Caux, B. Lefloch, A. Faure, A. Castets, and D. R. Flower, "$CH_3OH$ abundance in low mass protostars," *Astronomy & Astrophysics*, vol. 442, pp. 527–538, nov 2005.

[64] J. K. Jørgensen, F. L. Schöier, and E. F. Van Dishoeck, "$H_2CO$ and $CH_3OH$ abundances in the envelopes around low-mass protostars," *Astronomy & Astrophysics*, vol. 437, pp. 501–515, jul 2005.

[65] K. M. Menten, C. M. Walmsley, C. Henkel, and T. L. Wilson, "Methanol in the Orion region. I. Millimeter-wave observations.," *Astronomy & Astrophysics*, vol. 198, pp. 253–266, June 1988.

[66] F. F. S. van der Tak, E. F. van Dishoeck, and P. Caselli, "Abundance profiles of $CH_3OH$ and $H_2CO$ toward massive young stars as tests of gas-grain chemical models," *Astronomy & Astrophysics*, vol. 361, pp. 327–339, Sept. 2000.

[67] R. Garrod, I. Hee Park, P. Caselli, and E. Herbst, "Are gas-phase models of interstellar chemistry tenable? The case of methanol," *Faraday Discussions*, vol. 133, pp. 51–62, nov 2006.

[68] W. D. Geppert, M. Hamberg, R. D. Thomas, F. Österdahl, F. Hellberg, V. Zhaunerchyk, A. Ehlerding, T. J. Millar, H. Roberts, J. Semaniak, M. af Ugglas, A. Källberg, A. Simonsson, M. Kaminska, and M. Larsson, "Dissociative recombination of protonated methanol," *Faraday Discuss.*, vol. 133, pp. 177–190, nov 2006.

[69] S. B. Charnley, M. E. Kress, A. G. G. M. Tielens, and T. J. Millar, "Interstellar Alcohols," *The Astrophysical Journal*, vol. 448, p. 232, jul 1995.

[70] R. Hudson and M. Moore, "Laboratory Studies of the Formation of Methanol and Other Organic Molecules by Water+Carbon Monoxide Radiolysis: Relevance to Comets, Icy Satellites, and Interstellar Ices," *Icarus*, vol. 140, pp. 451–461, aug 1999.

[71] H. Hidaka, N. Watanabe, T. Shiraki, A. Nagaoka, and A. Kouchi, "Conversion of $H_2CO$ to $CH_3OH$ by Reactions of Cold Atomic Hydrogen on Ice Surfaces below 20 K," *The Astrophysical Journal*, vol. 614, pp. 1124–1131, oct 2004.

[72] S. Ioppolo, H. M. Cuppen, and H. Linnartz, "Surface formation routes of interstellar molecules: hydrogenation reactions in simple ices," *Rendiconti Lincei*, vol. 22, p. 211, sep 2011.

[73] N. F. W. Ligterink, C. Walsh, R. G. Bhuin, S. Vissapragada, J. Terwisscha van Scheltinga, and H. Linnartz, "Methanol ice co-desorption as a mechanism to explain cold methanol in the gas-phase," *Astronomy & Astrophysics*, vol. 612, p. A88, apr 2018.

[74] M. E. Palumbo, A. Castorina, and G. Strazzulla, "Ion irradiation effects on frozen methanol ($CH_3OH$)," *Astronomy and Astrophysics*, vol. 342, pp. 551–562, 01 1999.

[75] A. Serrallach, R. Meyer, and H. Günthard, "Methanol and deuterated species: Infrared data, valence force field, rotamers, and conformation," *Journal of Molecular Spectroscopy*, vol. 52, pp. 94–129, jul 1974.

[76] E. Dartois, W. Schutte, T. R. Geballe, K. Demyk, P. Ehrenfreund, and L. D'Hendecourt, "Methanol: The second most abundant ice species towards the high-mass protostars RAFGL7009S and W 33A," *A&A*, vol. 342, pp. L32–L35, 1999.

