# Quantum and anharmonic effects in hydrogen-bonded crystals, surfaces and biological systems

Erika Fallacara

# PhD Thesis

Doctoral School 397

Physics and Chemistry of Materials

# Quantum and anharmonic effects in hydrogen-bonded crystals, surfaces and biological systems

To obtain the title of

PhD of Sorbonne Université

defended by

**Erika Fallacara**

on December 9th, 2022

---

*Prepared at* Institut des NanoSciences de Paris
*Under the supervision of* Fabio Finocchi and Michele Ceotto

**Jury**

| | | |
|---|---|---|
| *Reviewers* | Gloria Tabacchi | Università degli Studi dell'Insubria |
| | Marie-Pierre Gaigeot | Université d'Evry Val D'Essonne, Université Paris Saclay |
| *Examiners* | Rodolphe Vuilleumier | Sorbonne Université (Jury's President) |
| | Massimo Marchi | CEA Saclay |
| *Advisors* | Fabio Finocchi | INSP, CNRS, Sorbonne Université |
| | Michele Ceotto | Università degli Studi di Milano |

# ABSTRACT

Hydrogen bonds are weak intermolecular interactions that deeply perturb the chemical environment of atomic nuclei and contribute to the anharmonicity of the potential energy surfaces. Furthermore, they involve the hydrogen atom, which, even at ambient conditions, can manifest non negligible quantum properties due to its small mass. Both the chemical structure and the dynamics of hydrogen-bonded systems are influenced by the anharmonicity and the quantum nature of the nuclei. In this thesis we investigate different molecular systems containing hydrogen bonds by describing the nuclei by approximated quantum methods that overcome the classical and harmonic pictures.

The first system investigated is a phase transition in the solid state. Crystalline potassium hydroxide is characterized by the presence of weak hydrogen bonds and undergoes an order-disorder phase transition. The same transition happens in the deuterated crystal but the Curie temperature shifts up about 24 K due to the different proton and deuteron delocalization. Furthermore, a geometric H/D isotope effect characterizes the hydrogen bonds and it relates to the structural properties of the system.

The second study is about the adsorption of an organic molecule on an oxide surface. Formic acid is the simplest carboxylic acid and a promising hydrogen carrier material. Its adsorption on the $TiO_2$ anatase (101) surface presents competing adsorption configurations that are still debated. The molecular monodentate type of adsorption is characterized by the presence of a strong hydrogen bond, that stabilizes the molecule-surface interaction and make the proton shuttle between its stable position on the formic acid molecule and the surface.

The third study is about a complex biomolecular system. A crucial example of the importance of hydrogen bonds in biological systems is the pairing of nucleobases in DNA, where the hydrogen bonds contribute to the stability of the DNA double helix. The Watson and Crick conformation of guanine and cytosine dimer presents three hydrogen bonds. The proton delocalization is relevant even at 300 K in the gas-phase dimer. Simulated vibrational spectra were obtained.

# RÉSUMÉ

Les liaisons hydrogène sont des interactions intermoléculaires faibles qui perturbent profondément l'environnement chimique des noyaux atomiques et contribuent à l'anharmonicité des surfaces d'énergie potentielle. De plus, ils impliquent l'atome d'hydrogène, qui, même dans des conditions ambiantes, peut manifester des propriétés quantiques non négligeables en raison de sa faible masse. La structure chimique et la dynamique des systèmes avec liaison hydrogène sont influencées par l'anharmonicité et la nature quantique des noyaux. Dans cette thèse, nous étudions différents systèmes moléculaires contenant des liaisons hydrogène en décrivant les noyaux par des méthodes quantiques approchées qui vont au-delà de l'image classique et harmonique.

Le premier système étudié est une transition de phase à l'état solide. L'hydroxyde de potassium cristallin est caractérisé par la présence de liaisons hydrogène faibles et subit une transition de phase ordre-désordre. La même transition se produit dans le cristal deutéré mais la température de Curie augmente d'environ 24 K en raison de la délocalisation différente du proton et du deutéron. De plus, un effet isotopique H/D géométrique caractérise les liaisons hydrogène et est lié aux propriétés structurales du système.

La deuxième étude porte sur l'adsorption d'une molécule organique sur une surface d'oxyde. L'acide formique est l'acide carboxylique le plus simple et un matériau porteur d'hydrogène prometteur. Son adsorption sur la surface $TiO_2$ anatase (101) présente des configurations d'adsorption concurrentes qui font encore débat. Le type d'adsorption moléculaire monodentate est caractérisé par la présence d'une forte liaison hydrogène, qui stabilise l'interaction molécule-surface et font faire la navette au proton entre sa position stable sur la molécule d'acide formique et la surface.

La troisième étude porte sur un système biomoléculaire complexe. Un exemple crucial de l'importance des liaisons hydrogène dans les systèmes biologiques est l'appariement des nucléobases dans l'ADN, où les liaisons hydrogène contribuent à la stabilité de la double hélice de l'ADN. La conformation Watson et Crick du dimère de guanine et de cytosine présente trois liaisons hydrogène. La délocalisation du proton est pertinente même à 300 K dans le dimère en phase gazeuse. Des spectra vibrationnels simulés on été obtenus.

# ACKNOWLEDGEMENTS

First, I am deeply grateful to my supervisors, Fabio Finocchi and Michele Ceotto, who embraced the idea of supervising together this thesis more than three years ago. Their encouragement and scientific guidance has been fundamental throughout all my research work and the preparation of this thesis. During my PhD, I had also two essential unofficial mentors, Philippe Depondt and Simon Huppert. My heartful thanks to them for their precious advises, the time and the dedication spent to teach me. I need to thank also all the present and former members of team Oxydes at INSP, in particular Slavica Stankic for our close and stimulating collaboration and Thomas Plé for his punctual technical help. Many thanks also to my collaborators from University of Milan, Marco Cazzaniga and Giacomo Botti, for the enriching discussions about semiclassical spectroscopy.

I would like to express my gratitude to the director of INSP, Massimiliano Marangolo, and the director of the doctoral school 397, Nadine Witkowski, for giving me the opportunity of pursuing this PhD. I am also thankful to the administrative staff of INSP and the doctoral school 397, in particular Cécile Lefebvre, Myriam Melois and Hakima Si Bachir for their help and kindness.

Proceeding in the list, I would like to express how much I am grateful to all the awesome people I connected with during these years.

I cannot imagine how my PhD would have been without my office mate Niccolò, who has made my daily routines the best I could ask for. I will always remember with a smile all our funny moments and eternal discussions at work and be forever grateful for cheering me up in the most difficult times. A special thanks to my other companions during this journey, Silvia, Luis and Federica for their constant support and friendship. I would like also to thank the PhD students of the lab for all the nice shared lunch breaks, beers, raclette and pizza dinners, Lune, Benoît, Angeline, Brieuc, Perla, Oliviero, Angus, Thibaut, Sarah, Thomas and Marine.

I need to thank also all my friends met at Cité Universitaire, where I spent several confinement periods at the beginning of the pandemic. Thanks to Mehdi, Ikenna, Dylan, Rutger, Massi, Aphrodite, Clara, Naomi, Tommaso, Anna, Martin, Abdallah, Kostas, Enrico and Sofia and many others for making me feel home when I needed it the most.

Moving abroad has inevitably driven me far from my dearest ones. I have to thank my closest friends in Milan for their long lasting friendship throughout all these years, to name just a few Stella, Nadia, Ilaria, Luca and Valentina. And I have simply no words to thank my parents and my brother Andrea for always believing in me and support me in all the decisions I have made, despite being far apart during difficult times.

Completing a PhD has been one of toughest journey of my life. There is one special person, who has accompanied me throughout all this complicated journey from day zero, my boyfriend Davide. To him my immense gratitude for facing my darkest moments, constantly pushing me to never give up and be the best version of myself.

# CONTENTS

# INTRODUCTION

Hydrogen bond (A–H$\cdots$B) is a non-covalent interaction between a hydrogen atom, covalently bonded with atom A, and the B atom, which has a lone pair of electrons. Atoms A and B, that are called the donor and acceptor respectively, are electronegative atoms such as fluorine, nitrogen or oxygen; their typical reciprocal distance is in the range 2.5-3.3 Å. Hydrogen bond generally occurs between separate molecules (intermolecular), but it can also be found in the same molecule (intramolecular). Within a binding energy ranging from 4 to 50 kJ/mol [1], it is weaker than the ionic and covalent bond but stronger than the van der Waals dispersion forces, playing a main role in the stability of diverse chemical systems, from condensed matter to biomolecules. For instance, it is responsible for making the density of ice lower than that of water and indispensable in the DNA building blocks of life. However, it is also present in inorganic matter, for example in the hydrous minerals that compose the Earth's mantle, as brucite.

Both structural and dynamical properties of hydrogen-bonded systems depend on the nature of the hydrogen bond and the potential energy surface experienced by the atomic nuclei. The latter are often treated in numerical simulations within the harmonic and classical approximations. However, the harmonic and classical description of hydrogen bonds is often not adequate. First of all, it is well known that hydrogen bonds make the potential energy surfaces of the protons rather anharmonic and result in an enforced coupling between the different vibrational modes, which is totally neglected within the harmonic approximation. A first way to overcome the harmonic approximation is via classical molecular dynamics simulations, that include anharmonicity, temperature and dynamical effects. However, in molecular dynamics one solves the Newton's equations of motions where the nuclei are treated classically. As we will show in the following, this approximation is often not accurate enough.

It is common belief that quantum effects are important only at very low temperatures. However, due to its light mass, the proton can manifest a high degree of quantum delocalization even at room temperature. Indeed, at 300 K the thermal de Broglie wavelength of the proton is approximately 1 Å, the same length scale of the inter-atomic distances such as covalent and hydrogen bonds. The physico-chemical properties and the dynamics of hydrogen and deuterium nuclei can be strongly impacted by nuclear quantum effects (NQEs). They include zero-point energy fluctuations, tunneling, enhanced anharmonicity, large sensitivity to isotope substitution, that cannot be described in

terms of classical mechanics. Due to its intrinsic quantum nature, the hydrogen atom explores more the anharmonic part of the potential energy surface, resulting in a enhanced anharmonicity, which can be measured in the frequencies and the shape of the signals. When hydrogen is substituted with deuterium, the lower zero-point energy of the latter will drive the system to explore a smaller part of the potential energy surface. This effect is entirely absent in classical mechanics. More generally, the H/D isotope effect can affect also other properties, as kinetic rates and the geometry, and isotopic substitution is a fundamental tool to have a direct comparison with the experiments since they are frequently performed in the presence of deuterated species.

In the recent years, NQEs have stimulated a growing interest in a large variety of phenonema [2–6], which often involve hydrogen-bonded systems. It is now clear that behind the hydrogen bond, a so much important interaction in nature, there is a great complexity. Let us consider one of the most important systems containing hydrogen bonds, water. Water has a phase diagram that is strongly affected by NQEs [7–10] and it is a prime example. For instance, the phase transition in ice at $\simeq$ 65 GPa from phase VII to phase X, which involves the symmetrization of O–H$\cdots$O hydrogen bonds, is completely missed by classical mechanics, which predicts a transition pressure of about 100 GPa [10]. On the contrary the inclusion of NQEs is able to recover the correct description of the phase transition. A less striking effect in the physical properties of water is the isotope effect on its melting point: it shifts up to 3.81°C in heavy water, $D_2O$, and to 4.48°C in tritiated water, $T_2O$ [11].

NQEs are not easy to predict and they strongly depend on the balance between the quantum wave-particle duality of the hydrogen atom and the localization induced by the potential, which is in turn affected by the anharmonicity. This thesis is devoted to the investigation of quantum and anharmonic effects in complex chemical systems presenting hydrogen bonds. The interplay between quantum delocalization of the proton and the atomic confinement is very different for each investigated system, that we will charecterize in detail. For an adequate description of the nuclear motion, we used advanced approximated quantum methods within the Born-Oppenheimer separation between the electronic and nuclear degrees of freedom using either *ab initio* or force fields approaches for the electronic structure. For the calculation of time-independent properties we used mostly path integral molecular dynamics [12] methods, which are based on the formulation of quantum mechanics by Feynman. As an alternative, quantum baths approaches via Langevin thermostat with colored noise can be used to reproduce the statistical quantum properties [13, 14]. To have a direct comparison with the experiments, we calculated vibrational spectra through semiclassical methods based on the initial value representation theory [15], which address both anharmonic and quantum effects in the vibrational spectra.

The thesis is organized as follows:

- In the first part we will focus on the theoretical background of the employed methods and computational approaches.

  In chapter 1, we will provide a description of the classical treatment of the nuclei as a standard procedure in molecular simulations. We will firstly recall the Born-Oppenheimer approximation and the concept of the potential energy surface, then we will briefly describe the density funtional theory and the molecular mechanics approaches and discuss the harmonic

*vs* the anharmonic picture from a classical point of view.

In chapter 2, we will present some of the advanced methods that incorporate a quantum description of the nuclei starting from the Feynman formulation of quantum mechanics and proceed to semiclassical techniques applied to vibrational spectroscopy and quantum bath approaches.

- In the second part we will investigate selected hydrogen-bonded systems.

  In chapter 3, we will discuss a quantum-driven phase transition, the antiferroelectric to paraelectric phase transition in KOH and KOD crystals. The phase transition was experimentally characterized by a H/D isotope effect on the Curie temperature, but in the absence of an explanation of the phase transition mechanism. We will explain in details the mechanism of the phase transition and show how thermal and H/D isotope effects affect the structural stability of the crystal.

  In chapter 4, we will present the case of formic acid adsorption on the $TiO_2$ anatase (101) surface, which has still not reached a general consensus from both experimental and theoretical studies due to the presence of competing adsorption configurations. By combining a theoretical investigation with infrared experiments on anatase nanopowders, we will characterize the different adsorption species and we will specifically focus on a type of adsorption that manifests the presence of a strong hydrogen bond between the molecule and the surface.

  Lastly, in chapter 5 we will present a study concerning the guanine and cytosine (GC) base pair. Several studies have been focused on the proton transfer reactions along the hydrogen bonds present in the GC base pair, that are hypothesized to be the source of genetic point mutations. A lower number of studies has been published with the inclusion of the physiological conditions as well quantum and anharmonic effects. Starting from the gas-phase dimer, we will investigate the role of the solvent by analyzing both the equilibrium and vibrational properties, as well providing a benchmark of the force field employed.

# Part I

# Theoretical background

CHAPTER 1

# THE POTENTIAL ENERGY SURFACE AND CLASSICAL TREATMENT OF NUCLEI

This chapter outlines a brief and pedagogical description of the computational approaches used in the current thesis as a necessary introduction to the next chapter, where the methods for including a quantum description of the nuclei are illustrated.

## 1.1 Full quantum electron-nuclear problem

The full quantum mechanical information of a molecular system formed by $M$ nuclei with coordinates $\mathbf{R} = \{\mathbf{R}_1, \cdots, \mathbf{R}_M\}$ and $N$ electrons with coordinates $\mathbf{r} = \{\mathbf{r}_1, \cdots, \mathbf{r}_N\}$ can be obtained via the non-relativistic time-dependent Schrödinger equation (TDSE):

$$i\hbar\frac{\partial}{\partial t}\psi(\mathbf{r}, \mathbf{R}, t) = \hat{\mathcal{H}}\,\psi(\mathbf{r}, \mathbf{R}, t) \tag{1.1}$$

The electronic-nuclear Hamiltonian operator $\hat{\mathcal{H}}$ is an Hermitian operator and defines the total energy of the system. It reads as:

$$\hat{\mathcal{H}} = \hat{T}_{\text{nuc}}(\mathbf{R}) + \hat{T}_{\text{el}}(\mathbf{r}) + \hat{V}_{\text{nuc,nuc}}(\mathbf{R}) + \hat{V}_{\text{el,el}}(\mathbf{r}) + \hat{V}_{\text{nuc,el}}(\mathbf{r}, \mathbf{R}) \tag{1.2}$$

in which we have considered no spin-orbit interactions and the absence of external fields. The first and second term correspond to the kinetic energies of the nuclei and the electrons respectively, the third and the fourth terms are the nuclei-nuclei and the electrons-electrons energy potential, while the fifth term is the electrons-nuclei potential energy (eq. (1.3)).

$$\hat{T}_{\text{nuc}} = -\frac{\hbar^2}{2}\sum_{a=1}^{M}\frac{1}{m_a}\nabla_a^2 \qquad \hat{T}_{\text{el}} = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{N}\nabla_i^2$$
$$\hat{V}_{\text{nuc,nuc}} = \sum_{a=1}^{M}\sum_{b=1}^{M}\frac{Z_a Z_b e^2}{|\mathbf{R}_a - \mathbf{R}_b|} \qquad \hat{V}_{\text{el,el}} = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{1.3}$$
$$\hat{V}_{\text{nuc,el}} = -\sum_{a=1}^{M}\sum_{i=1}^{N}\frac{Z_a e^2}{|\mathbf{r}_i - \mathbf{R}_a|}$$

The Hamiltonian in eq. (1.2) does not depend on the time $t$, i.e. it describes a conservative quantum system for which the potential is not a function of the time. Thus, it is possible to make the variable separation:

$$\psi(\mathbf{r}, \mathbf{R}, t) = u(t)\psi(\mathbf{r}, \mathbf{R}) \tag{1.4}$$

which leads to the following time-independent Schrödinger equation:

$$\hat{\mathcal{H}}\,\psi(\mathbf{r}, \mathbf{R}) = E\psi(\mathbf{r}, \mathbf{R}) \tag{1.5}$$

However, eqs. (1.1) and (1.5) are second order differential equations in space domain, for which an analytical solution is impossible for many-body systems. In order to perform molecular simulations, an approximation of the Hamiltonian of eq. (1.2) and the many-body wavefunction $\psi(\mathbf{r}, \mathbf{R})$ is, therefore, needed. One way to firstly simplify the problem is to separate the electrons and the nuclear variables. However, the molecular Hamiltonian defined in eq. (1.2) is not separable due to the electron-nuclei potential energy term, which depends on both the nuclei and the electronic coordinates.

### 1.1.1 Born-Oppenheimer approximation

To tackle the problem of the non-separability of the Hamiltonian of eq. (1.2), the nuclear degrees of freedom can be considered stationary with respect to the electronic ones due to their much heavier masses ($m_{\text{nuc}} \gg m_{\text{el}}$). This latter is central in the Born-Oppenheimer (BO) approximation [16], a fundamental assumption that is used in most of electronic structure calculations and molecular dynamics simulations. Therefore the nuclear and electronic motion can be adiabatically decoupled through a separation of variables, by rewriting the wavefunction $\psi(\mathbf{r}, \mathbf{R})$ as:

$$\psi(\mathbf{r}, \mathbf{R}) = \chi_{\text{nuc}}(\mathbf{R})\psi_{\text{el}}(\mathbf{r}; \mathbf{R}) \tag{1.6}$$

where $\chi_{\text{nuc}}$ corresponds to the nuclear wavefunction and $\psi_{\text{el}}(\mathbf{r}; \mathbf{R})$ is the electronic wavefunction at a fixed nuclear configuration $\mathbf{R}$, i.e. the dependence on the nuclear coordinates is purely parametric. Since the nuclear are considered stationary, the electrons will just adapt adiabatically to a shift of the nuclear coordinates. In this way the full quantum electron-nuclear problem is recast into two problems. The first task is to solve the following electronic time-independent Schrödinger

equation.

$$\hat{\mathcal{H}}_{\mathrm{el}}\,\psi_{\mathrm{el}}(\mathbf{r};\mathbf{R}) = \varepsilon_{\mathrm{el}}(\mathbf{R})\psi_{\mathrm{el}}(\mathbf{r};\mathbf{R}) \tag{1.7}$$

where $\hat{\mathcal{H}}_{\mathrm{el}}$ is the electronic Hamiltonian operator defined as:

$$\hat{\mathcal{H}}_{\mathrm{el}} = \hat{T}_{\mathrm{el}}(\mathbf{r}) + \hat{V}_{\mathrm{el,el}}(\mathbf{r}) + \hat{V}_{\mathrm{nuc,el}}(\mathbf{r};\mathbf{R}) + \hat{V}_{\mathrm{nuc,nuc}}(\mathbf{R}) \tag{1.8}$$

and $\varepsilon_{\mathrm{el}}(\mathbf{R})$ is the eigenvalue associated to the electronic Hamiltonian and is called adiabatic potential energy surface. $\varepsilon_{\mathrm{el}}(\mathbf{R})$ describes the relation between the total energy and a particular nuclear arrangement of the system. The calculation of this surface is a crucial step for determining the properties of the chemical systems. The last part of the BO approximation regards the nuclear problem. By substitution of eq. (1.6) into eq. (1.1), then multiplication from the left by $\psi^{*}_{\mathrm{el}}(\mathbf{r};\mathbf{R})$ and integration over the electronic coordinates [17], the quantum motion of the nuclei is given by:

$$\left[\hat{T}_{\mathrm{nuc}}(\mathbf{R}) + \varepsilon_{\mathrm{el}}(\mathbf{R})\right]\chi_{\mathrm{nuc}}(\mathbf{R},t) = i\hbar\frac{\partial}{\partial t}\chi_{\mathrm{nuc}}(\mathbf{R},t) \tag{1.9}$$

Therefore, the nuclei move in an effective potential $V_{\mathrm{eff}} = \varepsilon_{\mathrm{el}}(\mathbf{R})$, which is the solution of the electronic problem, i.e. the potential energy surface at a fixed electronic state. Standard molecular dynamics, which is a powerful tool to explore the properties of a molecular system, is based on the assumption that eq. (1.9) can be approximated by classical Newton equations with the drawback of including no quantum effects for the atomic nuclei. The classical treatment of the nuclei in molecular simulations will be discussed in details in section 1.4. A more realistic picture is obtained by describing the nuclei within quantum approximations, as done in this thesis. Chapter 2 is dedicated to the illustration of these methods.

### 1.1.2 Potential energy surface

The quantum pontential energy surface (PES), which originates from the Born-Oppenheimer approximation and is specific to each electronic state[1], is an essential tool for obtaining chemical information about the system of interest, including molecular geometry, energy barriers and vibrational properties. By choosing a suitable number of nuclear configurations, it is possible to locally reconstruct the PES. This is generally a complex function that depends on one or more coordinates and presents critical points such as global/local minima and saddle points (fig. 1.1). The calculation of the PES can be done either "on-the-fly" via the computation of the electronic energies or analytically via advanced fitting scheme. An accurate description of the potential energy surface is of crucial importance in molecular simulations. However, the construction of global high-dimensional potential energy surfaces from *ab initio* calculations represent still nowadays an actual challenge.

In particular, the resolution of the many-body electrons problem of eq. (1.7) is not a trivial task due to the already mentioned impossibility to obtain a closed-form solution of a complex second order

[1] Although the BO surface is specific to each electronic state, here we will consider the ground electronic state at each atomic configuration, which is a good framework for the problems here tackled.

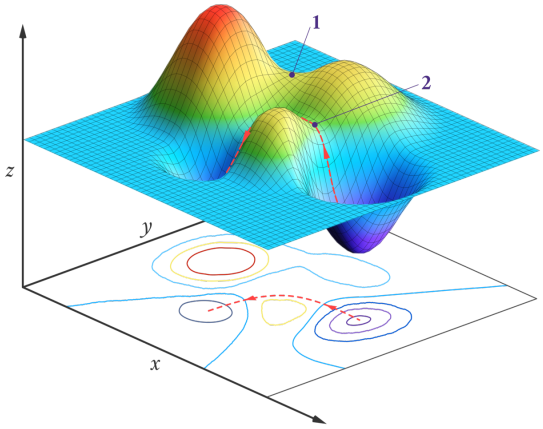



Figure 1.1: A 2D contour map and corresponding potential energy surface for a hypothetical endothermic reaction. Reproduction from ref. [18].

differential equation. Electronic structure packages deal with this task, providing approximated methods to solve the many-body electronic problem. They are called *ab initio* tecniques since they do not need, differently from other techniques such as semiempirical and molecular mechanics methods, *a priori* assumptions – for example information about the chemical bonds. There are two main approaches in the electronic structure calculations. One is wavefunction based starting from the Hartree–Fock self-consistent field procedure, the "post-Hartree-Fock" methods for the electrons correlation energy term such as Møller–Plesset perturbation, and the high level of theory methods, Configuration Interaction and Coupled Cluster, which are considered the most accurate. The other approach is density functional theory (DFT), which has been established as the most popular method for quantum electronic structure calculations of molecules and solids. We will discuss in details of DFT in the following section (see section 1.2). However, large-scale atomistic computer simulations cannot be performed via *ab initio* methods due to the too demanding computational cost. An alternative is to use interatomic potentials functions to compute the total energy and forces of the interested system. Traditional interatomic potentials are derived from a molecular mechanics approach (see section 1.3.1). In the recent years there has been a rapidly growing interest in constructing fast and accurate interatomic potentials instead via machine-learning techniques starting from *ab initio* datasets [19, 20].