# Conclusion

This thesis is part of the Astro-Chemical origin (ACO) European project (https://aco-itn.oapd.inaf.it/project), an innovative training network (ITN) (Call: H2020-MSCA-ITN-2018) that has for goals to unveil the early history of the Solar system by studying the chemical composition of today forming Solar-like planetary systems and comparing it with that of the Solar System primitive bodies. In this context, my thesis was part of the work package WP3, targeting the understanding of the chemical processes in the exotic interstellar conditions *via ad hoc* quantum chemistry computation and laboratory experiments, in order to establish a reliable network of reactions occurring in the forming Solar-like systems.

My task in WP3 was to develop a methodology to build computational model of the icy grain mantles in order to: (i) provide a structural accurate grain mantle model in which the H-bonding features were properly represented, without resorting to conditions not present in the ISM; (ii) characterize the structural and electrostatic features of the icy grain as a function of size and growing conditions; (iii) accurately compute binding energy (BE) distributions of astrochemically relevant interstellar molecules adsorbed at the icy mantle. Since computational chemists started to model the icy grain mantles, a multitude of models have been developed using periodic (either crystalline or amorphous ice) and cluster approaches. Furthermore, the icy mantle amorphization process, needed to ensure a good agreement with the observations, was usually carried out by using a procedure that can hardly occur in the conditions of the ISM (heating at high T followed by a rapid cooling process). As for the adopted computational methods, almost all different levels have been applied, from classical force field methods to semiempirical QM methods up to DFT and post Hartree-Fock approaches. Up to now, we are not aware of any attempt to provide the Astrochemical community with a universal model of the ice mantle, coupled with a universal method to build

it within certain structural constraints. We think achieving those goals will allow a transferable know-how in the Astrochemical community, facilitating the progress in our understanding of the adsorption phenomena occurring at the icy mantle in the ISM. To fulfill those goals we propose ACO-FROST, a series of public domain Python codes, allowing everyone to build grain models of different sizes, shapes, and chemical composition, using mainly the new GFN-xTB family of accurate semiempirical quantum mechanical methods. Furthermore, ACO-FROST can also handle the delicate step of orienting a selected molecule towards the surface of the icy grain to study its adsorption features (binding energy distribution, structure, vibrational features, etc.) with an accuracy matching the CCSD(T) golden standard for quantum chemistry.

The key difference of icy models coming from ACO-FROST with respect to those of literature is the set of rules allowing their own growing. In ISM, water is synthesized *in situ* by radical-radical reactions between pre-adsorbed H atoms and incoming O atoms from the gas-phase. This process has never been simulated by quantum mechanical approach (the only one ensuring an accurate account of the bond making/breaking energy balance) due to the high cost of running AIMD on clusters of significant size. The fundamental step there, is to understand how the large energy of the newly formed OH bonds is dissipated within the grain, a process requiring long time evolution. This aspect is key in determine the features of the final grain structure, as hot spots may induce local crystallization to occur in an environment at 10 K. As we cannot afford running long AIMD, we developed a set of rules, coded in ACO-FROST, which imitate the physical processes described above, by using pre-formed $H_2O$ molecules. ACO-FROST, as described in the specific chapters of this thesis, is able to built up icy grains evolving in the growing process the amorphous nature of the ice in a natural manner, coherent with the thermodynamic conditions of the ISM. We tested the grain features on two clusters, LC (1000 $H_2O$ molecules) and SC one (200 $H_2O$ molecules). The study of the LC showed a rapid convergence of its properties toward those of bulk water, allowing to adopt the smaller SC one for the simulating the adsorption of relevant molecules. We studied $NH_3$, $H_2O$, and $CH_3OH$, three molecules relevant in the astrochemical context. Another added value of the icy grain from ACO-FROST is the unbiased location of the adsorbate at almost all available adsorption sites of the grain surface. This ensures that a BE distribution, rather than a single BE value, can be extracted from the simulation. This is essential, for instance, in the prediction of the abundance

of solid and gaseous water in protoplanetary disks, using numerical models based on BE values from quantum mechanical simulations. We found that when using only one BE, almost all water is in gaseous form at a distance of $\geq 0.3$ au, at variance with an amount of 10% of still frozen water at the same distance when using the BE distribution. This has profound consequence in the fate of water during the subsequent formation of a planetary system. Also, a BE distribution is crucial when a simulated TPD spectra has to be compared with the experimental ones.