## 1.2 Density functional theory

The resolution of the many-body electrons problem for molecular and condensed matter systems has been historically one of the major issues in quantum physics and chemistry. The wavefunction based methods, using variational method schemes, allow an exact numerical solution of the electronic problem, but, although very accurate, they are limited to low-dimensional systems due to the unfavourable scaling computational cost with the system size. The big turning point came in 1964, when P. Hohenberg and W. Kohn proved through two theorems [21] that the many-body electrons problem can be tackled in terms of electron density instead of wavefunctions. The electron density, which is a much more simple object than the wavefunction, can describe all the properties of the

system at the electronic ground state.[2] For a system of $N$ electrons it is:

$$n(\mathbf{r}) = N \int d\mathbf{r}_2 ... d\mathbf{r}_N \, |\psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 \tag{1.10}$$

The two theorems by Hohenberg and Kohn can be resumed in the following two points. The ground state energy of the system can be defined as a function of the electron density:[3]

$$E[n] = T[n] + U[n] + V_{\text{ext}}[n] \tag{1.11}$$

and it is a minimum that is attained when $n = n_0(r)$ among all the possible densities.

$$\min(E[n]) = E[n_0] \tag{1.12}$$

In this way DFT reduces the 3N-dimensional problem of the electronic Schrödinger equation to an equivalent 3-dimensional problem of the electron density. The ground-state energy is, in principle, obtained through a minimization scheme of the energy functional $E[n]$ at constant number of electrons $N$, i.e. by solution of the following constrained minimization problem:

$$\frac{\delta}{\delta n(\mathbf{r})} \left[ E[n] - \mu \int d\mathbf{r} \, n(\mathbf{r}) \right] = 0 \tag{1.13}$$

where $\mu$ is the Lagrange multiplier associated with the conservation of the number of electrons and corresponds to the chemical potential of the electrons system.

### 1.2.1 Kohn-Sham equations

After a year of the publication of Hohenberg and Kohn theorems, Kohn and Sham (KS) proposed a different approach to DFT [23]. Their idea was that the many-body ground-state interacting electrons system can be replaced by the ground-state of an equivalent virtual system of non-interacting electrons having the same energy density of the original system. The KS wavefunction of non-interacting electrons system is a single Slater determinant as constructed from $N$ orthonormal single-particle orbitals $\phi_j(\mathbf{r})$, that are in general not the true electronic orbitals. The electron density is therefore rewritten as:

$$n(\mathbf{r}) = \sum_j f_j |\phi_j(\mathbf{r})|^2 \tag{1.14}$$

in which $\psi_j(\mathbf{r})$ are the occupied orbitals and $f_j$ their corresponding occupation factor ($0 \leq f_j \leq 2$). The kinetic energy term can be approximated to the kinetic energy of the virtual non-interacting electrons system $T[n] \approx T_s[n]$. Solving the minimization problem of eq. (1.13) leads to the one-electron set of equations:

[2] Further information about DFT and its implementation can be found in ref. [22].

[3] $T[n] = \langle \psi_0 | \hat{T}_{\text{el}} | \psi_0 \rangle$, $U[n] = \langle \psi_0 | \hat{V}_{\text{el,el}} | \psi_0 \rangle$ and $V_{\text{ext}}[n] = \langle \psi_0 | \hat{V}_{\text{nuc,nuc}} + \hat{V}_{\text{nuc,el}} | \psi_0 \rangle$ (see the terms in eq. (1.3)) are the functionals representing the expectation values of the kinetic, electron-electron and nuclear-nuclear + electron-nuclear operators, respectively.

$$\left[\frac{\hbar^2\nabla^2}{2m} + v_{\text{eff}}(\mathbf{r})\right]\phi_j(\mathbf{r}) = \varepsilon_j\phi_j(\mathbf{r}) \tag{1.15}$$

which are known as the Kohn-Sham equations and can be solved via self-consistent iterations. $v_{\text{eff}}(\mathbf{r})$ is the optimized effective Kohn-Sham potential which reads:

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + v_{\text{xc}}(\mathbf{r}) \tag{1.16}$$

The term $v_{\text{ext}}(\mathbf{r})$ is specific for the system and corresponds to the external potential defined by the functional $V_{\text{ext}}[n]$ of eq. (1.11). The term $v_{\text{H}}(\mathbf{r})$ is the classical Coulomb electrostatic potential (Hartree energy):

$$v_{\text{H}}(\mathbf{r}) = \frac{\delta E_{\text{H}}[n]}{\delta n(\mathbf{r})} = \frac{e^2}{2}\int d\mathbf{r}'\,\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{1.17}$$

The term $v_{\text{xc}}(\mathbf{r})$ corresponds to the exchange-correlation energy:

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[n]}{\delta n(\mathbf{r})} \tag{1.18}$$

### 1.2.2 Exchange-correlation functional

The exchange-correlation functional $E_{\text{xc}}[n]$ appearing in eq. (1.18) can be formally defined as:

$$E_{\text{xc}}[n] = (T[n] - T_s[n]) + (U[n] - E_{\text{H}}[n]) \tag{1.19}$$

i.e. as the difference between the true unknown energy functional $E[n] = T[n] + U[n]$ and the known approximated functional $E^{(\text{appr.})}[n] = T_s[n] + E_{\text{H}}[n] + E_{\text{xc}}[n]$. The form of the exchange-correlation functional is at the core of DFT methods. Hundreds of approximations to $E_{\text{xc}}[n]$ are available [24, 25]. Among them, we recall the main ones, which fall into a few classes. We will limit our discussion to the approximations employed in this thesis.

Table 1.1: Classification of some DFT methods and their relative dependencies of the exchange-correlation functional.

| | |
|---|---|
| LDA | $n(\mathbf{r})$ |
| GGA | $\nabla n(\mathbf{r})$ |
| meta-GGA | $\nabla^2 n(\mathbf{r}), \tau$ |
| hybrid | $\epsilon_x$ |
| generalized RPA | $\phi_i$ |

The local Density Approximation (LDA) is the simplest approach: the exchange-correlation energy, $v_{\text{xc}}$, is approximated to the exchange-correlation energy of an homogenous electron gas (HEG), $v_{\text{xc}}^{\text{HEG}}$, having the same density at the point $\mathbf{r}$:

$$E_{\rm xc}^{\rm LDA}[n] = \int d\mathbf{r}\; v_{\rm xc}^{\rm HEG}(\mathbf{r}) n(\mathbf{r}) \tag{1.20}$$

This approximation works well for several materials but it models hydrogen bonds badly. An improvement of LDA method can be obtained by describing the electron density in term of its gradient $\nabla n(\mathbf{r})$. These are known as generalized gradient approximation (GGA) methods and belong to the family of semi-local functionals.

$$E_{xc}^{\rm GGA}[n] = E_{\rm xc}^{\rm LDA}[n] + \int d\mathbf{r}\; e_{\rm xc}^{\rm GGA}(n(\mathbf{r}))\nabla n(\mathbf{r}) \tag{1.21}$$

The function $e_{\rm xc}^{\rm GGA}$ can be expressed in many different forms. We recall the one presented by Perdew, Burke and Ernzerhof (PBE) [26], which reproduces the strength of hydrogen bonds in several systems within an accuracy of 1 kcal/mol. Thus, we employed PBE for DFT calculations in this thesis (see chapter 3, chapter 4) since we are mostly interested in the description of hydrogen bonds and not to long-range interactions or strong exchange and correlation effects. Next, meta-GGA methods, which are an extension of GGA, incorporate the Laplacian of the density $\nabla^2 n(\mathbf{r})$ or the kinetic energy density $\tau$ into the the exchange-correlation functional. Another category is that of the hybrid functionals, which combine a part of the exact exchange from Hartree-Fock and density based exchange–correlation terms via parameters obtained from either *ab initio* or empirical sources. One popular version of hybrid functionals and very accurate for molecular systems is Becke-3-parameter-Lee-Yang-Parr (B3LYP) [27].

## 1.3 Interatomic potentials

Although their accurate description for electronic degrees of freedom, *ab initio* methods are not feasible for very large systems. Standard plane-wave based DFT methods, that we employed in this thesis, scale up as $\sim N^3$ with $N$ being the number of electrons of the system. It is therefore clear that *ab initio* methods cannot be applied to large systems as biomolecular structures, which often require the inclusion of the solvent. The use of interatomic potentials – mathematical functions that yield the potential energy for a given nuclear configuration – allows the treatment of such large and complex systems, otherwise forbidden by *ab initio* approaches.

### 1.3.1 Molecular mechanics

One traditional approach to treat biological systems is via molecular mechanics (MM), which refers to a group of methods based on Newton mechanics. In a MM approach the atoms are modeled as balls (nucleus+electrons systems) held together by springs. The balls-springs system is governed by classical empirical functions constituting a so-called force field (FF):

$$U = \underbrace{U_{\rm bond} + U_{\rm angle} + U_{\rm b\theta} + U_{\rm oop} + U_{\rm torsion}}_{\rm bonded} + \underbrace{U_{\rm vdW} + U_{\rm ele}}_{\rm non-bonded} \tag{1.22}$$

The first five terms describe the short-range interactions: bond stretching, angle bending, bond-

angle cross term, out-of-plane bending and torsional rotation. The last two terms correspond to the non-bonded energy contribution given by the long-range van der Walls and electrostatic interactions. The various existing force field present different ways to calculate the above energy terms. Examples of popular force fields are AMBER [28, 29], OPLS [30] and CHARMM [31]. In all FF approaches, the energy terms are calculated empirically, i.e. they are based on set of parameters derived from either experiments or *ab initio* calculations.

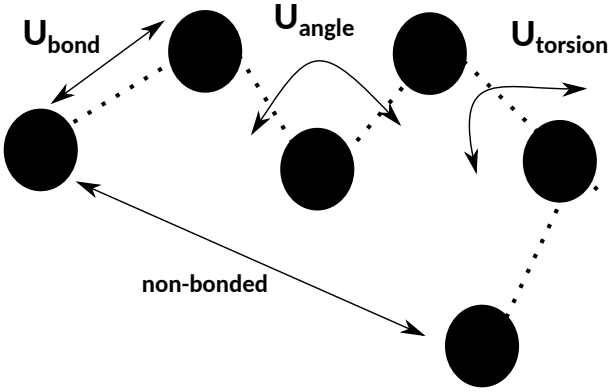



Figure 1.2: Illustration of the fundamental force field energy terms.

One of the main challenge of MM approaches is to find a good compromise between accuracy and transferability. The parametrization itself is often a time demanding procedure and sometimes includes an elaborate fitting scheme. To make the parametrization easier, a common practice is to chose the same parameters for the same functional groups, which become atom types in a FF formulation. The applicability of a FF are limited to the group of molecules for which the set of parameters are available. Furthermore, in standard force fields it is not possible to break and form chemical bonds. For these reasons, despite their significantly lower computational cost, MM approaches are less flexible and transferable than *ab initio* methods.

As concerned the level of accuracy, the most delicate term to calculate in eq. (1.22) are the non-bonded interactions. In particular, traditional force field represent the electrostatic potential in terms of fixed partial charges model, which, however, neglects the polarization. To tackle this issue, one can rely to hybrid quantum mechanics and molecular mechanics (QM/MM) approaches, which partition the total system into a classical MM region and a electronically important region treated at *ab initio* level. Instead, for a pure MM treatment, a great progress has been made in the recent years with the development of the so-called polarizable force fields. In this respect, the polarizable force field AMOEBA[4], which we employed in this thesis (see chapter 5), is largely used in biomolecular simulations. The AMOEBA model – atomic multipole optimized energetics for biomolecular applications – was introduced in 2002 [33] and has been recently developed for proteins [34] and nucleic acids [35]. In the AMOEBA model the short-ranged valence interactions of eq. (1.22) are given by:

[4] A detailed description of the energy terms of the AMOEBA force field can be found in reference [32].

$$U_{\text{bond}} = K_{\text{b}}(b - b_0)^2\Big[1 - 2.55(b - b_0) - 3.793125(b - b_0)^2\Big]$$

$$\begin{aligned} U_{\text{angle}} = \; & K_\theta(\theta - \theta_0)^2\Big[1 - 0.014(\theta - \theta_0) + 5.6 \times 10^{-5}(\theta - \theta_0)^2 \\ & -7.0 \times 10^{-7}(\theta - \theta_0)^3 + 2.2 \times 10^{-8}(\theta - \theta_0)^4\Big] \end{aligned} \tag{1.23}$$

$$U_{\text{b}\theta} = K_{\text{b}\theta}[(b - b_0) + (b' - b'_0)](\theta - \theta_0) \qquad U_{\text{oop}} = K_\chi \chi^2$$

$$U_{\text{torsion}} = \sum_n^N K_{n\phi}\Big[1 + \cos(n\phi \pm \text{phase})\Big]$$

The bond and angle energy terms differ from the traditional force fields version by including anharmonic contributions via polynomials of order $> 2$. The torsion energy term is modeled as Fourier series up to $N$-fold, e.g. $N = 3$ for proteins and nucleic acids. The van der Walls term is described by the Halgren's buffered 14-7 function [36], instead of the traditional Lennard-Jones 12-6 function:

$$U_{\text{vdW}}(ij) = \varepsilon_{ij}\left(\frac{1+\delta}{\rho_{ij}+\delta}\right)^7 + \left(\frac{1+\gamma}{\rho_{ij}^7+\gamma} - 2\right) \tag{1.24}$$

where the $\varepsilon_{ij}$ parameter is the well depth and $\rho_{ij} = (r_{ij}/r_{ij}^0)$ with $r_{ij}$ and $r_{ij}^0$ being the $i$-$j$ separation and $i$-$j$ minimum energy distance. The electrostatic energy term includes both permanent and induced multipoles contributions:

$$U_{\text{ele}} = U_{\text{ele}}^{\text{perm}} + U_{\text{ele}}^{\text{ind}} \tag{1.25}$$

$U_{\text{ele}}^{\text{perm}}$ is generated from interaction between atomic multipoles. The charge distribution is modeled by placing on each atomic center a point monopole (partial charge), a dipole vector and a quadrupole tensor. $U_{\text{ele}}^{\text{ind}}$ is the potential energy term due to dipole-induced interactions. The induced dipole of each atom is calculated as the product of its atomic polarizability and the external electrostatic field generated by the permanent multipoles and the induced dipoles of all the other atoms.

## 1.4 Classical treatment of nuclei

A simple approximation to the dynamics of the nuclei is via classical mechanics. This is the standard approach used in *ab initio* molecular dynamics (AIMD), where the electronic structure is calculated via quantum mechanical approaches but the nuclei are treated as classical particles at the ground-state electronic surface.

### 1.4.1 Classical dynamics

From the Born-Oppenheimer approximation, it is possible to the derive classical molecular dynamics by considering the nuclei as point masses. We start from the Schrödinger equation for the nuclear degrees of freedom eq. (1.9) and derive classical dynamics of the nuclei as per reference [17]. The nuclear wavefunction can be rewritten in "polar" representation:

$$\chi_{\text{nuc}}(\mathbf{R}, t) = \mathcal{A}(\mathbf{R}, t) e^{\frac{i}{\hbar}\mathcal{S}(\mathbf{R},t)} \tag{1.26}$$

in which $\mathcal{A}(\mathbf{R}, t)$ and $\mathcal{S}(\mathbf{R}, t)$ are respectively the amplitude factor and phase function and are considered to be real with $\mathcal{A}(\mathbf{R}, t) \geq 0$. In particular the function $\mathcal{S}(\mathbf{R}, t)$ corresponds to the action. By substitution of eq. (1.26) into eq. (1.9), we obtain:

$$\left[ -\frac{\hbar^2}{2} \sum_{a=1}^{M} \frac{1}{m_a} \nabla_a^2 + \varepsilon_{\text{el}}(\mathbf{R}) \right] \mathcal{A}(\mathbf{R}, t) e^{\frac{i}{\hbar}\mathcal{S}(\mathbf{R},t)} = i\hbar \frac{\partial}{\partial t} \mathcal{A}(\mathbf{R}, t) e^{\frac{i}{\hbar}\mathcal{S}(\mathbf{R},t)} \tag{1.27}$$

After separating the real and imaginary part of $\chi_{\text{nuc}}$, we obtain the following equations of motion for the nuclei:

$$\frac{\partial \mathcal{S}}{\partial t} + \sum_{a=1}^{M} \frac{1}{2m_a} (\nabla_a \mathcal{S})^2 + \varepsilon_{\text{el}} = \hbar^2 \sum_{a=1}^{M} \frac{1}{2m_a} \frac{\nabla_a^2 \mathcal{A}}{\mathcal{A}} \tag{1.28}$$

$$\frac{\partial \mathcal{A}}{\partial t} + \sum_{a=1}^{M} \frac{1}{m_a} (\nabla_a \mathcal{A})(\nabla_a \mathcal{S}) + \sum_{a=1}^{M} \frac{1}{2m_a} \mathcal{A}(\nabla_a^2 \mathcal{S}) = 0 \tag{1.29}$$

which are re-written exactly in terms of the real quantities $\mathcal{S}$ and $\mathcal{A}$ instead of $\text{Re}\chi_{\text{nuc}}$ and $\text{Im}\chi_{\text{nuc}}$. Eq. (1.29) can be rewritten by multiplying by $2\mathcal{A}$ from the left:

$$\frac{\partial \mathcal{A}^2}{\partial t} + \sum_{a=1}^{M} \frac{1}{m_a} \nabla_a (\mathcal{A}^2 \nabla_a \mathcal{S}) = 0 \tag{1.30}$$

By defining the nuclear probability density $\rho$ as $\rho = |\chi_{\text{nuc}}|^2 = \mathcal{A}^2$ and the associated current density as $\mathbf{J}_a = \mathcal{A}^2 \nabla_a \mathcal{S}$, the following continuity equation is obtained:

$$\frac{\partial \rho}{\partial t} + \sum_{a=1}^{M} \nabla_a \mathbf{J}_a = 0 \tag{1.31}$$

which is independent of $\hbar$ and ensures at each point the conservation of the particle probability density of the nuclei, $|\chi_{\text{nuc}}|^2$, in the presence of a flux. For the purposes of the discussion, we are more interested to the relation for the phase $\mathcal{S}$, eq. (1.28), whose term on the right side depends explicitly on $\hbar$ and vanished in the classical limit $\hbar \to 0$:

$$\frac{\partial \mathcal{S}}{\partial t} + \sum_{a=1}^{M} \frac{1}{2m_a} (\nabla_a \mathcal{S})^2 + \varepsilon_{\text{el}} = 0 \tag{1.32}$$

The above expression actually corresponds to the Hamilton-Jacobi equation:

$$\frac{\partial \mathcal{S}}{\partial t} + \mathcal{H}(\mathbf{R}_a, \nabla_a \mathcal{S}) = 0 \tag{1.33}$$

with $\mathcal{H}(\mathbf{R}_a, \nabla_a \mathcal{S})$ being the classical Hamilton function. By defining $\mathbf{P}_a$ the nuclear conjugate momenta of nuclear positions $\mathbf{R}_a$:

$$\mathbf{P}_a \equiv \nabla_a \mathcal{S} \tag{1.34}$$

the classical Hamiltonian can we written as:

$$\mathcal{H}(\mathbf{R}, \mathbf{P}) = T(\mathbf{P}) + V(\mathbf{R}) \tag{1.35}$$

in which $T(\mathbf{P})$ is the nuclear kinetic energy $T(\mathbf{P}) = \sum_{a=1}^{M} \frac{1}{2m_a}\mathbf{P}_a^2$ and $V(\mathbf{R})$ is the effective potential $V(\mathbf{R}) = \varepsilon_{\mathrm{el}}(\mathbf{R})$ acting on the nuclei. The classical total energy conservation ($dE_{\mathrm{tot}}/dt = 0$ ) reads as:

$$\frac{\partial \mathcal{S}}{\partial t} = -T(\mathbf{P}) - \varepsilon_{\mathrm{el}}(\mathbf{R}) = -E_{\mathrm{tot}} = \mathrm{const} \tag{1.36}$$

The Newtonian equations of motion for a nucleus $a$, $\dot{\mathbf{P}}_a = -\nabla_a V(\mathbf{R}_a)$, can be written as:

$$m_a \ddot{\mathbf{R}}_a = -\nabla_a \varepsilon_{\mathrm{el}}(\mathbf{R}_a(t)) \tag{1.37}$$

The classical motion of the nuclei is driven by the effective potential $\varepsilon_{\mathrm{el}}(\mathbf{R})$, which corresponds to the Born-Oppenheimer potential energy surface defined in eq. (1.7). The above equation can be implemented in a simulation, giving the so-called Born-Oppenheimer molecular dynamics scheme.

### 1.4.2 Static and dynamical properties

From a molecular dynamics simulation it is possible to derive many properties of the system, including both static, i.e. time-independent, and dynamical, i.e. time-dependent, ones. To do so, a statistical mechanics approach is needed. Let us consider, for simplicity, a classical system of $M$ fixed number of particles in one-dimensional case with positions $x = \{x_i\}$ and conjugate momenta $p = \{p_i\}$. The total energy is given by the classical Hamiltonian $\mathcal{H}(x, p)$.

$$\mathcal{H}(x, p) = \sum_{i=1}^{M} \frac{p_i^2}{2m_i} + V(x) \tag{1.38}$$

with $m_i$ being the mass of the $i$-esim particle.

**Static properties** Suppose we want to describe the thermodynamic equilibrium of the classical system defined by the Hamiltonian of eq. (1.38). At the thermal equilibrium the probability density is stationary. In the canonical ensemble (NVT), the thermal equilibrium will be given by the

following Boltzmann distribution:

$$\rho_{\mathrm{cl}}(x,p) = \frac{e^{-\beta\mathcal{H}(x,p)}}{\sum_i e^{-\beta E_i}} = \frac{e^{-\beta\mathcal{H}(x,p)}}{\mathcal{Z}_{\mathrm{cl}}} \tag{1.39}$$

where $\beta$ is the inverse of the thermal energy, $\beta = 1/k_B T$, and $E_i$ is the total energy of the microstate $i$. The denominator corresponds to the classical partition function of the system, which, in the continuous case, is rewritten as the phase space integral:

$$\mathcal{Z}_{\mathrm{cl}} = \int dx\, dp\, e^{-\beta\mathcal{H}(x,p)} \tag{1.40}$$

Let $A$ be the observable associated to the static property that we are interested in. An estimation of property $A$ can be evaluated through its thermal average via integration over a sufficiently large number of points weighted according to the Boltzmann distribution:

$$\langle A\rangle_{\mathrm{cl}} = \int dx\, dp\, \rho_{\mathrm{cl}}(x,p) A(x,p) \tag{1.41}$$

**Time-dependent properties** The description of dynamical phenomena requires a more difficult procedure due to the time-dependency of the process involved. Time-correlation functions (TCF) represent a tool to evaluate the dynamical properties of a system such as vibrational spectra and transport coefficients. Given two arbitrary observables $A$ and $B$ within a classical dynamic, their associated time correlation function $C_{AB}(t)$ is defined as:

$$C_{AB}(t) = \langle A(0)B(t)\rangle = \int dx\, dp\, \rho_{\mathrm{cl}} A(x,p) B(x_t, p_t) \tag{1.42}$$

TCF are real and even functions in time, therefore they satisfy:

$$C_{AB}(t) = \langle A(0)B(t)\rangle = \langle A(t)B(0)\rangle \qquad C_{AB}(t) = C_{AB}(-t) \tag{1.43}$$

In the case $A = B$, the time correlation function $C_{AA}(t)$ is said an auto-correlation function. The direct calculation of TFC from eq. (1.42) is not computationally efficient since it requires to run several trajectories from different configurations sampled from the density function $\rho_{\mathrm{cl}}$. Generally MD simulations rely on the ergodicity hypothesis, so that a single trajectory can be employed for both sampling the initial distribution and to compute the correlations. Along a single trajectory, the time correlation function for $A$ and $B$ can be calculated with:

$$C_{AB}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, A(x_t) B(x_{t+\tau}) \tag{1.44}$$

with $T$ being the total time length of the simulation.

**Wiener–Khintchine theorem** Let us consider the power spectrum or spectral density, $S_{xx}(\omega)$, of a random time series $x(t)$. $S_{xx}(\omega)$ is a measure of the power signal over frequency and it is

defines as:

$$S_{xx}(\omega) = \lim_{T\to\infty} \frac{1}{T}|\tilde{x}(\omega)|^2 \tag{1.45}$$

in which $\tilde{x}(\omega)$ is the Fourier transform of the process $x(t)$:

$$\tilde{x}(\omega) = \int_{-\infty}^{+\infty} dt\, e^{-i\omega t} x(t) \tag{1.46}$$

The Wiener–Khintchine theorem makes the connection between the spectral density $S_{xx}(\omega)$ and the autocorrelation function of the associated process, $C_{xx}(t)$:

$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} dt\, e^{-i\omega t} C_{xx}(t) \tag{1.47}$$

i.e. the autocorrelation function of $x(t)$ is the inverse Fourier transform of the spectral density. This is an import result allowing to compute the power spectrum for a specific process starting from its autocorrelation function.

### 1.4.3 Harmonic *vs* anharmonic picture

The traditional approach to study molecular vibrations and lattice dynamics is via the harmonic approximation. Firstly, critical points of the PES are identified through geometry optimization algorithms. Once identified the interested critical point, for example a local minimum, the Hessian $\partial^2 E/\partial x_i \partial x_j|_0$ is computed[5], which enters in a Taylor expansion of the potential at the second order of the displacements. Therefore, in the harmonic approximation only the curvature of the PES at the critical points is required. The harmonic eigenvectors and frequencies can be obtained by diagonalization of the Hessian, which for an equilibrium geometry we label as $\mathbf{H_{eq}}$.

$$\mathbf{W H_{eq} W^T} = \mathbf{\Lambda} \tag{1.48}$$

$\mathbf{W}$ is the matrix of change of basis with dimension $(N \times N)$ with $N$ being the n.o. degrees of freedom. Its columns $w_i$ with $i = 1, ..., N$ constitute the frequency eigenvectors.

The matrix $\mathbf{W}$ can be used for the normal modes coordinate transformation:

$$\mathbf{x} = \mathbf{W q} \tag{1.49}$$

in which $\mathbf{x}$ and $\mathbf{q}$ are respectively the Cartesian and normal modes coordinates. The normal mode eigenfunctions constitute a complete basis that can be used to described the different vibrational motion of a molecular system. We will use this representation in semiclassical spectroscopic calculations (see section 2.5.3). $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues $\lambda_i$, which are the square of the harmonic vibrational frequency for the $i$-mode, $\lambda_i = \omega_i^2$.

However, real potentials are anharmonic and this approximation is often too crude, completely ne-

[5] We remind that while using DFT methods, we do not dispose of an analytical PES.

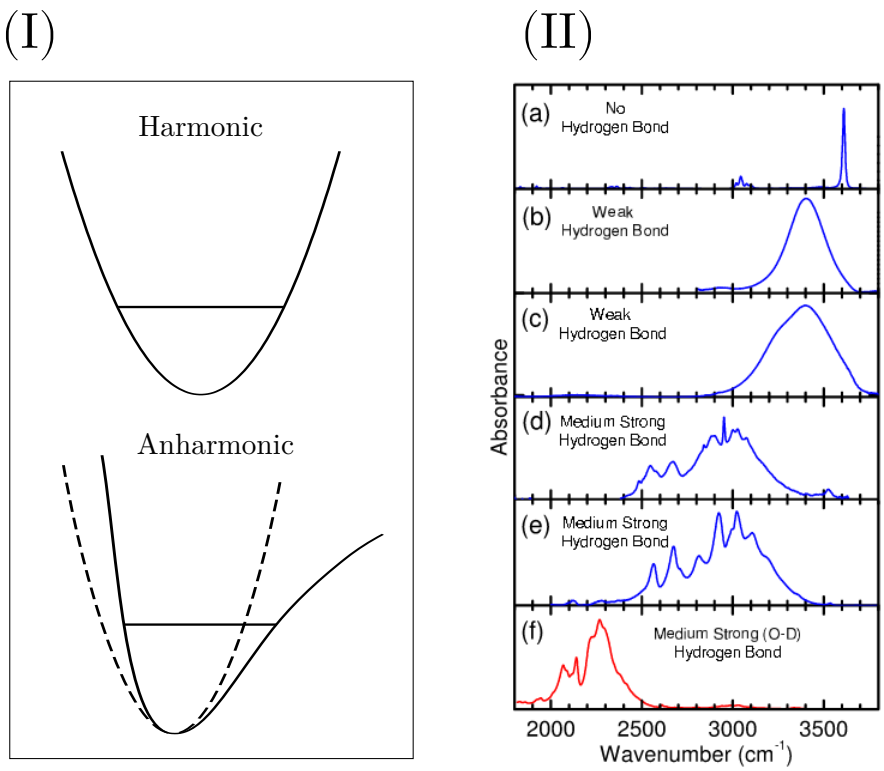



Figure 1.3: (I) Schematic representation of harmonic *vs* anharmonic potential of a diatomic molecule. Only the ground energy level is represented. (II) Infrared adsorption spectra showing O–H stretching adsorption bands of (a) uncomplexed phenol in $C_2Cl_4$, (b) the weak hydrogen bond of HOD in $D_2O$, (c) neat $H_2O$, (d) the medium strong hydrogen bond of PMME-H, (e) acetic acid dimer $(CD_3\text{-}COOH)_2$ and (f) the O-D stretching band of acetic acid dimer $(CD_3\text{-}COOD)_2$. Reproduction from ref [1].

glecting the coupling between the different modes and dynamical effects. Already for a diatomic molecule, harmonic approximation can completely fails (fig. 1.3-I) and for more complex systems is often inadequate. For example all the phenomena, in which the frequencies manifest a temperature-dependency such as diffusion, phase transitions and thermal expansion of crystals, cannot be described in terms of harmonic approximation. In particular, in the presence of hydrogen bonds the vibrational spectra are significantly affected (fig. 1.3-II). The adsorption band of the stretching of the donor group, for example a OH group, $\nu$(OH) presents respect to the "free" case[6] a red-shift and very often a spectral broadening and increased intensity.

To overcome the drawbacks of the harmonic approximation and get a more realistic picture, anharmonic contributions can be included via molecular dynamics simulations. Using the definition of eq. (1.47), one can compute the power spectrum from a molecular dynamics simulation as the Fourier transform of the velocity-velocity autocorrelation function $C_{vv}(\tau)$:

$$I(\omega) = \int_{-\infty}^{+\infty} d\tau \, e^{-i\omega\tau} C_{vv}(\tau) \tag{1.50}$$

From eq. (1.50) one can extract the relevant frequencies of the system at each temperature. The

[6] In the absence of the hydrogen bond.

above equation can be used to calculate the vibrational spectra, but with only the inclusion of anharmonicity and temperature. In the next chapter, we will discuss the methods that overcome the classical picture.

# CHAPTER 2

# FROM QUANTUM TO SEMICLASSICAL AND BATH APPROACHES

The present chapter outlines the simulation techniques used in this thesis and the formalism and approximations on which they are based. The presented methods rely on the Born-Oppenheimer approximation, as illustrated in the previous chapter, with a treatment of the electronic degrees of freedom either *ab initio* or by using fitted potential energy surfaces or force fields. These theories aim at providing an adequate quantum description of the thermodynamics and dynamics of various molecular systems, thus going beyond the classical approximation for the motion of the nuclei. At the beginning of the chapter, a short introduction of the theoretical framework around which these methods are developed is given. Next, the Feynman's path integral interpretation of quantum mechanics is presented along with the derivation of path integral molecular dynamics (PIMD) method. The third part of the chapter discusses the semiclassical approximation of Feynman's quantum propagator till the most recent developments of semiclassical initial value representation (SCIVR) theory applied to vibrational spectroscopy, followed by the quasi-classical approximation. Lastly, the quantum thermal bath (QTB) method, based on the Langevin equation of motion, is examined.

## 2.1 Quantum time evolution propagator

Let us recall the TDSE for the nuclei (eq. (1.9)). For simplicity sake, we rewrite this equation by considering just a particle with atomic mass in one dimension. The generalization to higher dimensions is straightforward. To facilitate the mathematical derivations discussed in this chapter,

This document is inspired by the following sources.
Mark Tuckerman *Statistical mechanics: theory and molecular simulation* (2010) [37].
Riccardo Conte and Michele Ceotto *Semiclassical Molecular Dynamics for Spectroscopic Calculations* (2020) [38].

we use the Dirac notation instead of wavefunctions. We indicate with $|\psi(t)\rangle$ the state vector that describes the quantum state of the particle. The nuclear TDSE is rewritten as:

$$\hat{\mathcal{H}}\,|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle \tag{2.1}$$

where the Hamiltonian is $\hat{\mathcal{H}} = \hat{p}^2/2m + V(\hat{x})$ with $m$ the mass of particle and $\hat{p}$ and $\hat{x}$ the momentum and position operator having commutation rule relationship $[\hat{p}, \hat{x}] = i\hbar$. The above equation has formal solution:

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|\psi(0)\rangle \tag{2.2}$$

The operator $e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}$ is unitary and describes the time evolution of the particle starting from a initial state $|\psi(0)\rangle$ ending to a final state $|\psi(t)\rangle$. It is known as the *quantum propagator*, $\hat{U}(t)$:

$$\hat{U}(t) = e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t} \tag{2.3}$$

It is useful to represent the quantum propagation in the coordinates basis via:

$$\psi(x', t) = \langle x'|\psi(t)\rangle = \langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|\psi(0)\rangle \tag{2.4}$$

By inserting the identity $\int dx\,|x\rangle\langle x| = \mathbb{I}$ in eq. (2.4), we obtain:

$$\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|\psi(0)\rangle = \int dx\,\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|x\rangle\langle x|\psi(0)\rangle \tag{2.5}$$

The quantity:

$$\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|x\rangle \equiv U(x', x, t) \tag{2.6}$$

is the probability amplitude for the particle to propagate from $x$ to $x'$ in a time $t$, i.e. the coordinate-space matrix elements of the quantum propagator, $U(x', x, t)$.

## 2.2 Density matrix

The conventional formulation of quantum mechanics in terms of wavefunctions is restricted to "pure states", i.e. quantum systems defined by a single state vector $|\psi\rangle$. To provide an ensemble representation in statistical mechanics, we introduce the density matrix operator formalism, which describes a statistical mixture of pure states, the so-called "mixed states". We introduce the density operator $\hat{\rho}$ for a pure state $|\psi\rangle$ as:

$$\hat{\rho} = |\psi\rangle\langle\psi| \tag{2.7}$$

The expectation value of a physical observable described by the operator $\hat{A}$ is given by:

$$\langle \hat{A} \rangle = \langle \psi | \hat{A} | \psi \rangle = \mathrm{Tr}[\hat{\rho}\hat{A}] \tag{2.8}$$

with $\mathrm{Tr}[\cdot]$ being the trace function. In a statistical ensemble a quantum state is expressed by an ensemble of microstates $|\psi_i\rangle$, which have probability $p_i$. The density operator for a mixture of pure states is:

$$\hat{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i| \tag{2.9}$$

where $\sum_i p_i = 1$. The expectation value of the $\hat{A}$ operator is therefore:

$$\langle \hat{A} \rangle = \sum_i p_i \langle \psi_i | \hat{A} | \psi_i \rangle = \sum_i p_i \, \mathrm{Tr}[\hat{\rho}_i \hat{A}] \tag{2.10}$$

The probability $p_i$ depends on the particular statistical mixture. Here, we refer to the canonical ensemble (NVT) and use as notation $\hat{\rho}(\beta)$. The statistical probabilities $p_i$ are expressed in terms of Boltzmann factor $p_i \propto e^{-\beta E_i}$.

$$\hat{\rho}(\beta) = \frac{\sum_i e^{-\beta E_i} \, |\psi_i\rangle\langle\psi_i|}{\sum_i e^{-\beta E_i}} = \frac{e^{-\beta \hat{\mathcal{H}}}}{\mathrm{Tr}[e^{-\beta \hat{\mathcal{H}}}]} \tag{2.11}$$

where the trace of $e^{-\beta \hat{\mathcal{H}}}$ corresponds to the quantum partition function $\mathcal{Z}$:

$$\mathcal{Z} = \mathrm{Tr}[e^{-\beta \hat{\mathcal{H}}}] \tag{2.12}$$

The thermal average of the observable $A$ is the expectation value of the corresponding operator $\hat{A}$ in the density matrix:

$$\langle \hat{A} \rangle = \mathrm{Tr}[\hat{\rho}\hat{A}] = \frac{1}{\mathcal{Z}} \mathrm{Tr}[e^{-\beta \hat{\mathcal{H}}} \hat{A}] \tag{2.13}$$

As we show for the time propagator operator in eq. (2.6), an equivalent representation for the canonical density operator in the coordinate basis is:

$$\rho(x', x, \beta) = \langle x' | e^{-\beta \hat{\mathcal{H}}} | x \rangle \tag{2.14}$$

## 2.3 Molecular vibrational spectroscopy

Here, we focus on the transition between the energy levels $E_n$ and $E_m$ characterized by a frequency transition $\omega_{nm}$:

$$E_n - E_m = \hbar \omega_{nm} \tag{2.15}$$

The lowest possible energy level $E_0$, i.e. the zero-point energy (ZPE) of the system, corresponds

to the ground state. Let $|\chi\rangle = \sum_n c_n|\psi_n\rangle$ a bound state wavepacket and $E_n$ the eigenvalues of the vibrational Hamiltonian defining the system. The vibrational power spectrum $I(E)$, i.e. the vibrational density of states (VDOS) at energy $E$, is given by:

$$I(E) = \sum_n |c_n|^2 \delta(E - E_n) \tag{2.16}$$

in which $\delta(E - E_n)$ is a Dirac function peaked at $(E - E_n)$. However, the calculation of the spectrum via eq. (2.16) is not straightforward since the energy eigenvalues are not easy to be determined. In the case of high-dimensional systems the vibrational density of states can be very large at low energies. Furthermore, the VDOS of crystals can display singularities. One way to calculate $I(E)$ by including anharmonicity and quantum effects is from trajectory-based methods. Indeed, it is possible to demonstrate that eq. (2.16) is equivalent to the Fourier transform of the survival amplitude of the reference state $|\chi\rangle$:

$$\begin{aligned} I(E) &= \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} dt\ \langle\chi|\chi(t)\rangle e^{\frac{i}{\hbar}Et} \\ &= \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} dt\ \langle\chi|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|\chi\rangle e^{\frac{i}{\hbar}Et} \\ &= \frac{1}{2\pi\hbar}\int_{-\infty}^{+\infty} dt\ \sum_{m,n} c_m^* c_n \langle\psi_m|\psi_n\rangle e^{\frac{i}{\hbar}(E-E_n)t} \\ &= \sum_{m,n} \delta_{mn}\delta(E - E_n) \\ &= \sum_n |c_n|^2 \delta(E - E_n) \end{aligned} \tag{2.17}$$

By using the first relation of the above equation is possible to evaluate numerically the power spectrum along several trajectories, which will be characterized by peaks located at the eigenenergies of the system. The power spectrum contains the information from all the vibrational eigenvalues. The information that is obtained from the power spectrum cannot be straightforwardly compared to experimental spectra. Indeed, each technique is characterized by a specific cross section, and the corresponding matrix elements have distinct expression. For example, in the case of an IR transition, the intensities $I_{abs}(E)$ are given by:

$$I_{abs}(E) \propto |\langle n|\hat{\mu}|m\rangle|^2 \tag{2.18}$$

where $\hat{\mu}$ is the dipole moment operator for the IR transition from vibrational state $n$ to $m$.

## 2.4 Feynman path integral formalism

Inspired by the work of P.A.M. Dirac [39], Richard Feynman introduced in 1948 a reformulation of quantum mechanics based on path integrals [40], later formalized in the book written with his former student Al. Hibbs in 1965 [41]. The underlying idea of path integrals (PI) is to incorpo-

rate the principle of least action of classical mechanics into quantum mechanics by recasting the quantum propagation as a sum over all possible paths that a particle can take between two points weighted by their respective action. In this section, the different interpretations of Feynman formalism are discussed. The so-called real-time representation gives the path integral version of the quantum propagator. In the continous-time limit, it is possible to derive a functional integral form, in which the weight of each path over the action appears. Finally, the imaginary-time representation introduces a link between statistical physics and the path integral, which can then be numerically implemented via techniques such as path integral molecular dynamics.

### 2.4.1 Real-time path integral

Let us recall the matrix elements of the quantum propagator (see eq. (2.6)) for a quantum particle in one-dimension evolving from the initial position $x$ to the final position $x'$ in an elapsed time $t$.

$$U(x', x, t) = \langle x' | e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} | x \rangle \tag{2.19}$$

The above matrix elements can be treated by dividing the total path of time length $t$ into infinitesimal intervals $\tau = t/n$. The final expression for the path integral of the quantum propagator is the following. The details of the mathematical derivation are reported in appendix, section A.1.

$$\langle x' | e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} | x \rangle = \lim_{n\to\infty} \left( \frac{m}{2\pi i\hbar\tau} \right)^{n/2} \int dx_1 \cdots dx_{n-1} \exp\left\{ \frac{i}{\hbar}\tau \sum_{j=0}^{n-1} \left[ \frac{m}{2} \left( \frac{x_{j+1} - x_j}{\tau} \right)^2 - U(x_j) \right] \right\} \tag{2.20}$$

where the $U(x, x', t)$ matrix is evaluated as a sum over all the possible paths leading from the starting point $x = x_0$ to the final point $x' = x_n$.

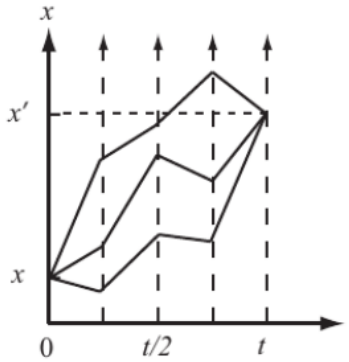



Figure 2.1: Representation of real-time path integral. Reproduction from ref. [37].

### 2.4.2 Functional integral representation

An elegant way to represent eq. (2.20) is via a functional integral, which originates from the following considerations. In the limit $n \to \infty$, which implies $\tau \to 0$, the time interval between the points $x_j$ and $x_{j+1}$ become infinitely small, while the number of points becomes infinite. The $x_j$ points

represent therefore a complete set of points defining the continuous function $x(s)$, that satisfies the boundary conditions:

$$\begin{cases} x(0) = x \\ x(t) = x' \end{cases} \tag{2.21}$$

Let us now focus on the exponential argument of eq. (2.20). In the limit $\tau \to 0$, we can assume:

$$\lim_{\tau \to 0} \frac{x_{j+1} - x_j}{\tau} = \frac{dx}{ds} = \dot{x}(s) \tag{2.22}$$

The whole argument of the exponential can now be considered as a Riemann sum of the integral:

$$\lim_{\tau \to 0} \tau \sum_{j=0}^{n-1} \left[ \frac{m}{2} \left( \frac{x_{j+1} - x_j}{\tau} \right)^2 - U(x_j) \right] = \int_0^t \left[ \frac{1}{2} m \dot{x}^2(s) - U(x(s)) \right] ds \tag{2.23}$$

where the integrand term appears as a difference between the kinetic and potential energy and corresponds, therefore, to the classical Lagrangian. The integral of the Lagrangian along the path $x(s)$ is, by definition, the action $\mathcal{S}[x]$,

$$\mathcal{S}[x] = \int_0^t \left[ \frac{1}{2} m \dot{x}^2(s) - U(x(s)) \right] ds \tag{2.24}$$

By introducing the functional $\mathcal{D}[x]$ as the following:

$$\mathcal{D}[x] = \lim_{n \to \infty} \left( \frac{m}{2\pi i \hbar \tau} \right)^{n/2} dx_1 \cdots dx_{n-1} \tag{2.25}$$

we can rewrite eq. (2.20) in the more compact way:

$$\langle x' | e^{-\frac{i}{\hbar} \hat{\mathcal{H}} t} | x \rangle = \int \mathcal{D}[x] \, e^{\frac{i}{\hbar} \mathcal{S}[x]} \tag{2.26}$$

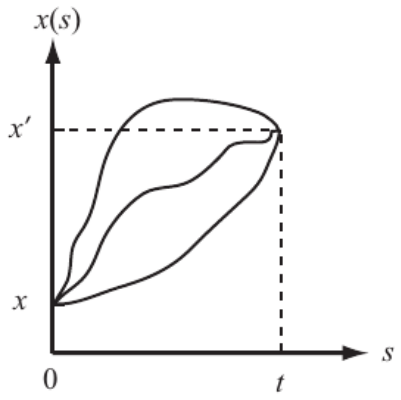



Figure 2.2: Representation of path integral in the continous-time limit. Reproduced from ref. [37].

which corresponds to the final functional form of the path integral. Here, all the possible paths are

formally weighted by their action by a phase factor $e^{\frac{i}{\hbar}\mathcal{S}[x]}$. However, the numerical resolution of the integral in eq. (2.26) is prohibitive, as the exponential scaling of the Schrödinger equation, due to the so-called *sign problem*. This arises by the presence of a strongly oscillating complex-valued action, that prevents to define a probability density so that the standard statistical methods cannot be applied. To overcome this problem, it is convenient to convert the PI formalism from real-time to imaginary-time, in which the phase factor is defined real positive, allowing a statistical treatment and therefore a numerical implementation.