In collaboration with Lorenzo Tinacci, a PhD student enrolled in the ACO project, the GFN2-xTB BE results were refined by using the ONIOM scheme, which combines B97D3 or B97-3c or even DLPNO-CCSD(T) for the model zone (the one including the adsorbate and a limited number of neighbours $H_2O$ molecules), while treating the real system (the whole grain plus the adsorbate) at GFN2-xTB level. We showed this procedure to be highly reliable and numerically robust and in our laboratory the application to many more molecules (including iCOM) is being carried out.

The work carried out in this thesis did not exploit the whole set of potentiality embedded in the ACO-FROST approach. One which is particular relevant in the present day is the simulation of the IR spectra of "dirty" icy grain, *i.e.* grain made up of a mixture of $H_2O$ molecules with other species known to be present at reduced percentage in the ice itself. ACO-FROST can easily built up a grain with a specific percentage of $CO_2$, just to mention an important contaminant of the ice mantle. We are studying the accuracy of GFN2-xTB to predict infrared spectra, with preliminary good feedback. Obviously, to be able to make accurate prediction of the "dirty" icy grain IR spectra would allow a comparison with the very recent experimental observation carried out by the James Webb Space Telescope (JWST)[1].

# References

[1] S. L. Grant, E. F. van Dishoeck, B. Tabone, D. Gasman, T. Henning, I. Kamp, M. Güdel, P.-O. Lagage, G. Bettoni, G. Perotti, V. Christiaens, M. Samland, A. M. Arabhavi, I. Argyriou, A. Abergel, O. Absil, D. Barrado, A. Boccaletti, J. Bouwman, A. C. o. Garatti, V. Geers, A. M. Glauser, R. Guadarrama, H. Jang, J. Kanwar, F. Lahuis, M. Morales-Calderón, M. Mueller, C. Nehmé, G. Olofsson, E. Pantin, N. Pawellek, T. P. Ray, D. Rodgers-Lee, S. Scheithauer, J. Schreiber, K. Schwarz, M. Temmink, B. Vandenbussche, M. Vlasblom, L. B. F. M. Waters, G. Wright, L. Colina, T. R. Greve, K. Justannont, and G. Östlin, "Minds. the detection of $^{13}CO_2$ with jwst-miri indicates abundant $CO_2$ in a protoplanetary disk," 2022.

# List of publications

## Main author

1. A. Germain and P. Ugliengo, "Modeling Interstellar Amorphous Solid Water Grains by Tight-Binding Based Methods: Comparison Between GFN-XTB and CCSD(T) Results for Water Clusters," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12253 LNCS, pp. 745–753, Springer, Cham, Jul. 2020.
DOI: 10.1007/978-3-030-58814-4_62

2. A. Germain, M. Corno, and P. Ugliengo, "Computing Binding Energies of Interstellar Molecules by Semiempirical Quantum Methods: Comparison Between DFT and GFN2 on Crystalline Ice," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12953 LNCS, pp. 632-645, Springer, Cham, sep 2021.
DOI: 10.1007/978-3-030-86976-2_43

3. A. Germain, L. Tinacci, S. Pantaleone, C. Ceccarelli, and P. Ugliengo, "Computer Generated Realistic Interstellar Icy Grain Models: Physicochemical Properties and Interaction with $NH_3$," *ACS Earth and Space Chemistry*, vol. 6, pp. 1286–1298, May 2022.
DOI: 10.1021/acsearthspacechem.2c00004

## Co-author

1. A. Rimola, S. Ferrero, A. Germain, M. Corno, and P. Ugliengo, "Computational surface modelling of ICES and minerals of interstellar in-

terest—insights and perspectives," *Minerals*, vol. 11, no. 1, pp. 1–25, 2021.
DOI: 10.3390/min11010026