### 2.4.3 Imaginary-time path integral

As anticipated in the previous section, the imaginary-time approach is a handful way to represent path integrals in statistical mechanics, allowing the incorporation of Feynman formalism in molecular simulations via Monte Carlo techniques or molecular dynamics. The tool required for this representation is the density matrix operator, which provides a link between quantum and statistical mechanics. Firstly, it must be noted that the canonical density matrix operator and the quantum propagator are strictly related. It is convenient to use the unnormalized canonical density matrix operator of eq. (2.11) by redefining $\hat{\rho}(\beta) = e^{-\beta\hat{\mathcal{H}}}$. Now, the relationship between $\hat{\rho}(\beta)$ and the time propagator defined in eq. (2.3) can be derived through the Wick rotation:

$$\hat{\rho}(\beta) = \hat{U}(-i\beta\hbar) \qquad \hat{U}(t) = \hat{\rho}\left(\frac{i}{\hbar}t\right) \tag{2.27}$$

which means that the thermalization of the system at an inverse temperature $\beta = it/\hbar$ is equivalent to the evolution of the system in the imaginary time $t = -i\beta\hbar$. Using the above relationship for eq. (2.20), the matrix elements of $\rho(x', x, \beta)$ are given by:

$$\langle x'|e^{-\beta\hat{\mathcal{H}}}|x\rangle = \lim_{n\to\infty}\left(\frac{m}{2\pi\beta_n\hbar^2}\right)^{n/2}\int dx_1\cdots dx_{n-1}\exp\left\{-\sum_{j=0}^{n-1}\left[\frac{mn}{2\beta\hbar^2}(x_{j+1}-x_j)^2+\frac{\beta}{n}U(x_j)\right]\right\} \tag{2.28}$$

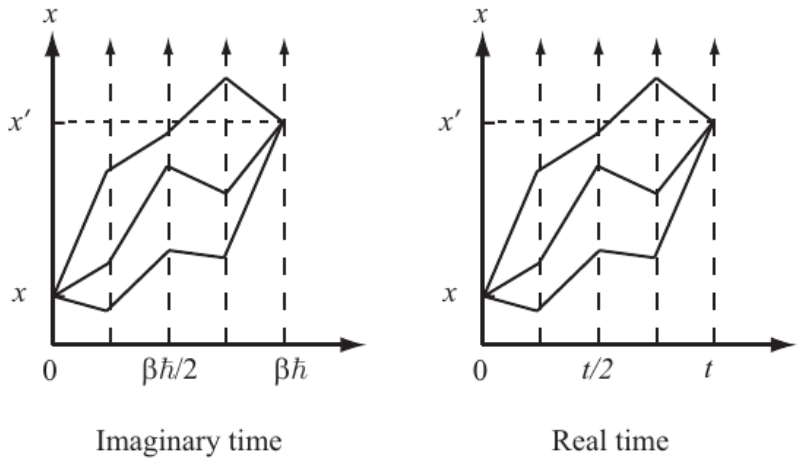



Figure 2.3: Imaginary-time *vs* real-time path-integral. Reproduction from ref. [37].

From eq. (2.28), the canonical partition function $\mathcal{Z}$ as per definition of eq. (2.12) can be derived. The evaluation of the trace in coordinate-space gives:

$$\mathcal{Z} = \int dx \, \langle x|e^{-\beta \hat{\mathcal{H}}}|x\rangle = \rho(x, x, \beta) \tag{2.29}$$

To solve eq. (2.29), the integration of eq. (2.28) has to be performed over the diagonal elements. The resulting integration is over closed paths ($\oint$) which begins and ends a the same point $x' = x$. The partition function is therefore given by:

$$\mathcal{Z} = \lim_{n\to\infty} \left(\frac{mn}{2\pi\beta\hbar^2}\right)^{n/2} \oint dx_0 \cdots dx_{n-1} \exp\left\{-\sum_{j=0}^{n-1}\left[\frac{mn}{2\beta\hbar^2}(x_{j+1}-x_j)^2+\frac{\beta}{n}U(x_j)\right]\right\} \tag{2.30}$$

The above equation is not straightforward to use. In the following the PIMD method, which is a combination of molecular dynamics and path-integral formalism, is discussed.

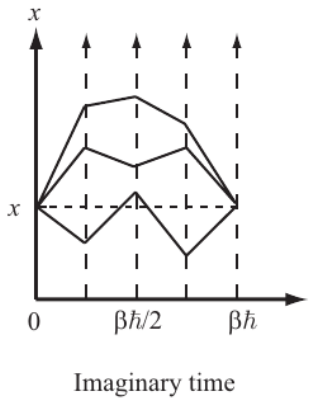



Figure 2.4: Represention for a discrete path sum for the canonical partition function. Reproduced from ref. [37].

**Path Integral Molecular Dynamics** In this tecnique, molecular dynamics is used as a tool to sample the canonical quantum partition function. The first PIMD simulations conducted in the '80s [42–44] employed empirical potential, later on, in 1996, Marx and Parinello opened the path to *ab initio* PIMD [12]. Different PIMD-based techniques have been developed during the years, allowing also an approximate description of quantum dynamical properties. We will focus our attention on the standard PIMD, which represents the reference method for describing NQEs. The idea behind PIMD is to link the quantum partition function to a classical one, by rephrasing eq. (2.30). Firstly, a finite description is adopted so that the limit $n \to \infty$ is removed for the purposes of numerical implementation. Then, it is convenient to insert $n$ times Gaussian integrals over fictitious momenta $p_j$ as conjugates of the positions $x_j$. By adopting this strategy, the final expression for the quantum partition function is:

$$\mathcal{Z}_n = \int dp_0 \cdots dp_{n-1} \oint dx_0 \cdots dx_{n-1} \, \exp\left\{ -\beta \left[ \sum_{j=0}^{n-1} \frac{p_j^2}{2\tilde{m}} + U_{\text{eff}} \right] \right\} \tag{2.31}$$

where we have introduced the following parameters. The frequency $\omega_n$ corresponds to the "chain" frequency of a cyclic polymer formed by $n$ particles, called beads (replicas). It is defined as:

$$\omega_n = \frac{\sqrt{n}}{\beta\hbar} \tag{2.32}$$

$U_{\text{eff}}$ is the effective potential experimented by the polymer:

$$U_{\text{eff}} = \sum_{j=0}^{n-1} \frac{1}{n} U(x_j) + \frac{1}{2} m\omega_n^2 (x_{j+1} - x_j)^2 \tag{2.33}$$

The constant $\tilde{m}$ introduced by the Gaussian integrals is a fictitious mass given by $\tilde{m} = mn/(2\pi\hbar)^2$ but it can be assigned as we like since it does not affect the equilibrium averages. Now, the quantum partition function of eq. (2.31) looks like the classical partition function of a $n$-particle system having a potential $U_{\text{eff}}$. The latter assumption is known as the *classical isomorphism*. The fictitious classical system is a cyclic polymer of $n$ replicas interacting via harmonic springs with force constants $m\omega_n$, and trapped in a potential $U(x)/n$ (see fig. 2.5).

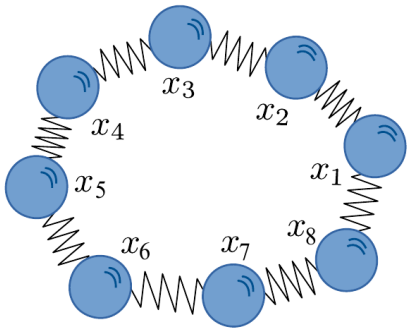



Figure 2.5: Schematic representation of cyclic ring polymer chain having $n = 8$ beads described by the partition function of eq. (2.31). Each bead is represented by the blue spheres and it is connected to its nearest neighbors via harmonic springs. Reproduction from reference [45].

By defining:

$$\hat{\mathcal{H}}_{\text{eff}} = \sum_{j=0}^{n-1} \frac{p_j^2}{2\tilde{m}} + U_{\text{eff}} \tag{2.34}$$

in eq. (2.31) as a classical-like effective Hamiltonian $\hat{\mathcal{H}}_{\text{eff}}$, we can derive a molecular dynamics scheme using the following Hamilton equations:

$$\dot{x}_j = \frac{\partial \hat{\mathcal{H}}_{\text{eff}}}{\partial p_j} = \frac{p_j}{\tilde{m}}$$
$$\dot{p}_j = -\frac{\partial \hat{\mathcal{H}}_{\text{eff}}}{\partial x_j} = -m\omega_n^2(2x_j - x_{j+1} - x_{j-1}) - \frac{1}{n}\frac{\partial U}{\partial x_j} \tag{2.35}$$

In the limit of infinite $n$, eq. (2.31) yields the exact quantum partition function, while when $n = 1$, the classical regime is recovered. For a finite number of $n$ replicas, PIMD converges to the exact expectation values when $n \sim \beta\hbar\omega_{\max}$, with $\omega_{\max}$ being the frequency of the fastest normal mode, making PIMD a computationally expensive technique. Thus, some accelerating approaches reducing the number of replicas have been developed [46, 47]. Once $n$ is large enough, the static averages at thermodynamic equilibrium are accurately computed. In contrast, the evaluation of dynamical properties, such as time-correlation functions, cannot be carried out exactly. Several path-integral based approximations, such as ring polymer molecular dynamics [48] and centroid molecular dynamics [49], have been developed. These methods have been satisfactorily applied to low-frequency scale related phenomena such as diffusion and transport properties [50]. However, when one wants to describe a high-frequency motion as molecular vibrations, these methods suffer from artificial resonances between the chain modes of the polymer and the vibrational modes coming from the system, leading to spurious frequencies [51, 52], a problem that can be eventually treated by the introduction of a Langevin thermostat attached to the internal modes of the polymer [53]. Although some significant improvements, these methods present still some limitations for vibrational spectroscopy.

## 2.5 Semiclassical vibrational spectroscopy

The exact quantum dynamics given by the TDSE and the Feynman's quantum propagator of eq. (2.26) is impossible to evaluate numerically, making the computation of time-dependent properties including NQEs rather challenging, especially in the case of molecular vibrations, for which the above mentioned method based on path integrals in imaginary-time (section 2.4.3) is not accurate. Quantum effects such as zero-point energy motion, resonances, overtones and combination bands are fundamental in vibrational spectroscopy and require an adequate quantum description. An answer to this challenge can be given by semiclassical methods, which are constructed from a superposition of classical trajectories with special initial conditions. In such a way, semiclassical (SC) methods provide an approximate description (albeit rather good in many cases) of the full quantum dynamics. From the first formulation of SC propagator by van Vleck, which can be directly derived by Feynman's path integral expression for the quantum propagator, semiclassical techniques applied to vibrational spectroscopy have been widely developed. In this section, the most recent improvements of SC spectroscopy are discussed, along with the quasi-classical trajectory approximation.

### 2.5.1 Semiclassical approximation to path integrals

One way to derive the semiclassical approximation is from Feyman's path integral formulation of the quantum propagator (eqs. (2.20) and (2.26)). In the classical limit $\hbar \to 0$, the exponential factor $e^{\frac{i}{\hbar}\mathcal{S}[x]}$ oscillates rapidly, thus the path integral will be dominated by the stationary phase paths, i.e. the critical points for which:

$$\frac{\delta \mathcal{S}}{\delta x_{cl}(s)} = 0 \tag{2.36}$$

where the critical points are labeled $x_{cl}$ since they satisfy the Hamilton principle of least action and are, therefore, classical paths as depited in fig. 2.6. The semiclassical approximation arises from accounting for these classical paths as the main contribution to the path integral.

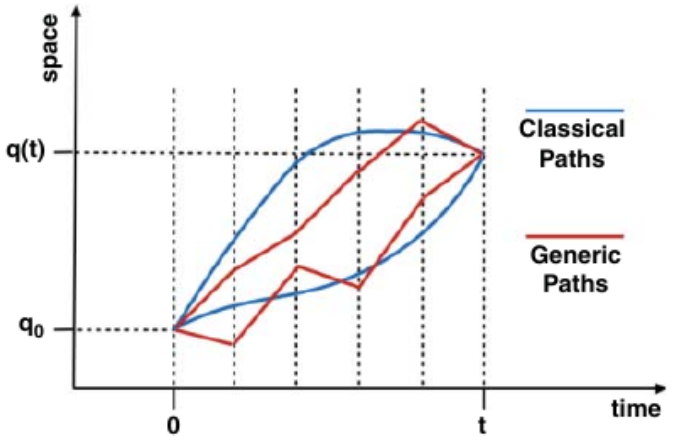



Figure 2.6: Representative paths (classical *vs* generic) of Feynman's path integral quantum propagator. Reproduction from ref. [38]. In the graph $\mathbf{q}_0$ and $\mathbf{q}(t)$ are respectively the initial and final space-points.

The mathematical tool used to derive the semiclassical propagator is the stationary phase approximation (SPA) [54], which for the one-dimensional case reads as:

$$\int dx\, e^{i\lambda f(x)} = \sum_{\{\bar{x}_j | f'(\bar{x}_j)=0\}} \sqrt{\frac{2\pi i}{\lambda f''(\bar{x}_j)}} e^{i\lambda f(\bar{x}_j)} \tag{2.37}$$

with $f(x)$ being an arbitrary oscillating function that is stationary for the critical points $\bar{x}_j$. The generalization of SPA to the multi-dimensional case is obtained via substitution of $f''(x)$ by the determinant of the second derivatives matrix of $f$ with respect to the positions, i.e. the Hessian. Further details are given in appendix, section A.2.

The applications of the SPA to the path integral is presented in the following. As anticipated before, in the limit $\hbar \to 0$, the PI integral is dominated by the stationary phase points $x_{cl}$, which correspond to the classical paths. A generic quantum path $x(s)$ can be expressed as:

$$x(s) = x_{cl}(s) + \xi(s) \tag{2.38}$$

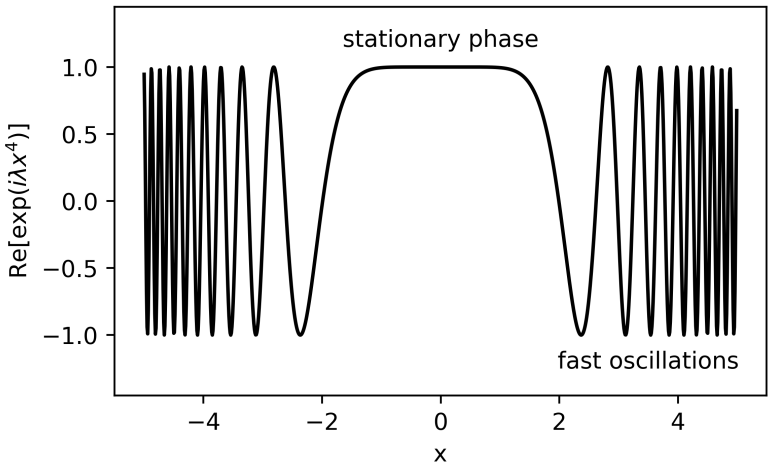



Figure 2.7: Example of a complex oscillating exponential $f(x) = \exp(i\lambda x^4)$ with $\lambda = 10^{-1}$. Only a small region around the critical point $x = 0$ contributes to the integral.

where $x_{cl}(s)$ is one of the classical paths and $\xi(s)$ the fluctuations around it. The expansion of the action around $x_{cl}$ at the second order in $\xi$ is given by:

$$\mathcal{S}[x] = \mathcal{S}[x_{cl}] + \frac{1}{2}\xi^T \mathbf{H}_{\mathcal{S}}\, \xi \tag{2.39}$$

where $\mathbf{H}_{\mathcal{S}} = [\partial \mathcal{S}_{cl}/\partial x_i \partial x_j]$ is the Hessian matrix of the classical action. In the above equation we have used a functional representation of the classical action: the classical action is a functional of the full classical path $x_{cl}$. For the purpose of the following derivations, it is convenient to replace the functional $\mathcal{S}[x_{cl}]$ with $\mathcal{S}_{cl}(x_t, x_0, t)$, where the classical action associated to the classical path $x_{cl}$ is expressed as a function of the initial and final position of the classical path, $x_0$ and $x_t$ points, respectively.[1] The semiclassical approximation for the PI quantum propagator is therefore given by:

$$\langle x_t | e^{-\frac{i}{\hbar}\hat{\mathcal{H}} t} | x_0 \rangle \propto \sum_{cl} \exp\left[\frac{i}{\hbar}\left(\mathcal{S}_{cl}(x_t, x_0, t) + \frac{1}{2}\xi^T \mathbf{H}_{\mathcal{S}} \xi\right)\right] \tag{2.40}$$

i.e. by a sum over all the possible classical paths connecting $x_0$ to $x_t$ points in a time $t$ with the inclusion of fluctuations around the classical action of each path up to the second order. The analytical resolution of eq. (2.40) leads to the van Vleck (VV) expression for the quantum propagator (1928) [55], which for one degree of freedom is:

$$U^{\mathrm{VV}}(x_t, x_0, t) = \sum_{cl} \left\{ -\frac{1}{2\pi i\hbar} \times \det\left(\frac{\partial^2 \mathcal{S}_{cl}(x_t, x_0, t)}{\partial x_t \partial x_0}\right)\right\}^{1/2} \exp\left[\frac{i}{\hbar}\mathcal{S}_{cl}(x_t, x_0, t)\right] \tag{2.41}$$

where the zero-th order term of the Taylor expansion of the action gives the oscillatory contribution $e^{\frac{i}{\hbar}\mathcal{S}_{cl}(x_t, x_0, t)}$, while the second order term, known as the van Vleck determinant, constitutes the pre-

[1] A different notation for the time evolution of a quantum particle is used with respect to section 2.4. The initial space point is labeled $x_0$ at a initial time $t_0 = 0$, while the final space point is labeled $x_t$ at final time $t$.

exponential factor describing the quantum fluctuations around the classical path. However, in the VV semiclassical propagator the prefactor, being a square root of a quantity without a definite sign, is, in general, complex and discontinuous. In order to preserve the continuity of the transformed propagator for trajectories close to the classical path, Maslov and Fedoriuk introduced the extra phase $-\frac{1}{2}i\pi\nu_{cl}$ [56], in which $\nu$ is called Maslov index:

$$U^{\mathrm{VVG}}(x_t, x_0, t) = \sum_{cl} \left\{ -\frac{1}{2\pi i\hbar} \times \det\left(\frac{\partial^2 \mathcal{S}_{cl}(x_t, x_0, t)}{\partial x_t \partial x_0}\right)\right\}^{1/2} \exp\left[\frac{i}{\hbar}\mathcal{S}_{cl}(x_t, x_0, t) - \frac{1}{2}i\pi\nu_{cl}\right] \tag{2.42}$$

The above equation is known as the van Vleck-Gutzwiller (VVG) formulation of the semiclassical propagator [57].

### 2.5.2 Initial value representation

It is possible to rewrite the van Vleck semiclassical propagator in terms of initial conditions by using the relation $\partial\mathcal{S}_{cl}(x_t, x_0, t)/\partial x_0 = -p_0$ in eq. (2.41):

$$U^{\mathrm{VV}}(x_t, x_0, t) = \sum_{cl} \left\{ \frac{1}{2\pi i\hbar} \times \det\left(\frac{\partial x_t}{\partial p_0}\right)^{-1}\right\}^{1/2} \exp\left[\frac{i}{\hbar}\mathcal{S}_{cl}(x_t, x_0, t)\right] \tag{2.43}$$

However, the above formulation has quite limited applications. This is due firstly by the fact that when $\partial x'/\partial p_0 \to 0$ the prefactor matrix becomes singular. Secondly, the search of all the possible classical paths satisfying the double boundary problem $x(0) = x_0 \wedge x(t) = x'$ is not trivial. To tackle these problems a semiclassical initial value representation (IVR) approach to the SC propagator was introduced by W.H. Miller in 1970 [15]. The idea behind the so-called SCIVR theory is to change the variables in the propagator in such a way to specify, for a generic path around the stationary one, the unique initial conditions $(x_0, p_0)$ that define the actual classical path. This can be done as it follows. We start with the time propagator operator expression $\hat{U}(t) = e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}$ and we insert two-times the identity $\int dx\, |x\rangle\langle x| = \mathbb{I}$, one for $|x_0\rangle$ and one for $|x_t\rangle$ vectors. Then, the resulting propagator coordinate-matrix is substituted with the corresponding VV SC propagator matrix of eq. (2.43):

$$\begin{aligned} e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} &= \int dx_0 \int dx_t\, |x_t\rangle\langle x_t| e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} |x_0\rangle\langle x_0| \\ &\approx \int dx_0 \int dx_t\, |x_t\rangle\, U^{\mathrm{VV}}(x_t, x_0, t)\, \langle x_0| \\ \approx \int dx_0 \sum_{cl} \int dx_t \left\{ \frac{1}{2\pi i\hbar} \times \det\left(\frac{\partial x_t}{\partial p_0}\right)^{-1}\right\}^{1/2} & e^{\frac{i}{\hbar}\mathcal{S}_{cl}(x_t, x_0, t)}\, |x_t\rangle\langle x_0| \end{aligned} \tag{2.44}$$

Now, it is convenient to use the "IVR trick" [58, 59], which consists in a change of variable integration from $x_t$ to $p_0$:

$$\sum_{cl} \int dx_t \rightarrow \int dp_0 \ \det\left(\frac{\partial x_t}{\partial p_0}\right) \tag{2.45}$$

The final expression for the VV propagator in the IVR representation is therefore:

$$e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} \approx \int dx_0 \int dp_0 \left[\frac{1}{2\pi i\hbar} \times \det\left(\frac{\partial x_t}{\partial p_0}\right)\right]^{1/2} e^{\frac{i}{\hbar}\mathcal{S}_t} \ |x_t\rangle\langle x_0| \tag{2.46}$$

In this way, the sum over all the possible classical paths is substituted with a phase space integral that can be numerically evaluated via a Monte Carlo integration. The SCIVR is therefore a very powerful approach which has been implemented many times since its original formulation by Miller. A quite popular version of SCIVR propagator is the one based on coherent states, a description firstly introduced in the work by Heller [60], from which Herman and Kluk drew inspiration in the later years [61]. The coherent states $|p', x'\rangle$ having a Gaussian shape of width $\gamma$ and centered on $(p', x')$ are chosen according to:

$$\langle x|p', x'\rangle = \left(\frac{\det(\gamma)}{\pi}\right)^{1/4} \exp\left[-\frac{1}{2}(x-x')^T\gamma(x-x') + \frac{i}{\hbar}p'^T(x-x')\right] \tag{2.47}$$

The survival amplitude for a reference state $|\chi\rangle$ using the semiclassical propagator of eq. (2.46) on the coherent states basis gives:

$$\langle\chi|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|\chi\rangle \approx \left(\frac{1}{2\pi\hbar}\right) \iint dx_0 dq_0 \ C_t(p_0, x_0) e^{-\frac{i}{\hbar}\mathcal{S}_t(p_0,x_0)} \langle\chi|p_t, x_t\rangle\langle p_0, x_0|\chi\rangle \tag{2.48}$$

where $C_t(p_0, x_0)$ is the prefactor:

$$C_t(p_0, x_0) = \sqrt{\left|\frac{1}{2}\left(\frac{\partial x_t}{\partial x_0} + \gamma^{-1}\frac{\partial p_t}{\partial p_0}\gamma - i\hbar\frac{\partial x_t}{\partial p_0} + \frac{i}{\hbar}\gamma^{-1}\frac{\partial p_t}{\partial x_0}\right)\right|} \tag{2.49}$$

The partial derivatives $\partial i/\partial j$ represent the elements of the monodromy matrix $\mathbf{M}_{ij}$:

$$\mathbf{M} = \begin{pmatrix} \partial p_t/\partial p_0 & \partial p_t/\partial x_0 \\ \partial x_t/\partial p_0 & \partial x_t/\partial x_0 \end{pmatrix} \tag{2.50}$$

$\mathbf{M}_{ij}$, also called stability matrix, measures the sensitivity of the trajectory to the initial conditions. Eqs. (2.48) to (2.50) represents the Herman and Kluk (HK) formulation of IVR propagator [61, 62], here reported for one degree of freedom for simplicity of sake. The survival amplitude can be numerically evaluated via a phase space integration by Monte Carlo techniques upon sampling of the initial conditions $(x_0, p_0)$.

### 2.5.3 Application to spectroscopy

For applications to vibrational spectroscopy it is necessary to extend the problem to $N$ degrees of freedom and work in normal modes coordinates, on which vibrational calculations at the harmonic level are generally performed. As per notation compatible with the literature, the space coordinates are labeled as $q_j$, and the corresponding momenta $p_j$. The phase space vectors $\mathbf{q}$ and $\mathbf{p}$ are expressed:

$$\begin{cases} \mathbf{q} = (q_1, q_2, ..q_N) & N = \text{n.o. DOF (degrees of freedom)} \\ \mathbf{p} = (p_1, p_2, ..p_N) \end{cases} \tag{2.51}$$

We recall that (see eq. (2.17)) the quantum power spectrum $I(E)$ is given by the Fourier transform of the survival amplitude of a generic reference state $|\chi\rangle$:

$$I(E) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} dt \, \langle\chi|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}\,t}|\chi\rangle e^{\frac{i}{\hbar}Et} \tag{2.52}$$

**Herman-Kluk SCIVR** By using the formulation of Herman and Kluk for the semiclassical propagator (eq. (2.48)), the HK IVR formulation for the spectral density for $N$ degrees of freedom is:

$$I(E) = \left(\frac{1}{2\pi\hbar}\right) \int_{-\infty}^{+\infty} dt e^{\frac{i}{\hbar}Et} \left(\frac{1}{2\pi\hbar}\right)^N \iint d\mathbf{p}_0 d\mathbf{q}_0 \; C_t(\mathbf{p}_0, \mathbf{q}_0) e^{\frac{i}{\hbar}\mathcal{S}_t(\mathbf{p}_0,\mathbf{q}_0)} \langle\chi|\mathbf{p}_t, \mathbf{q}_t\rangle\langle\mathbf{p}_0, \mathbf{q}_0|\chi\rangle \tag{2.53}$$

Applications of the HK propagator have been limited to model systems as in references [62, 63]. The extension to molecular systems is limited due to the oscillating exponential in the integrand. One method to overcome this issue is to insert a time-averaging (TA) filter.

**Time-Averaging Filter** A time-average of the type $\frac{1}{T}\int_0^T dt$ can be inserted in the phase-space integrand of the HK semiclassical power spectrum eq. (2.53) as formulated by Kaledin and Miller [64] (2003). The resulting spectral density is:

$$I(E) = \left(\frac{1}{2\pi\hbar}\right)^N \iint d\mathbf{p}_0 d\mathbf{q}_0 \; \frac{1}{2\pi\hbar T} \left| \int_0^T dt \; \exp\left\{\frac{i}{\hbar}\left[\mathcal{S}_t(\mathbf{p}_0, \mathbf{q}_0) + Et + \phi_t\right]\right\} \langle\chi|\mathbf{p}_t, \mathbf{q}_t\rangle \right|^2 \tag{2.54}$$

where $\phi_t = \text{phase}[C_t(\mathbf{p}_0, \mathbf{q}_0)]$. In this way the convergence of the phase-space integrand is faster, reducing the number of the required trajectories. However, the computational cost of TA-SCIVR for molecular systems is still really high (a thousand of trajectories per degree of freedom), so that the applications are restricted to small molecular systems, for which pre-existing computed PES are available.

**Multiple Coherent States SCIVR (MC-SCIVR)** To overcome the issues with the TA-SCIVR, Ceotto and coworkers introduced, in 2009, the multiple coherent states SCIVR tecnique [65], that

replaces the Monte Carlo phase-space integration of eq. (2.54) with a sum over a few tailored trajectories. This led to a significant reduction of the computational cost, opening the path towards *ab initio* "on-the-fly" semiclassical spectroscopy in full dimensionality. A large variety of applications have been studied with successful outcomes [66–77]. The MC-SCIVR expression for the power spectrum reads:

$$I(E) = \left(\frac{1}{2\pi\hbar}\right)^N \frac{1}{2\pi\hbar T n_{\mathrm{traj}}} \sum_{j=1}^{n_{\mathrm{traj}}} \left| \int_0^T dt \exp\left\{\frac{i}{\hbar}\left[\mathcal{S}_t(\mathbf{p}_0^{(j)}, \mathbf{q}_0^{(j)}) + Et + \phi_t^{(j)}\right]\right\} \langle \chi^{(j)} | \mathbf{p}_t^{(j)}, \mathbf{q}_t^{(j)} \rangle \right|^2 \tag{2.55}$$

where $n_{\mathrm{traj}}$, the number of trajectories, is limited to carefully selected classical trajectories. This approach derives from the pioneering work of De Leon and Heller [78] (1983), which proves that accurate semiclassical results can be achieved even by means of a single trajectory if it is run at the exact quantum energy [78]. The idea of MC-SCIVR is therefore to use trajectories that have an energy nearby the correct quantum one, which, however, is unknown a priori. To this aim, the initial conditions $(\mathbf{q}_0, \mathbf{p}_0)$ are chosen according the following criteria [65, 79] .

$$\begin{cases} \mathbf{q}_0 = \mathbf{q}_{\mathrm{eq}} \\ \mathbf{p}_0 = 2m\sqrt{\hbar\boldsymbol{\omega}(\mathbf{n} + 1/2)} \end{cases} \tag{2.56}$$

in which $\mathbf{q}_{\mathrm{eq}}$ is the equilibrium geometry and $\boldsymbol{\omega} = (\omega_1, \omega_2, \cdots \omega_N)$ are the harmonic eigenvalues obtained via diagonalization of the Hessian matrix at $\mathbf{q}_{\mathrm{eq}}$. In such a way, the trajectories are started at the equilibrium position with a harmonic zero-point energy of $n$ quantum excitation. When adopting these initial conditions, it is possible to yield accurate results by means of a single trajectory ($n_{\mathrm{traj}} = 1$), as confirmed in the work of Gabas et al. [66], where single-trajectory MC-SCIVR was employed to investigate the multiple conformers of neutral glycine. Another important feature of eq. (2.55) is the nature of the reference state $|\chi\rangle$, which is chosen to enhance the signal of a selected vibrational mode. For each trajectory the reference state used is a combination of coherent states of the type :

$$|\chi\rangle = \prod_{k=1}^{N} \mid p_{0,k}, q_{0,k} \rangle + \varepsilon_k \mid -p_{0,k}, q_{0,k} \rangle \tag{2.57}$$

where index $k$ refers to the $k$-th mode components of vector $\mathbf{p}_0$ and $\mathbf{q}_0$. The coefficient $\varepsilon_k$ is equal to $\pm 1$ depending on the desired spectroscopic signal. For example, if it is $+1$ for all the $k$-th components, the spectroscopic signal enanched is the ZPE peak and even overtones. Instead, when one of the $k$-th component is switched to $-1$, the signal enanched is the fundamental of the $k$-th mode and the odd overtones [80].

**Divide and Conquer SCIVR (DC-SCIVR)** The development of a multiple coherent states approach has allowed *ab initio* semiclassical vibrational study of small size systems, like ammonia [71], up to medium size systems, as glycine molecule [66]. However, the direct application of this approach to systems with higher degrees of freedom is not fleasible due to the well-known problem of the "curse of dimensionality". As the dimensionality of the degrees of freedom in-

creases, it is more difficult to resolve the spectroscopic signal due to a decrease of the signal-to-noise ratio. Furthermore, to a large number of degrees of freedom it corresponds also a increased computational cost, which can be prohibitive via *ab initio*. These drawbacks have been tackled by adopting a divide and conquer strategy, introduced by Ceotto and coworkers in 2017 [81], where the high-dimensional system vibrational space is projected onto lower-dimensional subspaces. In DC-SCIVR method the power spectrum $I(E)$ is obtained as the composition of partial spectra $\tilde{I}(\tilde{E})$ computed in a $M$-dimensional subspace of the full $N$-dimensional space. In the case of a single-trajectory SC dynamics, the partial spectra $\tilde{I}(\tilde{E})$ are given by the following expression.

$$\tilde{I}(\tilde{E}) = \left(\frac{1}{2\pi\hbar}\right)^{M} \frac{1}{2\pi\hbar T} \left| \int_0^T dt \, \exp\left\{ \frac{i}{\hbar} \left[ \tilde{\mathcal{S}}_t(\tilde{\mathbf{p}}_0, \tilde{\mathbf{q}}_0) + \tilde{E}t + \tilde{\phi}_t \right] \right\} \langle \chi | \tilde{\mathbf{p}}_t, \tilde{\mathbf{q}}_t \rangle \right|^2 \tag{2.58}$$

where the projected quantities onto a $M$-dimensional subspace by a singular value decomposition, are indicated with $\sim$ symbol. All the terms of Eq. (2.58) can be exactly projected from the full dimensional space onto the subspace of interest with the exception of the the classical action $\tilde{\mathcal{S}}_t$ due to the fact that the potential is not generally separable. The potential energy, therefore, requires the following *ad hoc* procedure.

$$V_S(\tilde{\mathbf{q}}_t^M) = V\left(\tilde{\mathbf{q}}_t^M; \mathbf{q}_t^{(N-M)}\right) - V\left(\tilde{\mathbf{q}}_{\text{eq}}^M; \mathbf{q}_t^{(N-M)}\right) \tag{2.59}$$

which is exact for separable-potentials and a good approximation for the non-separable ones.

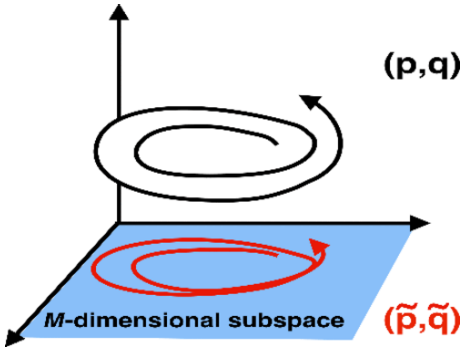



Figure 2.8: Pictorial representation of the projection procedure. Reproduction from reference [81].

The computational cost of a DC-SCIVR calculation is significantly reduced with respect to MC-SCIVR since the computation of the Hessian is made just with the information of the lower-dimensional subspace $M$-dimensional, allowing to obtain semiclassical spectra for systems with very large number of DOF.[2] Remarkable applications include biomolecular systems [82–84] and surface related phenomena [85, 86].

The most problematic part is the choice of the correct $M$-dimensional subspace. The subspaces are chosen so that the most coupled modes belong to the same subspace. There are different criteria to effectuate the partition of the vibrational space [87], among which we recall the "Hessian space-decomposition method". In this method, firstly the time average of the Hessian matrix $\bar{H}_{ij}$ along

[2] We can evaluate the computational cost with the number of degrees of freedom to the power of $q$. In the case of a DC-SCIVR calculation, $(K+M)^q > M^q$, where $M$ is the dimension of the subset of the full dimensional space with $(K+M)$ dimension; $q$ ($q > 1$) is the power at which the computational cost grows.

the dynamics (of length $n_{\text{steps}}$) is computed in normal modes representation:

$$\bar{H}_{ij} = \frac{1}{n_{\text{steps}}} \sum_{k=1}^{n_{\text{steps}}} |H_{ij}|^k \tag{2.60}$$

in which $i$ and $j$ refer to the $i$-th and $j$-th mode. Secondly, the averaged coupling strength between different modes is evaluated by the measure of the off-diagonals terms of $\bar{H}_{ij}$ ($i \neq j$) respect to a fixed threshold parameter $\epsilon$. The following considerations are adopted:

1. if $\bar{H}_{ij} \geq \epsilon$ : $i$ and $j$ belong to the same subspace;
2. if $\bar{H}_{ij} < \epsilon$ : $i$ and $j$ do not belong to the same subspace;
3. if $\bar{H}_{ij} < \epsilon \wedge l \ \exists \mid \bar{H}_{il} \geq \epsilon \wedge \bar{H}_{jl} \geq \epsilon$ : $i$, $j$ and $l$ belong to the same subspace.

The threshold $\epsilon$ is an arbitrary parameter and we cannot known *a priori* the best choice for it. Therefore it is necessary to test different values of $\epsilon$ and the optimal choice is discriminated with respect the quality of the spectroscopic signals and the subspace dimensionality suitable for numerical convergence.

### 2.5.4 Quasi-classical approximation and the classical limit

In the quasi-classical trajectory (QCT) approach the initial conditions follow a quantization scheme but the dynamics is treated classically. The idea is to give to each mode a precise vibrational energy by running a classical trajectory at that energy, thus improving the results of the pure classical picture. The initial conditions are quantized corresponding to eq. (2.56), i.e. with a harmonic estimate of the specific quantum excitation. In this way the trajectory, which is highly energetic, allows the exploration of a large portion of the PES so that anharmonic effects are accounted in the spectra. However, only the classical vibrational frequencies can be calculated and no quantum mechanical effects are included.

As in the classical case, the QCT spectra can be calculated via the Fourier transform of the velocity-velocity autocorrelation function. For simplicity here a one-dimensional problem is considered and as derived in reference [88]. By recalling the definition of time correlation functions calculated from a single classical trajectory (eq. (1.44)), the velocity-velocity autocorrelation function is expressed as:

$$C_{vv}(\tau) = \langle v(t)v(t+\tau) \rangle = \frac{1}{T} \int_0^T dt \, v(t)v(t+\tau) \tag{2.61}$$

with $T$ the total time of the simulation, which has to be sufficiently long. By substituting the above definition into the relation for the power spectrum eq. (1.50), we obtain:

$$I(\omega) = \int_{-\infty}^{+\infty} d\tau \, e^{-i\omega\tau} C_{vv}(\tau) = \frac{1}{T} \int_{-\infty}^{+\infty} d\tau \int_0^T dt \, e^{-i\omega\tau} v(t)v(t+\tau) \tag{2.62}$$

where the velocities $v(t)$ and $v(t+\tau)$ can be written as the inverse Fourier transform of $\tilde{v}(\omega')$ and

$\tilde{v}(\omega'')$:[3]

$$v(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega'\, e^{i\omega' t}\tilde{v}(\omega') \qquad v(t+\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega''\, e^{i\omega''(t+\tau)}\tilde{v}(\omega'') \tag{2.63}$$

By inserting eq. (2.61) into eq. (2.62) and using the two above expressions, $I(\omega)$ yields:

$$I(\omega) = \frac{1}{(2\pi)^2 T}\int_0^T dt \int_{-\infty}^{+\infty} d\omega'\, e^{i\omega' t}\tilde{v}(\omega')\int_{-\infty}^{+\infty} d\omega''\, e^{i\omega'' t}\tilde{v}(\omega'')\int_{-\infty}^{+\infty} d\tau\, e^{i(\omega''-\omega)\tau} \tag{2.64}$$

Therefore[4],

$$\begin{aligned} I(\omega) &= \frac{1}{(2\pi)^2 T}\int_0^T dt \int_{-\infty}^{+\infty} d\omega'\, e^{i\omega' t}\,\tilde{v}(\omega')\int_{-\infty}^{+\infty} d\omega''\, e^{i\omega'' t}\,\tilde{v}(\omega'')\,2\pi\delta(\omega''-\omega) \\ &= \frac{1}{2\pi T}\int_0^T dt \int_{-\infty}^{+\infty} d\omega'\, e^{i\omega' t}\,\tilde{v}(\omega')\,e^{i\omega t}\,\tilde{v}(\omega) \\ &= \frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\omega' \int_0^T dt\, e^{i(\omega'+\omega)t}\,\tilde{v}(\omega')\,\tilde{v}(\omega) \end{aligned} \tag{2.66}$$

Now, taking the limit of $T \to +\infty$, the last term of eq. (2.66) can be rewritten as:

$$I(\omega) = \lim_{T\to+\infty}\frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\omega' \int_0^{+\infty} dt\, e^{i(\omega'+\omega)t}\,\tilde{v}(\omega')\,\tilde{v}(\omega) \tag{2.67}$$

Taking advantage of the parity of the integrand in eq. (2.67), we rewrite:

$$\begin{aligned} I(\omega) &= \lim_{T\to+\infty}\frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\omega'\,\frac{1}{2}\int_{-\infty}^{+\infty} dt\, e^{i(\omega'+\omega)t}\,\tilde{v}(\omega')\tilde{v}(\omega) \\ &= \lim_{T\to+\infty}\frac{1}{2\pi T}\int_{-\infty}^{+\infty} d\omega'\,\frac{1}{2}\,\tilde{v}(\omega')\,\tilde{v}(\omega)\,2\pi\delta(\omega'+\omega) \\ &= \lim_{T\to+\infty}\frac{1}{2T}\int_{-\infty}^{+\infty} d\omega'\,\frac{1}{2}\,\tilde{v}(\omega')\,\tilde{v}(\omega)\,\delta(\omega'+\omega) \end{aligned} \tag{2.68}$$

The above integral is resolved in:

$$I(\omega) = \frac{1}{2T}\tilde{v}(\omega)\tilde{v}(-\omega) \tag{2.69}$$

[3] The convention used for the Fourier transform and its inverse is defined as the following.

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} dt\, f(t)\, e^{-i\omega t} \qquad f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega\, \tilde{f}(\omega)\, e^{i\omega t}$$

[4] The Dirac delta function is defined as the following.

$$\delta(x'-x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dt\, e^{i(x'-x)t} \tag{2.65}$$

Since $v(t)$ is real, $\tilde{v}(-\omega) = \tilde{v}^*(\omega)$. The final working formula for the quasi-classical vibrational spectra is:

$$I(\omega) = \frac{1}{2T}|\tilde{v}(\omega)|^2 = \frac{1}{2T}\left|\int_0^T dt\, e^{i\omega t} v(t)\right|^2 \tag{2.70}$$

## 2.6 Quantum baths

The purpose of this section is to briefly recall the QTB method, which can be employed to evaluate both static and dynamical properties as an alternative to state-of-the-art PIMD methods. It combines classical trajectory-based dynamics with an approximate quantum phase-space sampling. The QTB uses a generalized Langevin equation, in which the quantum Bose-Einstein energy distribution is imposed instead of that given by the classical equipartition. This is realized by a convenient adjustment of the random and friction forces. The great advantage of this technique is the low computational cost, which is comparable to standard Langevin molecular dynamics, thus, much less than the cost of PIMD-based methods.

### 2.6.1 Classical Langevin thermostat

In 1908 Paul Langevin introduced a model for describing the Brownian motion [89], which is an erratic motion of a "heavy" particle in a fluid composed of much "lighter" particles [90]. The standard Langevin equation of motion for a one-dimensional particle of mass $m$ reads:

$$m\ddot{x} = -\frac{dV}{dx} - m\gamma\dot{x} + R(t) \tag{2.71}$$

where $-dV/dx$ are the inter-atomic forces due to the conservative potential $V(x)$, $\gamma$ is the friction coefficient and $R(t)$ is the Langevin stochastic force. $R(t)$ represents the incessant collisions between the heavy particle and the small particles constituting the fluid. It is described by a stationary stochastic process whose distribution is Gaussian with zero mean ($\langle R(t)\rangle = 0$) and satisfying the white noise property:

$$\langle R(t)R(t+\tau)\rangle = 2\gamma m k_B T \delta(t-t') \tag{2.72}$$

As a consequence, the power spectral density of the random force, $C_{RR}(\omega)$ is given by:

$$C_{RR}(\omega) = 2m\gamma k_B T \tag{2.73}$$

The effect of a thermal bath is modeled by the friction $m\gamma\dot{x}$ and the stochastic $R(t)$ forces (fig. 2.9): the energy is supplied from the bath into the system via the random force $R(t)$ and it is extracted from the system via the friction force $m\gamma\dot{x}$. The correct thermal equilibrium is obtained when the random and frictions forces are equilibrated. This ensures that the classical fluctuation–dissipation theorem (FDT, see in the appendix section A.3) is fulfilled.

The $\gamma$ friction coefficient present in both the friction and stochastic force measures the strength of

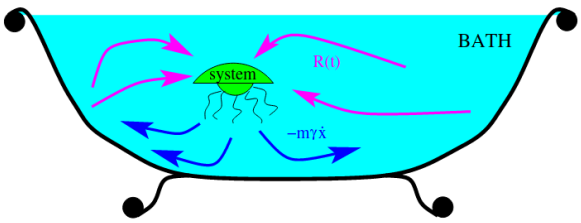



Figure 2.9: The Langevin equations eq. (2.71) as a thermal bath in classical molecular dynamics simulations. Reproduction from reference [91].

the coupling between the bath and the system. In principle this coupling should not have unwanted consequences on the results, provided it is not chosen too large.

### 2.6.2 Quantum thermal bath: the original formulation

To include quantum effects in MD simulations, Dammak and coworkers introduced in 2009 a modified version of the Langevin thermostat, the quantum thermal bath (QTB) method [13]. They used a modified Langevin equation, in which the random force $R(t)$ is characterized by the quantum FDT. The eq. (2.73) is modified in the QTB method via a power spectrum density $C_{RR}(\omega)$:

$$C_{RR}(\omega) = 2m\gamma\theta(\omega, T) \tag{2.74}$$

resulting in a colored stochastic force. In the above equation $\theta(\omega, T)$ corresponds to the quantum harmonic oscillator energy distribution:

$$\theta(\omega, T) = \hbar\omega \left[ \frac{1}{2} + \frac{1}{\exp(\hbar\omega/k_B T) - 1} \right] \tag{2.75}$$

In this way, the quantum statistical distribution of the energy, which includes the zero-point energy motion and quantum fluctuations, is enforced via a modified Langevin thermostat. In the limit of $\gamma \to 0$, the QTB gives the exact quantum distribution for a harmonic system by ensuring the quantum FDT. However, for anharmonic systems eq. (2.74) is not sufficient to enforce the quantum FDT. In this case, the coupling of the classical forces $-dV/dx$ with the different modes drives the system towards the classical equipartition of the energy. Therefore, the average thermal energy per mode calculated via QTB is generally lower than $\theta(\omega, T)$ when $\omega$ is large, while it is larger than $\theta(\omega, T)$ when $\omega$ is small. This is an unphysical phenomena, that consists in a energy redistribution from high to low frequency modes, and it is known as the zero-point energy leakage (ZPEL) [92]. The ZPEL can have a dramatic impact on the properties of the systems. For example, QTB cannot provide an adequate description of the solid-liquid phase transition of Neon clusters [14]. In this case, the potential is highly anharmonic with the particles mainly interacting via weak van der Waals forces. The energy leak due to the ZPEL causes an artifical excess of energy in the low frequency modes and completely destabilizes the cluster structure at low temperature, dissociating the weak interatomic bonds. Instead of a solid-like cluster at low temperature, the QTB predicts an abnormal liquid-like phase.

One way to treat the ZPEL is using high values for the friction coefficient $\gamma$ [92]. However, high values of $\gamma$ correspond to a strong coupling between the system and the bath, which can lead to an overdamped dynamic. The best approach is to chose $\gamma$ as a compromise to limit the ZPEL and a too strong coupling with the bath. Despite its drawbacks, QTB has been successfully applied for studying realistic systems [10, 93, 94].

### 2.6.3 Adaptive quantum thermal bath

An alternative and more accurate way to avoid ZPEL in QTB simulation is via its adaptive version (adQTB), recently introduced in 2019 [14]. This method has been successfully applied to the case study of Neon clusters, which we previously discussed, recovering the correct structure at low temperature [14]. The underlying idea of adQTB is to use the quantum FDT to quantify the amount of energy leaking from high to low frequencies and to adjust "on-the-fly" during the dynamics the parameters of the QTB thermostat. There are two ways to perform adQTB. One is by adjusting the coefficient of the friction force in eq. (2.71), by making it frequency-dependent and adapting it during the simulation. Within this adaptation method a Generalized Langevin Equation (GLE) is introduced [91], further details about it are illustrated in reference [14]. The other is via adaptation of the random force amplitude present in eq. (2.74). In this version of the adQTB (adQTB-r), which is the one we employed in this thesis, the power spectrum of $R(t)$, $C_{RR}(\omega)$ (eq. (2.74)) is replaced with:

$$C_{RR}(\omega) = 2m\gamma_r(\omega)\theta(\omega, T) \tag{2.76}$$

where $\gamma_r(\omega)$ are frequency-dependent coefficients which are adjusted "on-the-fly" during the simulation to enforce the quantum FDT. The friction parameter $\gamma$ is constant in the Langevin dynamics. The quantum FDT is rewritten as:

$$\mathrm{Re}[C_{vR}(\omega)] = m\gamma_r(\omega)C_{vv}(\omega) \tag{2.77}$$

in which $C_{vR}$ is the velocity-random force correlation function and $C_{vv}$ is the velocity-velocity autocorrelation function. By defining $\Delta_{\mathrm{FDT}}(\omega)$:

$$\Delta_{\mathrm{FDT}}(\omega) = \mathrm{Re}[C_{vR}(\omega)] - m\gamma_r(\omega)C_{vv}(\omega) \tag{2.78}$$

The quantum FDT is enforced for $\Delta_{\mathrm{FDT}} = 0$ for all $\omega$. In this way, the ZPEL for each frequency can be directly estimated and corrected by adaptation of the coefficients $\gamma_r(\omega)$. The general procedure is the following. The dynamics is divided in segments of a few hundreds fs. The simulation begins with $\gamma_r(\omega) = \gamma$, then $\Delta_{\mathrm{FDT}}$ is estimated at each segment of the dynamics via eq. (2.78) and $\gamma_r(\omega)$ is adjusted to have $\Delta_{\mathrm{FDT}} = 0$ (if $\Delta_{\mathrm{FDT}} < 0$ , $\gamma_r(\omega)$ is decreased, while for $\Delta_{\mathrm{FDT}} > 0$, $\gamma_r(\omega)$ is increased). Typically, $\gamma_r$ should increase for the high-frequency modes (that loose energy through ZPEL), and decrease at low frequencies.