2. L. Tinacci, A. Germain, S. Pantaleone, S. Ferrero, C. Ceccarelli, and P. Ugliengo, "Theoretical Distribution of the Ammonia Binding Energy at Interstellar Icy Grains: A New Computational Framework," *ACS Earth and Space Chemistry*, vol. 6, pp. 1514–1526, Jun. 2022.
DOI: 10.1021/acsearthspacechem.2c00040

3. L. Tinacci, A. Germain, S. Pantaleone, C. Ceccarelli, N. Balucani, and P. Ugliengo, "Theoretical water binding energy distribution and snow-line in protoplanetary disks," submitted to *The Astrophysical Journal* Dec. 2022.

# Appendix

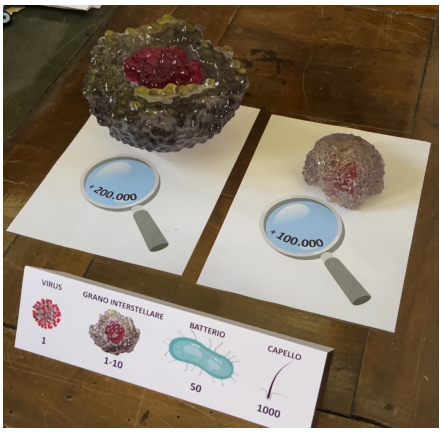
## ACO-VR project: The journey of water

During my thesis, I collaborated with my ACO colleagues to organise and build an astrochemical virtual reality (VR) project about the journey of water to Earth. The project was created for outreach purposes, and has already been displayed at the open days of the Torino University's Department of chemistry (May 2022), the UNIGHT (United citizens for research - Notte Europea delle Ricercatrici e dei Ricercatori) 2022 in Torino (October 2022), and at the Genoa Science Festival (October 2022) [1]. The project will also be presented at the ACO Congress in Toulouse (June 2023).

In the following section, I present the project, which includes its goals, its impact, and the different steps of the VR experience. The latter is offered for free on the Oculus AppLab store under the name "ACO Water Experience", meanwhile the source code can be found at the following link: https://github.com/Marcus-Keil/ACO_VR_Project. Other outreach products can be found in the ACO website (https://aco-itn.oapd.inaf.it/outreach-products). I have also, together with Lorenzo Tinacci, another PhD involved in the ACO project, built a real model of an icy grain rendered as a 3D printed model. The core of the grain is detachable to highlight the fine structure of a grain (see videos here and here). In addition, we prepared some drawings highlighting important molecules in the astrochemical context to be built by using SNATOMS (https://snatoms.com) a very nice magnetically based molecular builder, easily used also by children. Some of the activities in which ACO students have participated are illustrated in Figure 5.32.

[1] The INAF outreach group upload a youtube video on our project, at the following link: https://www.youtube.com/watch?v=vgziBE3QOVg.

**(a)** Torino University's Department of chemistry open days (May 2022).




**(b)** Torino University's Department of chemistry open days (May 2022).

**(c)** Torino UNIGHT (October 2022).

**(d)** Genoa Science Festival (October 2022).

**Figure 5.32:** Some of the activities carried out during several ACO outreach activities.

# The journey of water to Earth

*Do you know where our water originally came from?*
*Water is an essential part of life, but where it comes from is still a mystery.*
*This is one of the puzzles that Astrochemistry is trying to solve.*
*With this VR experience, we will walk you through our origins and try to solve this riddle.*

### Context

Astrochemistry is a relatively new science, characterized by its highly interdisciplinary nature. Sciences that are often perceived to be separate from one another collaborate in order to achieve a common goal: understanding the formation of the Universe using chemistry.

The public does not really know about the existence of this science; it cannot be part of the school curriculum due to its cutting edge nature. Our goal as ACO students is to shed light on astrochemistry and to fill the gap between society and this science. Science festivals and similar events need more modern content and methods of outreach.