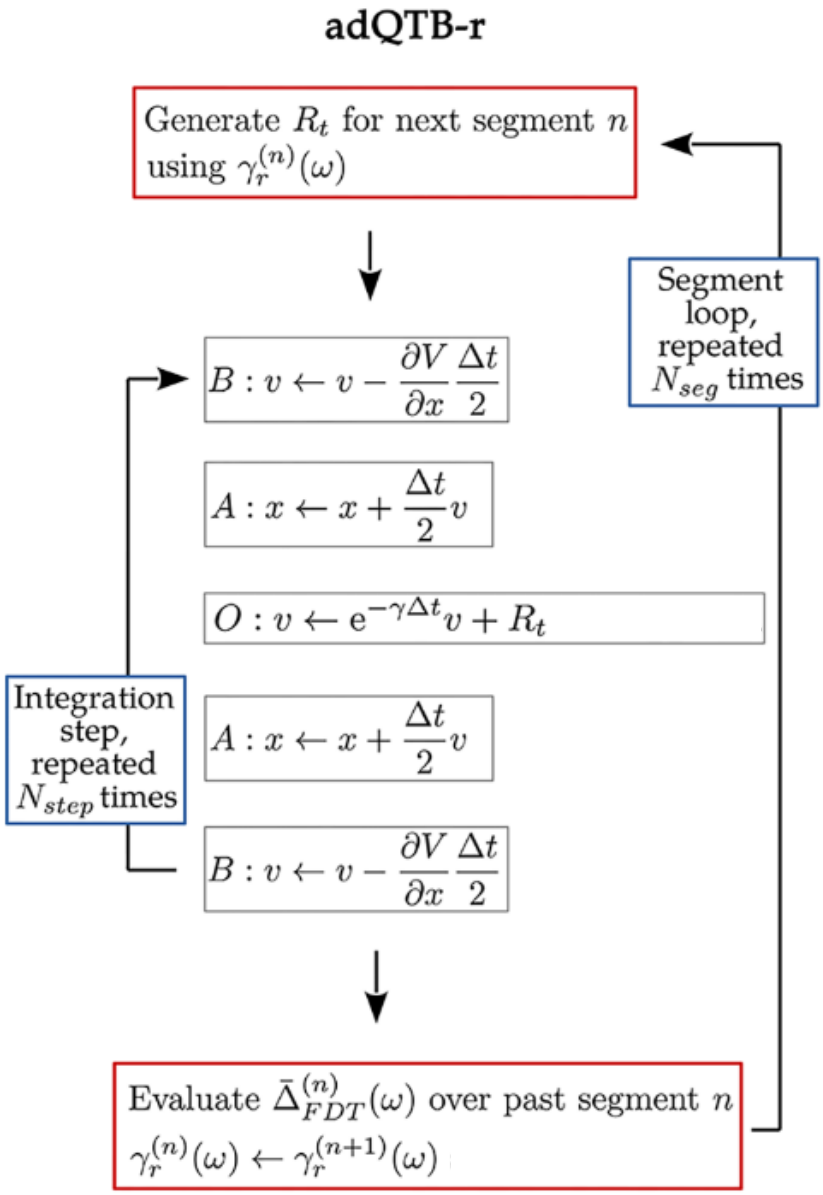



Figure 2.10: Flux diagram of the adQTB-r algorithm. The figure describes the modifications introduced with respect to the standard BAOAB algorithm for the Langevin equation [95, 96]. The adQTB trajectories are decomposed in a series of $N_{seg}$ segments of $N_{step}$ time steps each; the segments have a duration $\tau = N_{step}\Delta t$. One-dimensional notations are used for simplicity. Reproduction from reference [14].

# APPENDIX A

## A.1 Derivation of Feynman path integral

The matrix elements of the quantum propagator for a one-dimensional quantum particle evolving from the initial position $x$ to the final position $x'$ in an elapsed time $t$ is given by:

$$U(x', x, t) = \langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|x\rangle \tag{A.1}$$

By recalling that the kinetic and potential operator do not commute:

$$[K(\hat{p}), V(\hat{x})] \neq 0 \tag{A.2}$$

To overcome this issue, we can use the Trotter product formula, which states that for arbitrary matrices $A$ and $B$,

$$e^{A+B} = \lim_{n\to\infty}(e^{A/n}e^{B/n})^n \tag{A.3}$$

The application of the Trotter formula to the quantum propagator operator gives:

$$e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t} = \lim_{n\to\infty}[e^{-\frac{i}{\hbar}K(\hat{p})\tau}e^{-\frac{i}{\hbar}U(\hat{x})\tau}]^n \tag{A.4}$$

where $\tau = t/n$. We define $e^{-\frac{i}{\hbar}K(\hat{p})\tau}e^{-\frac{i}{\hbar}U(\hat{x})\tau} = \hat{\Omega}$. The propagator matrix elements become:

$$\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|x\rangle = \lim_{n\to\infty}\langle x'|\hat{\Omega}^n|x\rangle = \lim_{n\to\infty}\langle x'|\hat{\Omega}\hat{\Omega}\hat{\Omega}\cdots\hat{\Omega}|x\rangle \tag{A.5}$$

The set of states $\int dx\,|x\rangle\langle x| = \mathbb{I}$ can be inserted between each factor $\hat{\Omega}$, i.e. $(n-1)$ insertions.

$$\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|x\rangle = \lim_{n\to\infty}\int dx_1\cdots dx_{n-1}\,\langle x_n|\hat{\Omega}|x_{n-1}\rangle\langle x_{n-1}|\cdots|x_1\rangle\langle x_1|\hat{\Omega}|x_0\rangle \tag{A.6}$$

where the intermediate integrations over $x_1, ..., x_{n-1}$ constitute the sum over all possible paths linking the two endpoints $x$ and $x'$.

Consider the matrix element $\Omega(x_{j+1}, x_j)$:

$$\Omega(x_{j+1}, x_j) = \langle x_{j+1}|\hat{\Omega}|x_j\rangle = \langle x_{j+1}|e^{-\frac{i}{\hbar}K(\hat{p})\tau}e^{-\frac{i}{\hbar}U(\hat{x})\tau}|x_j\rangle \tag{A.7}$$

To simplify, the identity $\int dp\,|p\rangle\langle p| = \mathbb{I}$ can be inserted.

$$\begin{aligned}\langle x_{j+1}|\hat{\Omega}|x_j\rangle &= \int dp\,\langle x_{j+1}|e^{-\frac{i}{\hbar}K(\hat{p})\tau}|p\rangle\langle p|e^{-\frac{i}{\hbar}U(\hat{x})\tau}|x_j\rangle \\ &= \int dp\,e^{-\frac{i}{\hbar}\frac{p^2}{2m}\tau}\langle x_{j+1}|p\rangle\langle p|x_j\rangle e^{-\frac{i}{\hbar}U(x_j)\tau}\end{aligned} \tag{A.8}$$

By recalling:

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{\frac{i}{\hbar}px} \tag{A.9}$$

The relation eq. (A.9) is substituted into eq. (A.8), giving:

$$\langle x_{j+1}|\hat{\Omega}|x_j\rangle = \frac{1}{2\pi\hbar}\int dp\,e^{\frac{i}{\hbar}\left[-\frac{\tau}{2m}p^2+(x_{j+1}-x_j)p-\tau U(x_j)\right]} \tag{A.10}$$

The above integral is a Gaussian integral that can be solved in the following way:

$$\int dx\,e^{-ax^2+bx} = \sqrt{\frac{\pi}{a}}e^{b^2/4a} \tag{A.11}$$

By defining:

$$a = \frac{i\tau}{2m\hbar} \qquad b = \frac{i}{\hbar}(x_{j+1}-x_j) \tag{A.12}$$

Eq. (A.10) becomes:

$$\langle x_{j+1}|\hat{\Omega}|x_j\rangle = \left(\frac{m}{2\pi i\hbar\tau}\right)^{1/2}e^{\frac{i}{\hbar}\tau\left[\frac{m}{2\tau^2}(x_{j+1}-x_j)^2-U(x_j)\right]} \tag{A.13}$$

Now, the matrix element $\Omega(x_{j+1}, x_j)$ of eq. (A.13) are substituted into eq. (A.6):

$$\langle x'|e^{-\frac{i}{\hbar}\hat{\mathcal{H}}t}|x\rangle = \lim_{n\to\infty}\left(\frac{m}{2\pi i\hbar\tau}\right)^{n/2}\int dx_1\cdots dx_{n-1}\sum_{j=0}^{n-1}e^{\frac{i}{\hbar}\tau\left[\frac{m}{2}\left(\frac{x_{j+1}-x_j}{\tau}\right)^2-U(x_j)\right]} \tag{A.14}$$

## A.2 Stationary phase approximation

Suppose one wants to calculate the integral of one-dimensional oscillator:

$$I \equiv \int dx\, e^{i\lambda f(x)} \tag{A.15}$$

where $\lambda$ is a parameter. When $\lambda$ is very large, the integral is highly oscillatory and it is dominated by the stationary points $\bar{x}$ which gives:

$$f'(\bar{x}) = 0 \tag{A.16}$$

A good approximation is to expand the phase around the stationary point $\bar{x}$ at the second order:

$$f(x) = f(\bar{x}) + \frac{1}{2} f''(\bar{x})(x - \bar{x})^2 \tag{A.17}$$

in which the first order term $f'(\bar{x}) = 0$. This approximation is valid if $f''(\bar{x})$ is not too small. By setting $y = (x - \bar{x})$, the integral of eq. (A.15) becomes a Gaussian integral:

$$\int dx\, e^{i\lambda f(x)} \approx e^{i\lambda f(\bar{x})} \int dy\, e^{iy^2 \lambda f''(\bar{x})/2} = \sqrt{\frac{2\pi i}{\lambda f''(\bar{x})}} e^{i\lambda f(\bar{x})} \tag{A.18}$$

In the case there is more than one stationary point, eq. (A.18) becomes:

$$\int dx\, e^{i\lambda f(x)} = \sum_{\{\bar{x}_j | f'(\bar{x}_j)=0\}} \sqrt{\frac{2\pi i}{\lambda f''(\bar{x}_j)}} e^{i\lambda f(\bar{x}_j)} \tag{A.19}$$

By applying the previous assumptions, the SPA can be generalized to $n$-dimensional integral over coordinate $x = (x_1, x_2, \cdots, x_n)$. The following integral is considered:

$$I \equiv \int d^n x\, e^{i\lambda f(x)} \tag{A.20}$$

The function $f(x)$ is expanded at the second order:

$$f(x) = f(\bar{x}) + \frac{1}{2} \sum_{k,l} (x_k - \bar{x})(x_l - \bar{x}) \frac{\partial^2 f(\bar{x})}{\partial x_k \partial x_l} \tag{A.21}$$

in which $\partial^2 f(\bar{x})/\partial x_k \partial x_l$ constitutes the Hessian matrix of $f$ at $\bar{x}$. Therefore, the final expression of SPA for a $n$-dimensional integral is:

$$\int d^n x\, e^{i\lambda f(x)} = \sum_{\{\bar{x}_j | f'(\bar{x}_j)=0\}} \left(\frac{2\pi i}{\lambda}\right)^{n/2} \det \left(\frac{\partial^2 f(\bar{x}_j)}{\partial x_k \partial x_l}\right)^{-1/2} e^{i\lambda f(\bar{x}_j)} \tag{A.22}$$

## A.3 Fluctuation-dissipation theorem

Linear response theory provides a mathematical relationship - the fluctuation-dissipation theorem - that connects the response of a system to an external perturbation and the fluctuations about the equilibrium in absence of the perturbation. A general form of the fluctuation-dissipation theorem reads as:

$$C_{vv}(\omega) = 2k_B T \, \mathrm{Re}[\chi(\omega)] \, \Phi(\beta, \omega) \tag{A.23}$$

where $C_{vv}(\omega)$ is the Fourier transform of the velocity-velocity autocorrelation function, $\chi(\omega)$ is the susceptibility characterizing the linear response of velocity $\Delta v(t)$ to a small perturbative force $\Delta F(t)$:

$$\Delta v(\omega) = \chi(\omega) \Delta F(\omega) \tag{A.24}$$

The term $\Phi(\beta, \omega)$ is a distribution function of the thermal energy. For the classical case, where there is equipartition of the energy, $\Phi(\beta, \omega) = 1$, while for quantum systems:

$$\Phi(\beta, \omega) = \frac{\beta\hbar\omega}{2} \coth\left(\frac{\beta\hbar\omega}{2}\right) \tag{A.25}$$

Finally, we notice that $2\mathrm{Re}[\chi(\omega)]$ corresponds to the vibrational density of states in the case of a harmonic system having an energy distribution in which a vibrational mode at frequency $\omega$ is thermalized with an average kinetic energy $\Phi(\omega) k_B T/2$.

# Part II

# Selected applications on crystals, adsorbed molecules and nucleobases

CHAPTER 3

# THE ANTIFERROELECTRIC TO PARAELECTRIC PHASE TRANSITION IN POTASSIUM HYDROXIDE

## 3.1 Introduction

In 1920, ferroelectricity was discovered for the first time by Valasek in the Rochelle salt crystal [97]. Ferroelectricity is the spontaneous electric polarization of certain materials, that can be switched under the application of an external electric field. Antiferroelectricity is intrinsically related to ferroelectricity and can be explained in terms of the microscopic order of the dipoles. While in ferroelectric materials the dipoles are parallel, leading to an overall spontaneous polarization, in antiferroelectric materials the dipoles are arranged antiparallel generating no macroscopic polarization. Typical hysteresis loops for conventional ferroelectric, antiferroelectric and paraelectric materials are represented in fig. 3.1. Over the years, research interest in the topic has grown rapidly, motivated by the large number of applications involving these materials. These include, for example, dynamical capacitors, non-linear optics, volatile memories and high energy storage devices.

The family of perovskites $ABO_3$ is one of the most studied type of ferroelectrics. A common practise in this field is to model the temperature-dependent properties of the material via effective Hamiltonians, developed from first principle calculations [98]. These are functions of the order parameter governing the phase transition and, generally, also of other quantities describing the coupling with other degrees of freedom. In the Landau theory of phase transitions, the high-symmetry phase, which corresponds to the paraelectric one, can be described by a single well potential, while the lower-symmetry ordered phase by a double well potential. Among the other types of ferroelectrics, a special attention has been given to materials presenting hydrogen bonds – hydrogen-bonded ferroelectrics. Conventional hydrogen-bonded ferroelectrics are associated to displacive phase transitions but a coexistence of both order-disorder and displacive effects is very

common [99]. Potassium dihydrogen phosphate (KDP) and its deuterated analog, being characterized by the presence of strong hydrogen bonds (O-O $< 2.5$ Å), have been widely studied during the years, along with related similar materials (KDP-crystals).

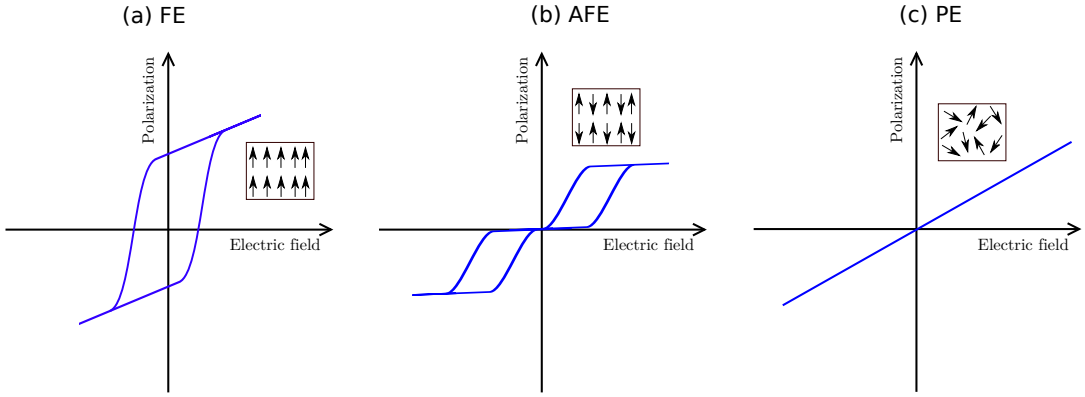



Figure 3.1: Schematic representation of different dielectric response to external electric field: (a) ferroelectric (FE), (b) antiferroelectric (AFE), (c) paraeletric (PE) materials.

More generally, nuclear quantum effects can have a significant impact on the properties of ferroelectrics and antiferroelectric systems. One critical example is quantum paraelectricity, which is the suppression of low temperature ordering phase transition by quantum fluctuations [100]. Concerning hydrogen bonded materials, the influence of the hydrogen bonds can be highlighted by replacing H with D isotope. A giant hydrogen/deuterium isotope effect has been found in KDP due to the increase of about 100 K of the Curie temperature upon deuteration [101]. Besides KDP crystals, there are other hydrogen-bonded ferroelectrics that manifest interesting properties. Alkali hydroxides (AOH, A = Li, Na, K, Rb, Cs) and their deuterated analogs (AOD) are characterized by proton ordering phase transitions at ambient pressure and in temperature-range 150-310 K – except LiOH/LiOD and NaOH – with antiferroelectric order in NaOD, KOH/KOD and CsOH/CsOD and ferroelectric order in RbOH/RbOD [102–115]. Differently from the KDP crystals, AOX crystals present weak and long hydrogen bonds, which are oriented along $b$-axis direction of a monoclinic or orthorombic structure (see for example the structure of proton ordered KOH in fig. 3.2).

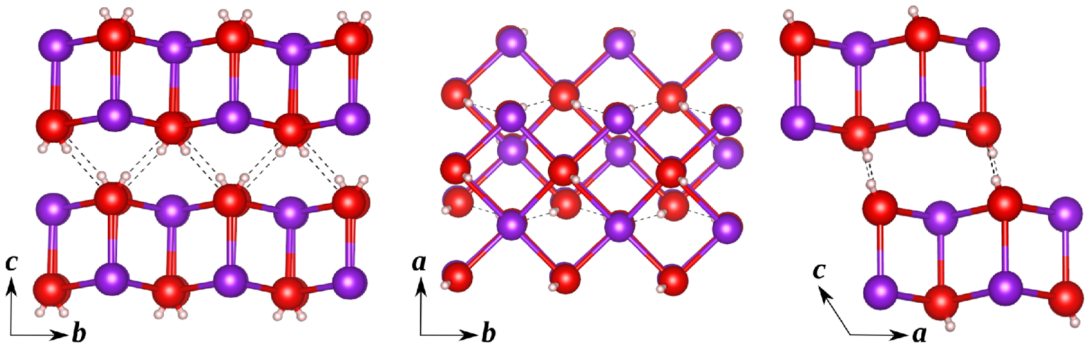



Figure 3.2: Sketch of the monoclinic potassium hydroxide structure in an antiferroelectric configuration. K, O and H atoms are respectively colored purple, red and white with ionic radii according to reference [116]

Regarding the NaOH crystal, it is important to stress that the protonated material manifests no or-

dering phase transition while the deuterated does. This anomaly can be explained in terms of the different zero-point energy quantum fluctuations of NaOH and NaOD [117]. In the protonated compound NaOH, the large zero-point fluctuations overcome the weak hydrogen bonds and destroy the antiferroelectric order. The system in this phase can be described as a as quantum paraeletric. On the contrary, when quantum fluctuations are reduced (NaOD), the system is driven into an antiferroelectric phase at low temperature. K, Rb and Cs hydroxides display a less striking isotope effect, where the critical temperature of the deuterated crystal is larger than the protonated one. Beyond the isotope effect, pressure has a crucial role since it can favor the same kind of phase transitions also in the hydroxides, that do not display ordering at ambient pressure [118–122]. In NaOH, for example, the external pressure has the same effect as deuteration, so that an antiferroelectric phase transition can be recovered [117].

In the present chapter we address, among the different AOX crystals, the ambient pressure low-temperature antiferroelectric to high-temperature paraelectric phase transition of potassium hydroxide and deuteroxide, phase-IVa (AFE) $\leftrightarrow$ phase-II (PE).[1] This phase transition has been characterized via different experimental techniques including neutron scattering and x-ray powder diffraction [113, 114, 123, 124], Raman and infrared spectroscopy [125–127], calorimetric [108, 113] and NMR [115] measurements. The Curie temperature is at 233 K in KOH and shifts up to 257 K in KOD [113]. To explain the isotopic shift, a model of tunneling of the H/D atoms between two minima in a double-well potential was proposed [108, 113], but no microscopical description of the phase transition has been provided up to now. The H/D isotope effect on $T_c$ cannot be clearly described in terms of classical dynamics. Therefore, in this study we highlight how the description including both thermal and nuclear quantum effects is essential to unravel the nature and mechanism of the phase transition [128]. Nuclear quantum effects were modeled via path integral molecular dynamics scheme and electronic structure calculations were performed using DFT/PBE. Firstly, the assignment of space groups of the low-temperature (low-T) and high-temperature (high-T) phases will be discussed based on the experimental literature. Secondly, the potential energy surface for monoclinic potassium hydroxide will be illustrated. A particular focus will be given on the structural instabilities of the crystal and the identification of the order parameter of the phase transition. Then, we will discuss the thermal and quantum description beyond the T = 0 K picture for the structural properties, the proton/deuteron order and the hydrogen bonds features. Finally, the model of the phase transition via adQTB is compared to path integral results. The present study can be used as a general reference for mild isotope effects in alkali hydroxides proton-ordering phase transition.

## 3.2 Symmetry of phase IVa and phase II

The first x-ray powder diffraction analysis conducted on KOH at standard conditions was made by Ibers and Kumamoto in 1960 [124], which assigned to the hydroxide a monoclinic structure with symmetry $P2_1$. The determination of heavy atoms positions suggested a monoclinically distorted NaCl arrangement with oxygen atoms forming a zig-zag chain parallel to $b$-axis. In the following years, Jacobs and coworkers corrected through x-ray powder diffraction $P2_1$ to $P2_1/m$ space group, which presents the (a)-(c) mirror plane as an additional element of symmetry [123]. However, x-

[1] The notation of the phases is the one used by Krobok and Holzapfel [119].

ray diffraction was unable to determine the hydrogens positions of KOH at room temperature. The great contribution to the issue was the discovery of the low-temperature phase transition of KOH by Bastow and coworkers in 1986 [113]. Both low-temperature (IVa) and room-temperature (II) phases were structurally characterized through neutron powder diffraction by which, thanks to the advantage of larger sensitivity to light atoms, hydrogen positions were determined. The existence of zig-zag chains of oxygen atoms along $b$-axis was confirmed in both phases and, in addition, hydrogen atoms were found to form weak asymmetric hydrogen bonds O–H$\cdots$O which hold the chains together (see fig. 3.2). The new phase (IVa) was assigned to $P2_1/a$ space group. Other neutron powder diffraction and calorimetric studies on the deuterated hydroxide, KOD, between 16 K and its melting point 646 K confirmed the phase transition [114]. Furthermore, it was found that at high temperatures (above 523 K in KOD), the hydrogen bonds break and the overall structure is cubic NaCl type [114].

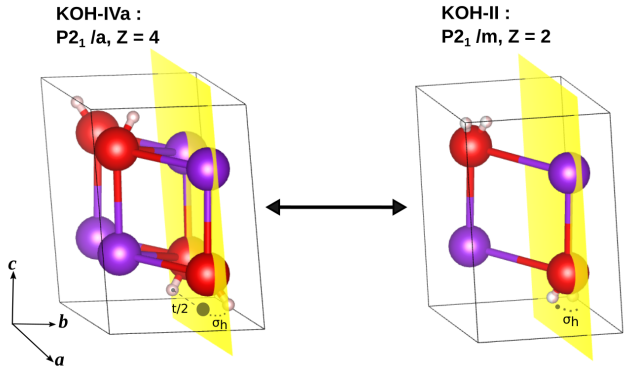



Figure 3.3: Sketch of the symmetry changing for KOH-IVa↔KOH-II phase transition using crystallographic data from reference [113].

To summarize, the low temperature phase transition in KOH and KOD crystals involves two phases: the ordered low-T antiferroelectric phase (phase IVa) and the disordered high-T paraelectric phase (phase II). The two phases are monoclinic with phase IVa having $P2_1/a$ space group and $Z = 4$ and phase II having $P2_1/m$ space group and $Z = 2$. The hydrogen atoms of phase II were identified with half-occupancy at Wyckoff sites $4f$, which are related to the mirror plane $y = \pm 1/4$. A dynamical disorder for the hydrogen positions along this mirror plane was hypothesized. The transition from phase II to phase IVa causes a breaking of the symmetry from mirror plane to glide plane along $a$-axis (fig. 3.3).

## 3.3 Determination of the potential energy surface

In order to obtain a first picture of the relevant geometrical parameters for the phase transition, we have performed geometry optimization calculations at T = 0 K, without including any thermal and nuclear quantum effects. Three monoclinic arrangements can be identified: ferroelectric, antiferroelectric and paraelectric (see fig. 3.4). The structure of monoclinic KOH is in agreement with the experimental findings. A bilayered structure of K and O heavy atoms is formed along the $c$-axis, which is a common feature for all three phases. The difference between each configuration is determined by the orientation of the OH groups. The ordered FE and AFE phases are characterized by dipole chains of OH groups along the $b$-axis. Neighboring OH dipoles chains with respect

to the $a$-axis can have the same (FE) or opposite (AFE) orientations. The overall bilayered structure of KOH is stabilized by the presence of weak hydrogen bonds (O-O = 3.0 Å in both FE and AFE) along the dipoles chains. On the contrary, in the PE phase there is no formation of hydrogen bonds and the OH groups are oriented normally to the $b$-axis direction over Wickoff positions $2e$, corresponding to fractional coordinates $\pm 1/4$.

Table 3.1: Lattice parameters and O-O distance of monoclinic potassium hydroxide from DFT optimization at T = 0 K. $c \sin \beta$ represents the projection of the $c$-axis along $z$-axis. Lengths are reported in Å and angles in degrees.

| | | $a$ | $b$ | $c$ | $\beta$ | $c \sin \beta$ | O-O |
|---|---|---|---|---|---|---|---|
| Exp. [113] | | 7.892 | 3.945 | 5.947 | 114.24 | 5.423 | (not reported) |
| PBE | FE | 3.982 | 4.009 | 5.568 | 107.674 | 5.305 | 3.00 |
| | AFE | 7.958 | 4.014 | 5.770 | 113.230 | 5.302 | 3.00 |
| | PE | 3.963 | 4.028 | 5.782 | 101.634 | 5.663 | 3.47 |
| PBE+D2 | FE | 3.885 | 4.024 | 5.414 | 111.023 | 5.054 | 2.82 |
| | AFE | 7.767 | 4.026 | 5.415 | 111.046 | 5.054 | 2.82 |
| | PE | 3.903 | 3.954 | 5.569 | 102.535 | 5.436 | 3.26 |
| PBE+D3 | FE | 3.928 | 4.033 | 5.525 | 110.928 | 5.161 | 2.90 |
| | AFE | 7.852 | 4.037 | 5.527 | 110.987 | 5.160 | 2.90 |
| | PE | 3.945 | 4.013 | 5.679 | 101.882 | 5.557 | 3.39 |

The geometrical features of monoclinic KOH can be better understood by the analysis of the optimized lattice constants at T = 0 K. Table 3.1 report the lattice constants and O-O distance between the layers, i.e. the distance between the donor and acceptor hydrogen bonds sites, using different levels of DFT: PBE *vs* PBE plus the inclusion of Van der Walls corrections via Grimme either two-body (D2) [129, 130] or three-body (D3) [131] techniques. The relevant structural parameters are mainly $c$ and $\beta$. In particular, the projection of the $c$-vector along the $z$-axis, i.e. $c \sin \beta$, is directly correlated to the O-O distance. $c \sin \beta$ indeed measures the inter-layer distance and, therefore, the strength of the hydrogen bonds. The experimental $c \sin \beta$ of the IVa phase, calculated from reference [113], is equal to 5.423 Å. The AFE configuration shows a smaller inter-layer distance than the experimental value by 0.12 Å, within the PBE approximation, which worsens when including corrections D2 and D3. The general trend is that D2 and D3 approaches strengthen the hydrogen bonds (O-O distances are shorter) and, therefore, $c \sin \beta$ is smaller. We did not proceed to further advanced calculations using D2 and D3 corrections, since they worsen accordance with the experiments. Lastly, we highlight that the FE configuration has the same $c \sin \beta$ as the AFE in all the three approaches, meaning that there is no substantial difference between AFE and FE in the strength of hydrogen bonds. Regarding the PE phase, $c \sin \beta$ is significantly larger compared to the hydrogen-bonded phases. This is due to the fact that in the PE phase the inter-layer distance is larger and no formation of hydrogen bonds is observed.

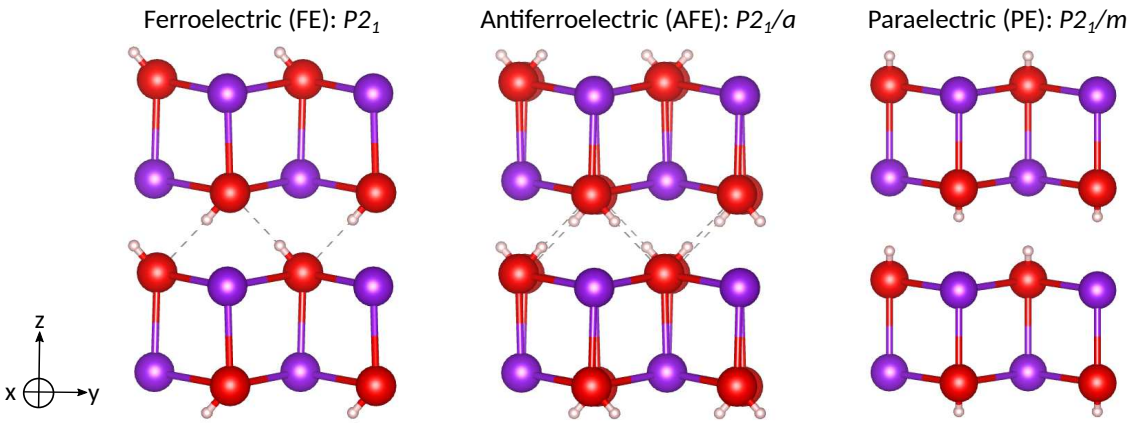



Figure 3.4: *Static* FE, PE, and AFE structures of monoclinic KOH.

The identification of the three monoclinic structures was a first step for reconstructing the energy landscape describing the phase transition. The analysis of the topology of the PES is a useful tool to understand at the microscopic scale the transformations in the crystal at T = 0 K. To reconstruct the PES we have used the following strategy. Several minimum energy paths by varying initial and final conditions were optimized via climbing nudged elastic band (CI-NEB) method [132]. The starting and final points for the NEB calculations were the FE, AFE and PE configurations.[2] The resulting energy landscape is illustrated in fig. 3.6-a, where the PES is plotted as a function of two reaction coordinates, $\theta_x$, and $\theta_{x+1/2}$. These correspond to the OH polar angle in the $(y, z)$ plane – defined as $\arctan(\theta) = \mathrm{OH}_y/|\mathrm{OH}_z|$ – at position $x$ and position $x + 1/2$, respectively. The orientation of the OH dipoles can be either clockwise, i.e. positive $\theta$, or anticlockwise, i.e. negative $\theta$.

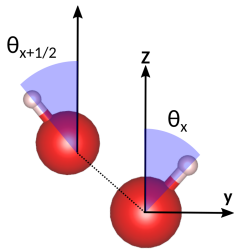



Figure 3.5: Sketch of consecutive $\theta$ angles in the $x$-axis direction. Their correlation characterizes the order of the phase transition.

A positive (negative) $\theta_x$,$\theta_{x+1/2}$ correlation implies the FE (AFE) arrangement, respectively. The case $\theta_x = \theta_{x+1/2} = 0^\circ$ corresponds to the pure paraelectric case. It is clear from the PES that FE and AFE structures correspond to local minima of the PES, while the PE phase is a local maximum. The FE phase is less stable than AFE by 3 meV, due to the increase in repulsion between the OH groups in parallel alignment.

The transition AFE↔FE can either involve a path through PE phase or directly through a transition state (TS). This is 0.18 eV higher in energy than the most stable minimum (AFE) and corresponds to a configuration in which one chain is very close to its slanted equilibrium position, while the

[2] See in the appendix, section B.1 for a detailed description of the procedure.

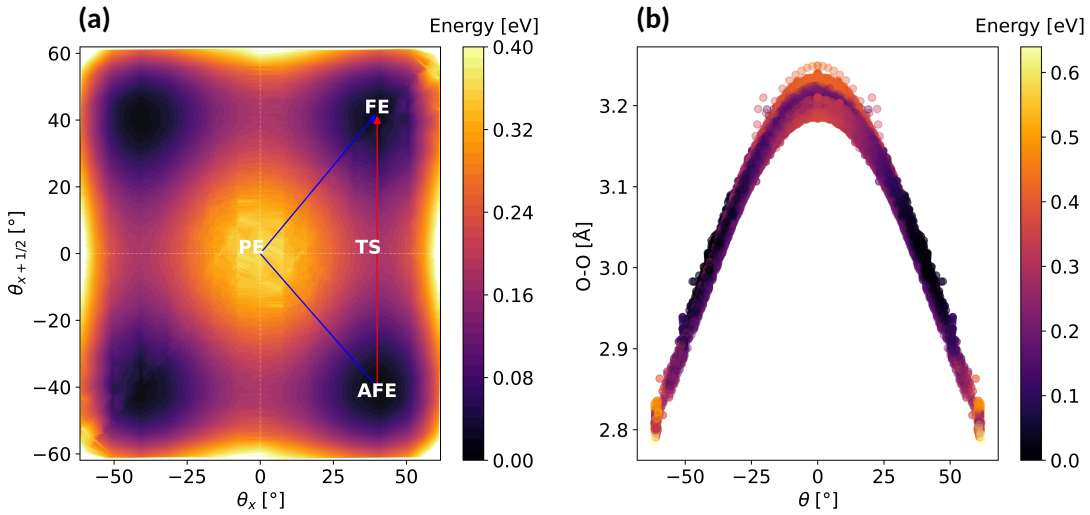



Figure 3.6: (a) Locally reconstructed potential energy surface with respect to $\theta_x$ and $\theta_{x+1/2}$ variables computed at constant volume (see lattice parameters of AFE phase at T = 0 K). (b) Correlation between the OH polar angle $\theta$ and the O-O distance along the PES. The color bar corresponds to the potential energy.

other chain has the OH dipole very close to crystal position $y = \pm 1/4$ which corresponds to the PE phase. The energy barrier for this direct path is actually close to half the total energy of the PE phase ($E_{PE}$ = 0.34 eV). Therefore, the *static* barrier for flipping of a single dipole chain ($\simeq$0.09 eV) would correspond to a rather high temperature ($\sim$ 1000 K).

Finally, by using an averaged value for $\theta$, we can look at the correlation between $\theta$ and the O-O distance (fig. 3.6-b). The two parameters are clearly strongly correlated. At values of $\theta$ corresponding to the equilibrium FE and AFE values ($30° \leq |\theta| \leq 40°$), the energy decreases with the O-O distance between 2.9 and 3.0 Å. For $\theta \to 0°$, the O-O distance approaches 3.2 Å, being in the PE or TS regime. The general trend is that long and weak inter-layer hydrogen bonds (large O-O) are associated to small OH bending angles.

## 3.4 Instability of static paraelectric phase

To better understand the instability of the *static* PE phase and its role on the equilibrium of FE and AFE configurations, we have firstly studied the lattice dynamics via the harmonic approximation using density functional perturbation theory (DFPT) [133]. Figure 3.7 reports the IR harmonic frequencies of FE, AFE and PE. As expected the PE phase, being a local maximum along the PES, has an imaginary frequency at $242i$ cm$^{-1}$. This frequency corresponds to a mode characterized as an O-H libration parallel to the $y$-axis direction (see fig. 3.8-a). Therefore the coordinate $\theta$ is the actual adequate parameter describing the phase transition, since it is associated with the bending motion of OH groups in the $(x, y)$ plane. By using a frozen phonon approach, the energy variation due to the displacements of the atoms along this mode can be computed at point $\Gamma$. As represented in fig. 3.8-a, the energy decreases towards a displacement along $y$-axis corresponding to $\sim \pm 0.12$ Å, an ordered arrangement of the dipoles. For simplicity, this displacement is represented with the PE phase having $Z = 2$ so that, by following the unstable OH mode, the system is driven to an

ordered FE state.

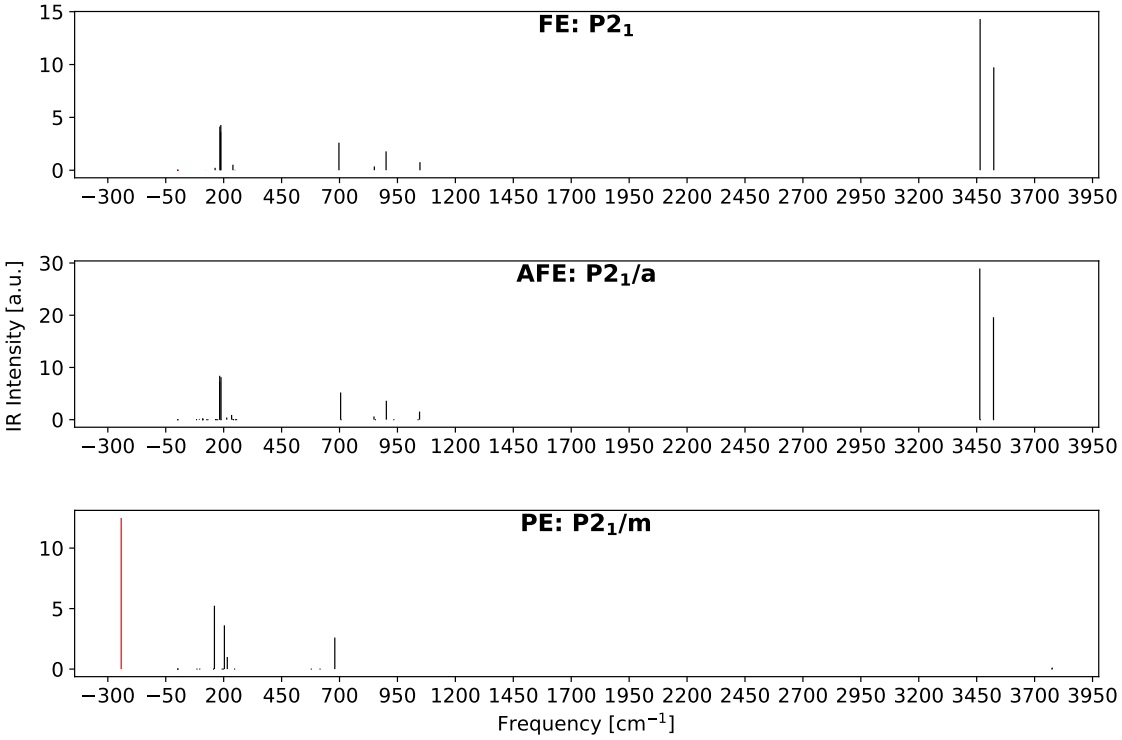



Figure 3.7: Harmonic IR spectrum for monoclinic KOH in the *static* FE, AFE and PE phases. The imaginary mode is represented with negative values.

Table 3.2: The diagonal elements of the effective charge tensor $Z^\star$ for monoclinic potassium hydroxide from DFPT calculations.

| | FE: $P2_1$ | | | AFE: $P2_1/a$ | | | PE: $P2_1/m$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $Z^\star_{xx}$ | $Z^\star_{yy}$ | $Z^\star_{zz}$ | $Z^\star_{xx}$ | $Z^\star_{yy}$ | $Z^\star_{zz}$ | $Z^\star_{xx}$ | $Z^\star_{yy}$ | $Z^\star_{zz}$ |
| K | 0.96 | 0.91 | 0.97 | 0.96 | 0.91 | 0.97 | 0.94 | 0.90 | 1.00 |
| O | −1.13 | −1.42 | −1.33 | −1.13 | −1.42 | −1.34 | −1.08 | −1.31 | −1.07 |
| H | 0.18 | 0.51 | 0.37 | 0.18 | 0.51 | 0.37 | 0.15 | 0.41 | 0.07 |

The corresponding dynamical charges for PE phase are $Z^\star_{yy}(\mathrm{H}) = 0.41$ and $Z^\star_{yy}(\mathrm{O}) = -1.31$; the unstable O-H libration mode therefore yields a non-null polarization $\parallel y$ and is IR active with a rather high intensity. Also noteworthy is the fact that the dynamical charges of H and O substantially differ between the PE, on the one side, and the FE/AFE configurations, on the other side; the latter are greater, which implies that the hydrogen bond can be easily polarized and that the dielectric constant is much enhanced in both FE and AFE configurations.

Another interesting way to describe the lattice dynamics of PE phase is by computing the phonon dispersion curves in a sufficiently large portion of the Brillouin zone (fig. 3.8-b). The O-H unstable bending mode is imaginary (negative) along the path Γ-Y-A-B, where $k_y = 0$, while it is real along B-D-E-C-Z path, where $k_y = \pi/b$. This means that real-space instability is associated with in-phase collective displacements of OH groups along the $y$-axis direction, leading to a FE or AFE arrangement. The T = 0 K DFT calculations clearly exclude the *static* paraelectric configuration as

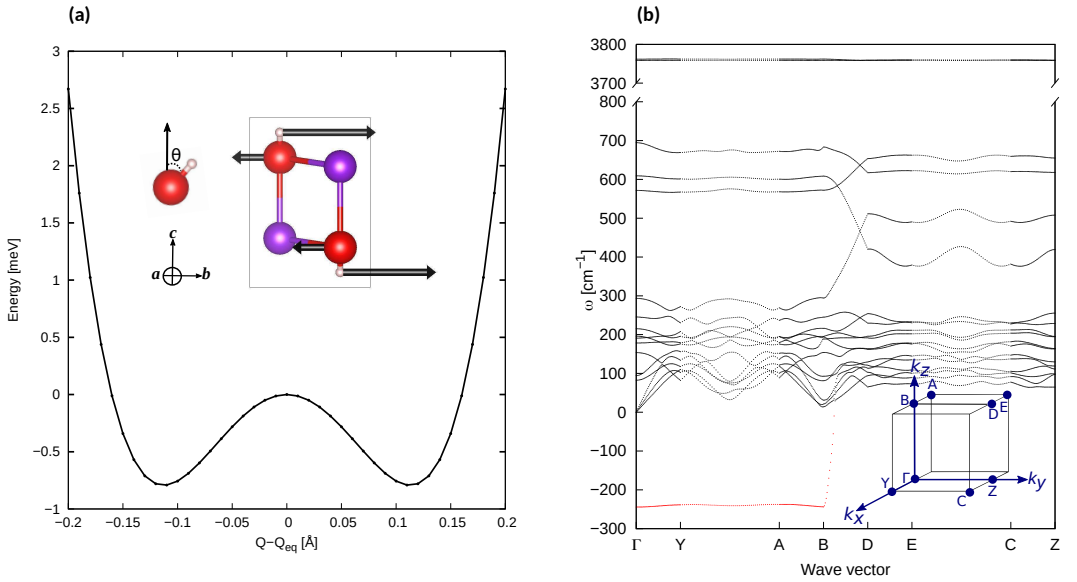



Figure 3.8: Panel (a): the variation of the total energy relative to the atoms displacements along the unstable O-H libration mode at point $\Gamma$ ($\nu = 242i$ cm$^{-1}$) in the *static* PE phase phase of monoclinic potassium hydroxide. Panel (b): representation of the phonon dispersion curves for the *static* PE phase of monoclinic potassium hydroxide, whose unit cell contains two formula units ($Z = 2$). The red branch corresponds to the unstable OH libration mode. Imaginary frequencies are plotted as negative.

a possible candidate for the high-temperature experimental phase II, due to its intrinsic instability.

## 3.5 Inclusion of thermal and nuclear quantum effects

### 3.5.1 Structural properties

Table 3.3 reports the experimental data of lattice constants and O-O distance in KOH and KOD with respect to temperature compared to our results from classical MD and PIMD at three temperatures: 77 K, corresponding to the IVa phase, 215 K a intermediate point close to the phase transition and 350 K, a temperature at which we can assert that KOH/KOD has made the transition from IVa phase to II phase. Starting from the experimental data, the IVa $\rightarrow$ II phase transition is characterized by a thermal expansion in both KOH and KOD crystals. This can be measured by $c \sin \beta$ which increases by $\sim$ 3% passing from phase IVa to phase II (see Table 3.3). By introducing thermal effects via classical molecular dynamics, a better agreement with the experimental $c \sin \beta$ at 77 K is obtained for the IVa phase with respect to the T = 0 K picture (see table 3.1). The agreement improves when introducing quantum effects via PIMD simulations. At high temperature, classical MD results are in good agreement with the experimental outcomes for phase II and the differences between the classical and KOH and KOD are less evident as compared to the low temperature case. Actually, KOH and KOD can be described using the same lattice parameters at high temperature, meaning that the isotope effect faints as the temperature grows. This is consistent with a classical regime at high temperature.

Table 3.3: Lattice parameters and O-O distance of monoclinic potassium hydroxide from *ab initio* calculations from classical MD (*cl*-KOH/D) and PIMD simulations (*q*-KOH and *q*-KOD), compared with the available experimental data. The reported experimental lattice parameters for the IVa phase are presented after a change of basis. Lengths are reported in Å and angles in degrees.[3]

| Experimental data | | $a$ | $b$ | $c$ | $\beta$ | $c\sin\beta$ | O-O |
|---|---|---|---|---|---|---|---|
| Phase IVa: $P2_1/a$, Z=4 | | | | | | | |
| 100 K | KOD[a] | 7.922 | 3.942 | 5.903 | 113.95 | 5.395 | 3.24 |
| 77 K | KOH[b] | 7.892 | 3.945 | 5.947 | 114.24 | 5.423 | (not reported) |
| Phase II: $P2_1/m$, Z=2 | | | | | | | |
| 300 K | KOD[a] | 3.965 | 3.999 | 5.728 | 104.23 | 5.552 | 3.45 |
| 293 K | KOH[b] | 3.951 | 3.999 | 5.750 | 103.58 | 5.589 | 3.33 |
| Theory (this work) | | $a$ | $b$ | $c$ | $\beta$ | $c\sin\beta$ | $\langle$ O-O $\rangle$ |
| 77 K | *cl*-KOH/D | 8.01 | 4.01 | 5.87 | 114.02 | 5.36 | 3.02 |
| | *q*-KOD | 8.04 | 4.02 | 5.98 | 114.25 | 5.45 | 3.09 |
| | *q*-KOH | 8.04 | 4.03 | 6.00 | 114.25 | 5.47 | 3.12 |
| 215 K | *cl*-KOH/D | 7.98 | 4.05 | 5.76 | 110.00 | 5.41 | 3.01 |
| | *q*-KOD | 8.02 | 4.05 | 5.86 | 111.00 | 5.47 | 3.05 |
| | *q*-KOH | 8.02 | 4.06 | 5.88 | 111.00 | 5.49 | 3.07 |
| 350 K | *cl*-KOH/D | 8.10 | 4.07 | 5.74 | 104.25 | 5.56 | 3.19 |
| | *q*-KOD | 8.15 | 4.08 | 5.77 | 104.25 | 5.59 | 3.21 |
| | *q*-KOH | 8.15 | 4.08 | 5.77 | 104.25 | 5.59 | 3.22 |

[a] From ref. [114]. For KOD-IVa at 100 K the transformation matrix is $\begin{pmatrix} 1 & 0 & \bar{1} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

[b] From ref. [113]. For KOH-IVa at 77 K the transformation matrix is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \bar{1} \end{pmatrix}$.