Virtual Reality is one of the most powerful and recent outreach tools that is growing in interest around the world. As part of the ACO work agreement, it had been decided to conceive and develop an astrochemical VR experience.

Among the astronomical fields, astrochemistry is among the newest. Thus, VR outreach projects like the one proposed here are very rare. In this way, we aim to enlighten the public about astrochemistry, its importance and its goals.

### Goals

- Make people aware of the existence of the ACO project.
- Introduce the audience to our scientific field: astrochemistry.
- Explain, simply and effectively, the cosmic history of water.
- Simply explain the process of formation of a Sun-like star system.
- Trigger the curiosity of the target audience towards the subject in question.

- Enable the audience to experience VR.
- Introduce the audience to the different disciplines within astrochemistry.

### Objectives and Impact

The main target audience will be: young people (for Europe) 15-34 years old that we can interact with at big science festivals, during proposed school activities and other public occasions.

A second level target audience will be scientists - peers - that will be targeted during science congresses. Our qualitative impact studies will focus on our major actions, namely participating in science festivals, European Researchers Night (ERN) and presenting the final product in conferences and congresses, as well as in secondary schools. These fields represent a range of event types and groups of people and as such need to be treated differently in terms of impact measuring.

The events are divided in two categories: firstly, the "drop-in" events gathering an important number of people, like science festivals and ERN; secondly, the presentation of the project to a limited number of people (a few dozen) in specific environments such as congresses, conferences and secondary schools. For the second event type, the groups of people targeted is also important, since we expect the impact made on those with high school-level education to be different compared to those at a higher education level.

### Description of the final product

The participant will arrive at the ACO corner at the scientific festivals. She/he will find a desk with ACO flyers and gadgets required for the VR experience. The participant will be welcomed by one of the ESRs. The ESRs will explain the experience and will help the participant to put on the Oculus and to start the VR journey. At the same time, while the participant is playing, other members of the audience will be able to see the experience on a big monitor (also in the ACO corner).

The VR experience consists of three different activities with short explanatory videos. It will be in English. The different steps of the VR experience will be as follows:

1. When the participant puts on the Oculus, she/he will choose one of the available languages in subtitles. An introductory video will explain the context of the upcoming experience via an avatar. *Estimated duration of the video: maximum 1.5 minutes.*

2. First activity: We start with the formation of water, as a VR activity on a first stage. Here, the user can build up the water molecules, within a molecular cloud, as an interactive experience. Pop-ups to show controls during this stage. *Estimated duration of the experience: 3 minutes maximum.*

3. Then, an intermediate video plays, showing how the water molecules are accreted into the ices, along with the evolution of the star. This second video will be focused on the build-up of amorphous water ices. This will range in time from the collapse of the prestellar core until the formation of a protoplanetary disk. *Estimated duration of the video: 30-60 seconds.*

4. Second activity: In this activity, the user will build the "small rocky bodies" together. Here, gravitational forces start to play an important role, with this part of the experience dedicated to emphasizing this concept and bringing the experience closer to reality. *Estimated duration of the experience: 3 minutes maximum.*

5. After this, a third video will show a zoomed-out view of the grains with the ice mantles formed on them (another minute or so), plus a time-lapse of the evolution of the stellar system with the bodies formed, from the protoplanetary disk phase until the formation of a young version of a Sun-like stellar system. *Estimated duration of the video: 1 minute maximum.*

6. A final interactive VR activity will focus on throwing the assembled bodies towards a dry-looking Earth, to create the oceans and explain the context. *Estimated duration of the experience: 2-3 minutes.*

I would like to thank my supervisor, Prof. Piero Ugliengo, without which this thesis would be a lot less legible.
Thanks also to Dr. Stefano Pantaleone, always there to help students in need and therefore dramatically overworked.
And special thanks to the entire ACO project, it was a wonderful experience filled with great memories. COVID did not allow us to live it to its full potential, but we did our best.

This thesis has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N°811312 for the project "Astro-Chemical Origins" (ACO).


Images of 3D molecule rendering were made with VMD. VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign.

This thesis was written using LaTeX and Overleaf.