### 3.5.2 Proton/deuteron ordering

The identification of the potential energy surface (section 3.3) and the dynamical matrix calculations (section 3.4) allowed to identify the order parameter of the phase transition, the polar angle $\theta$. However, in the classical finite temperature picture, we cannot precisely evaluate the thermal and isotope effects characterizing the phase stability of KOH and KOD systems.

The flipping of OH and OD groups at 77 K, 215 K and 350 K is described in terms of the probability distribution of the order parameter $\theta$ in fig. 3.9. These show a double-peak profile with equilibrium values (the maximum in probabilities) located at $\sim \pm 40°$, which corresponds to hydrogen bonds chains that are $\parallel \pm y$. The energy barrier corresponds to $\theta \to 0°$, a minimum in the probability. The population of states close to the barrier is very low at 77 K and the flipping events of OH and OD groups leading to opposite orientations are rare, consistently with the rather large energy barrier as found in the NEB calculations. By increasing the temperature, the $\theta \to 0°$ is much more probable. At all three temperatures, we notice that the PIMD KOD distributions are intermediate

[3] After a revision of our data, we found a discrepancy with respect the O-O distances. The table was corrected with respect to reference [128].

between the PIMD KOH and the classical ones, consistently with the fact that KOD behaves more classically than KOH.

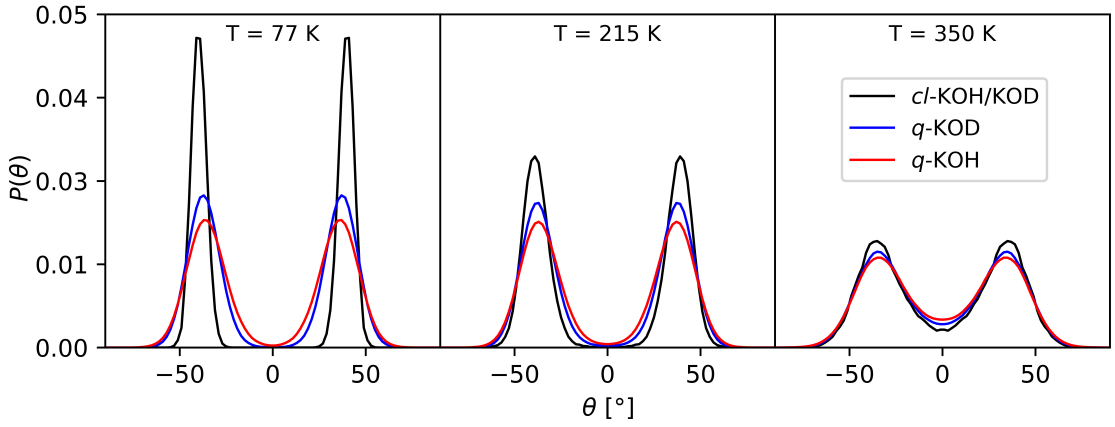



Figure 3.9: Probability distribution for the microscopic order parameter $\theta$ from *ab initio* classical (*cl*-KOH/D) and PIMD simulations (*q*-KOH and *q*-KOD). The distributions have been symmetrized.

In order to understand the flipping mechanism of the OH and OD groups, we have analyzed the trajectories of the centroid during the dynamics. Figure 3.10 represents the OH motion in KOH projected in the three different planes of the 3D space at 77 K, when there are no flipping events, and 350 K, when the flipping events are frequent. As expected, the flipping causes a change of direction in the $(y, z)$ plane. However, the flipping motion is not restricted just to the $(y, z)$ plane but it involves also a component in the $(x, y)$ plane. This is due to the monoclinic distortion of the crystal that leads the hydroxide to rotate in the $(x, y)$ plane.

To distinguish between the different proton and deuteron arrangements, it is convenient to reintroduce the reaction coordinates $\theta_x$ and $\theta_{x+1/2}$ and look at their correlation $\langle\theta_x, \theta_{x+1/2}\rangle$. The joint probability distribution of this couple of variables is shown in fig. 3.11. As anticipated before, positive correlations correspond to FE, while negative to AFE.

At 77 K, due to the rare events of flipping motion of OH and OD groups, the system is frozen in the starting point configuration, the AFE phase, with the classical picture giving much more localized AFE configurations. We know that the T = 0 K potential energy barrier for flipping of a single OH group is quite high (90 meV). At 350 K, where the number of flipping events is sufficiently large, we estimated a free-energy barrier of 18 meV for KOH. Concerning the barrier at 77 K, we predict that it should be placed between the potential energy barrier at T = 0 K (90 meV) and the quantum one at T = 350 K (18 meV) preventing the flipping of the OH and OD groups within our simulation time. At this temperature ($k_BT = 6.6$ meV), the flippings of OH/OD groups are rare events. Furthermore, by chosing a dynamic starting from FE configuration, we have computed the difference of the mean energies of the AFE and FE states, resulting in AFE being more stable of FE by about 5 meV, consistently with the T = 0 K PES, according to which the AFE minimum is more stable than FE minimum by 3 meV. The above assumptions are in accordance with the experiments, that predict the AFE phase at low temperature [113, 114].

At 350 K, the system is driven towards a dynamical disorder characterizing the II phase. The

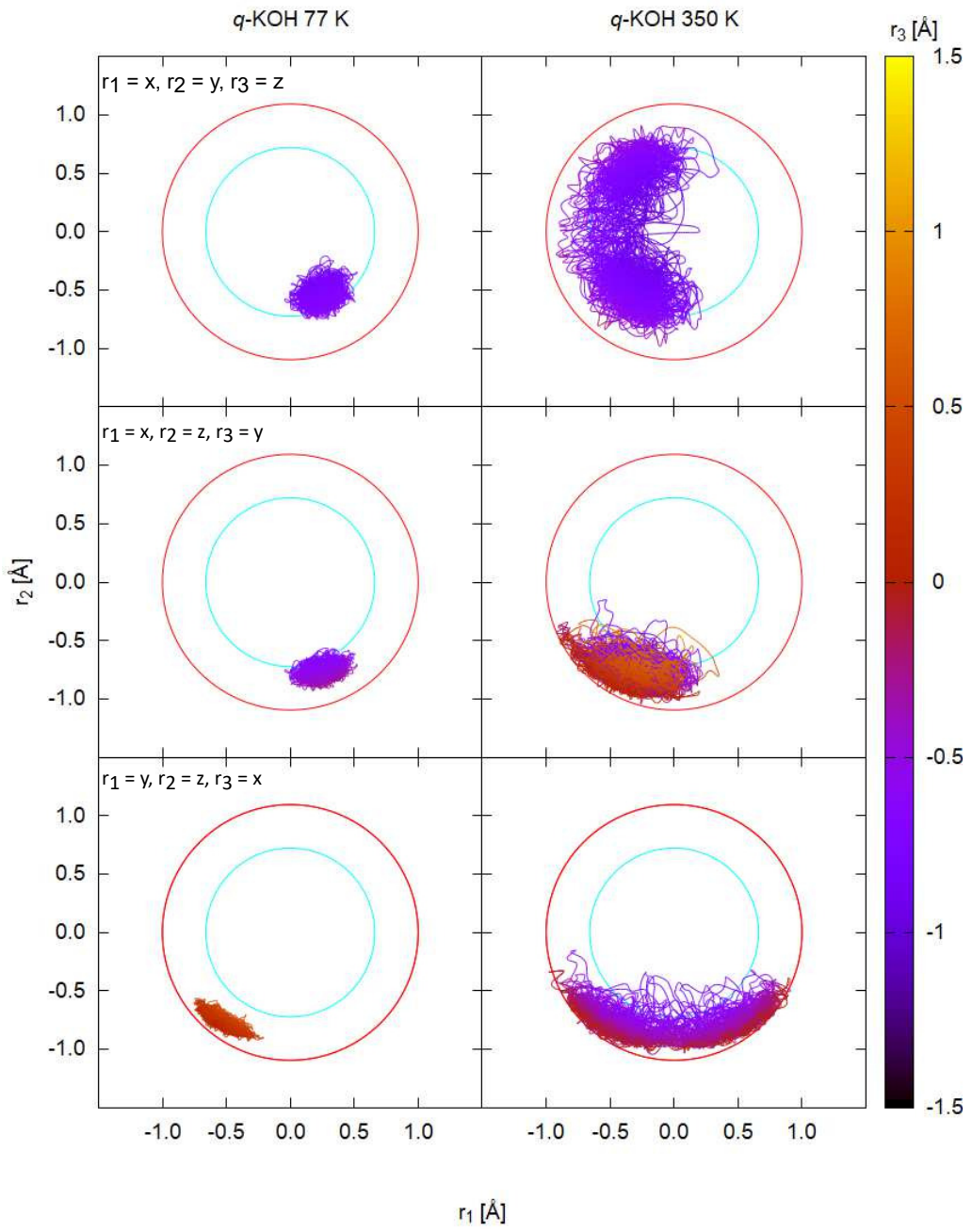



Figure 3.10: The centroid trajectory of an OH group in KOH projected in the $(x, y)$, $(x, z)$ and $(y, z)$ planes from *ab initio* PIMD simulations ($q$-KOH).

*static* PE is never reached, since $\theta$ is non-zero for all the OH/OD groups. Instead, both AFE and FE states are equally populated consistently with a model of half occupancy $4f$ of hydrogen and deuterium atoms [113, 114]. Therefore, phase II can be defined as a dynamical paraelectric phase associated with the disorder of OH and OD groups due to bending motion along the $b$-axis direction, which makes the hydrogen bonds break and form several times within the ps time scale. Thus, the hydrogen bonds play a crucial role in the phase transition.

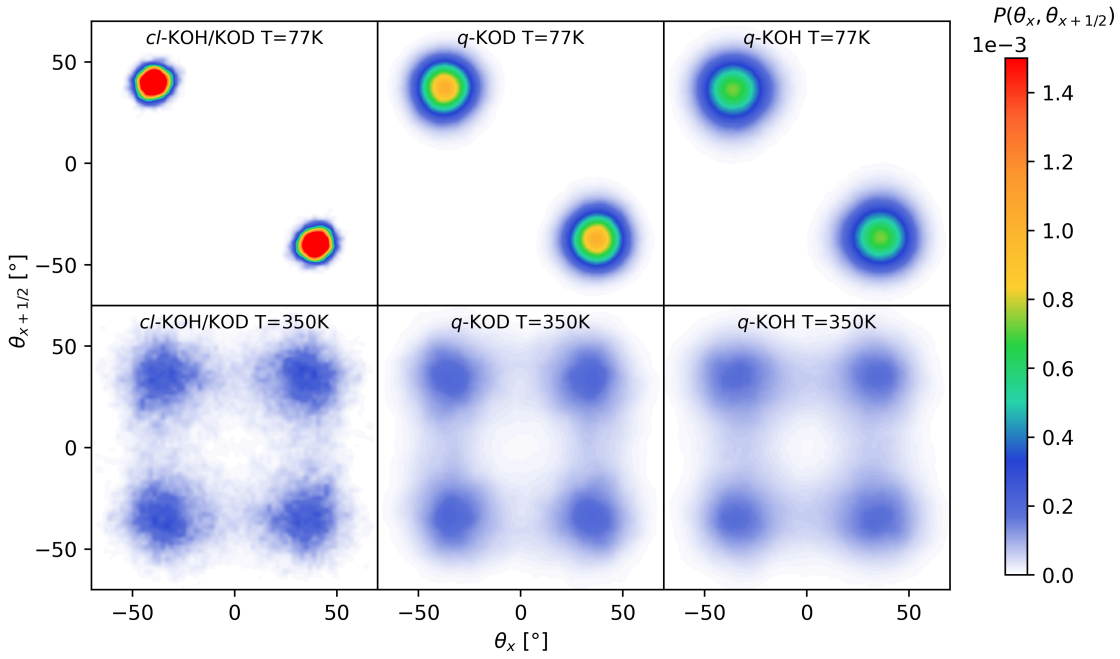



Figure 3.11: Joint probability distribution for $\theta_x$ and $\theta_{x+1/2}$ from *ab initio* classical (*cl*-KOH/D) and PIMD simulations (*q*-KOH and *q*-KOD). The distributions have been symmetrized.

### 3.5.3 Description of the hydrogen bonds

To have more insight on the role of the hydrogen bonds in the phase transition, we calculated the probability distributions of O$\cdots$X length at 77 K and 350 K, which are shown in Figure 3.12. Furthermore, we have computed the joint probability distribution for the hydrogen bond angle and the O-O distance, as shown in Figure 3.13. The major differences between the classical and quantum KOH and KOD are at low temperature. Nuclear quantum effects contribute to make the hydrogen bonds weaker, i.e. longer O$\cdots$X distances in the quantum case with respect to the classical one. Furthermore, the probability distribution of O$\cdots$D at 77 K is shifted to lower values with respect to O$\cdots$H. This geometric H/D isotope effect can be characterized as an inverse Ubbelohde effect [134]. The conventional Ubbelhode effect is the shortening of the hydrogen bond lengths in strong hydrogen-bonded systems. For strong hydrogen bonds, the deuterated system presents longer hydrogen bonds of about $\simeq 0.03$ Å. An example of this is given by the phase transition of ice VII to ice X under pressure [10, 135]. Nuclear quantum effects can act therefore differently depending on the strength of hydrogen bonds due to competing factors [136]. On the one hand, the zero-point motion along the O-H and O-D stretching modes strengthens the hydrogen bonds, on the other the fluctuations normal to the bond – OH and OD bending modes – severely weaken the hydrogen bonds. For strong hydrogen bonds, the stretching motion prevails over the bending one, thus leading to O$\cdots$D lengths longer than O$\cdots$H ones, on the contrary, in weak hydrogen bonds, the quantum fluctuations associated with the bending motion are more important and the O$\cdots$H lenght becomes longer than the O$\cdots$D lenght. Thus, in alkali hydroxides we observe an inverse Ubbelohde effect (O$\cdots$D lenght $<$ O$\cdots$H lenght) [137].

As a consequence, KOH exhibits larger angular quantum fluctuations at 77 K with respect to KOD, as shown in the joint probability distribution of $\angle$O-X$\cdots$O angle and the O-O distance,

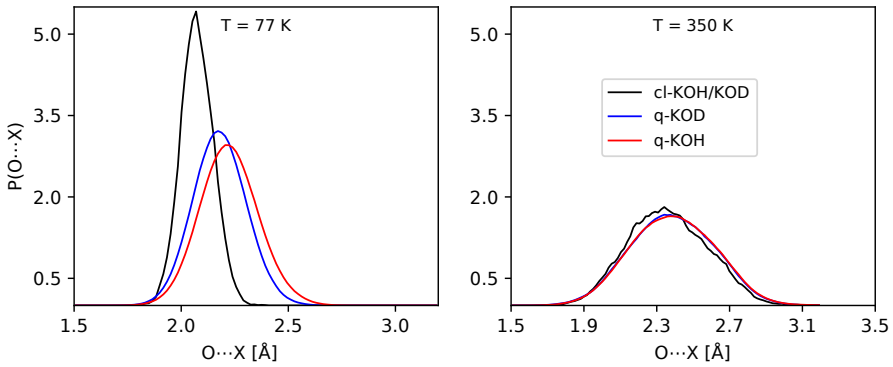



Figure 3.12: Probability distribution for the O· · · X length from *ab initio* classical (*cl*-KOH/D) and PIMD simulations (*q*-KOH and *q*-KOD).

$P(\angle \mathrm{O-X\cdots O}, \mathrm{OO})$ (fig. 3.13). Instead at 350 K, the classical and the quantum KOH and KOD probability distributions $P(\angle \mathrm{O-X\cdots O}, \mathrm{OO})$ look more alike to each other and both show significant contributions in the region $\angle$O-X· · · O $< 140°$, in which the hydrogen bonds are significantly deviating from linearity.

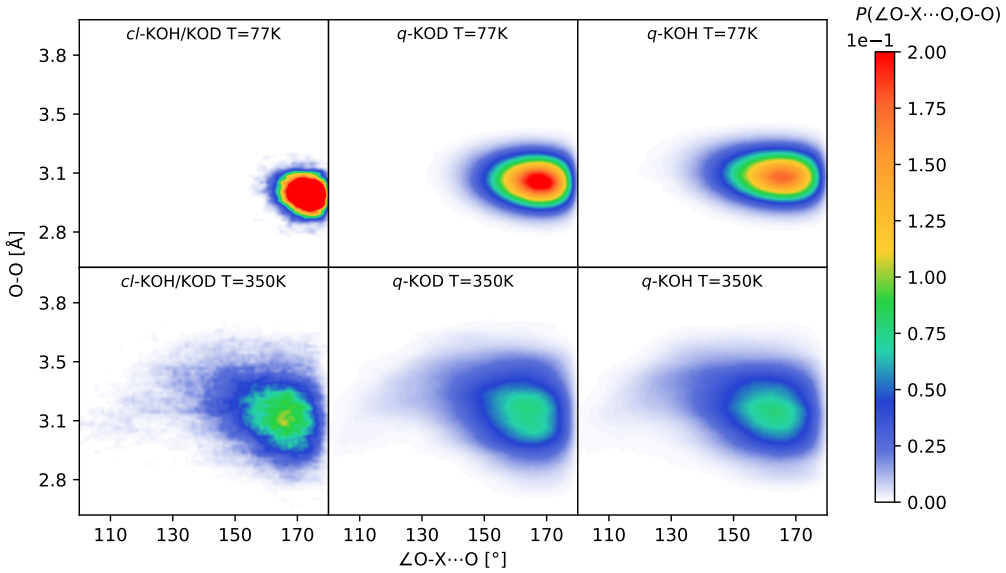



Figure 3.13: Joint probability distribution for the $\angle$O-X· · · O angle and the O-O distance from *ab initio* classical (*cl*-KOH/D) and PIMD simulations (*q*-KOH and *q*-KOD).[4]

## 3.6 Results from the adaptive quantum thermal bath

PIMD simulations are computationally expensive. To obtain, for example, quantities such as the transition temperature, one needs to study a larger system than the one currently investigated to reduce finite size effect. However, this is not affordable in PIMD simulations. To check whether the adQTB method is suitable for this kind of problem, we have compared its results with the PIMD ones for the KOH crystal.

[4] After a revision of our data, we found a discrepancy with respect the O-O distances. The figure was corrected with respect to reference [128].

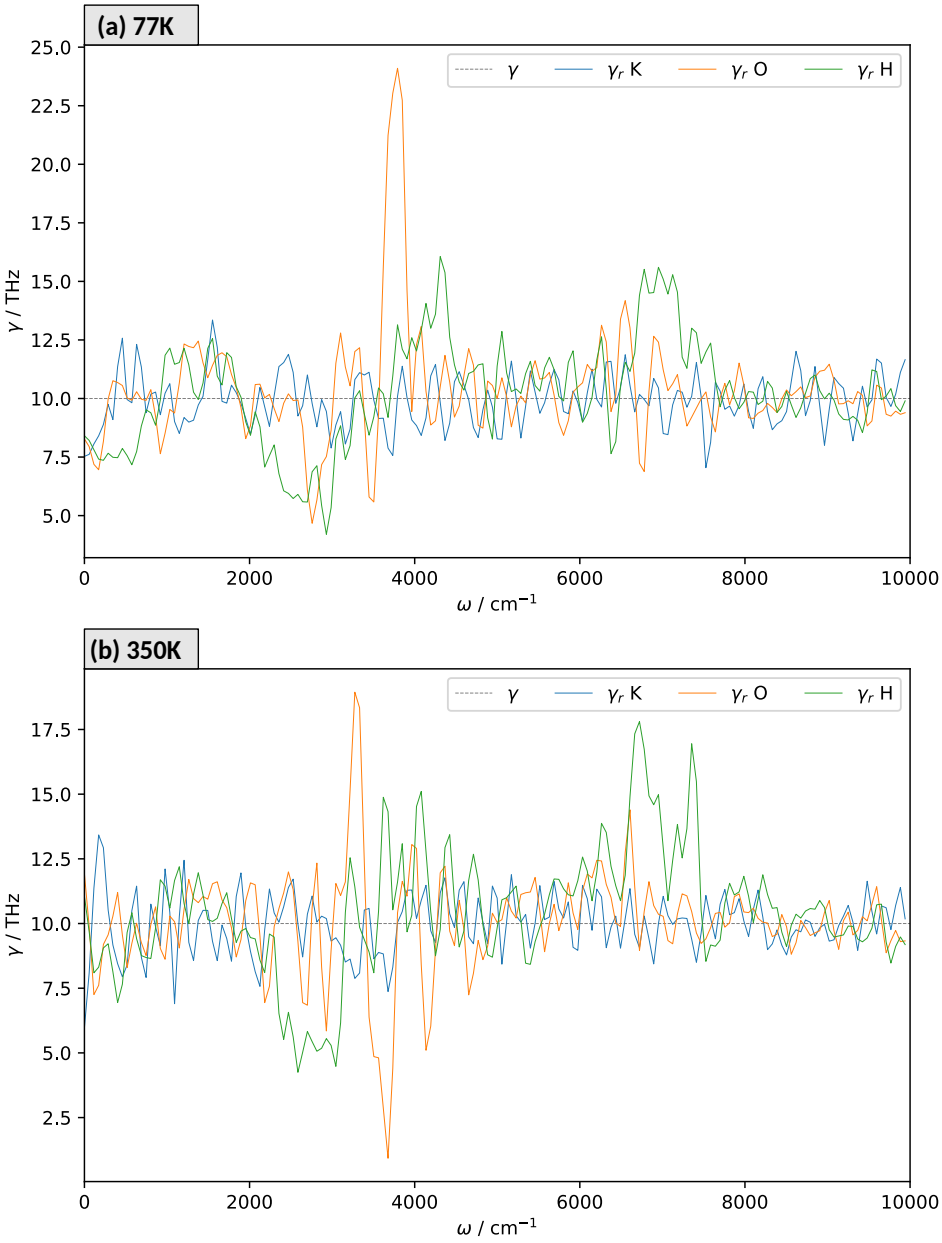



Figure 3.14: Adjusted $\gamma_r$ for K, O and H atoms. Panel (a): low-T. Panel (b): high-T.

The $\gamma_r$ coefficients of the adQTB (see eq. (2.76)) were adapted separately for K, O and H atoms starting from input value of 10 THz. They are shown in fig. 3.14 at 77 K and 350 K. The largest adjustment of $\gamma_r$ are at low temperature for the O atoms at stretching frequency $\sim$ 3700 cm$^{-1}$, while at high temperature the large contribution is also given by H atoms in the region 6500-7500 cm$^{-1}$. This corresponds to an overtone frequency. Concerning the low frequencies, we observe that, at 77 K, the $\gamma_r$ of H atoms is corrected to $\sim$ 7.5 THz, a significant difference compared to the input value, 10 THz. This is a clear sign of the correction of the zero-point energy leakage at low frequency. Without adaptation, the ZPE is leaking from the high-frequency stretching modes towards the low-frequency ones, which would result in an increased effective temperature for the low-frequency lattice modes and could induce structural distortions. At 350 K this effect is less

evident, since the system is found closer to the classical regime and the leakage is less significant.

The adQTB gives successful results regarding the radial distribution function of O and H pairs (fig. 3.15). The adQTB results are very close to the PIMD ones, with a slightly broadening in adQTB simulations and a tiny shift of the OH covalent peak to smaller distance values.

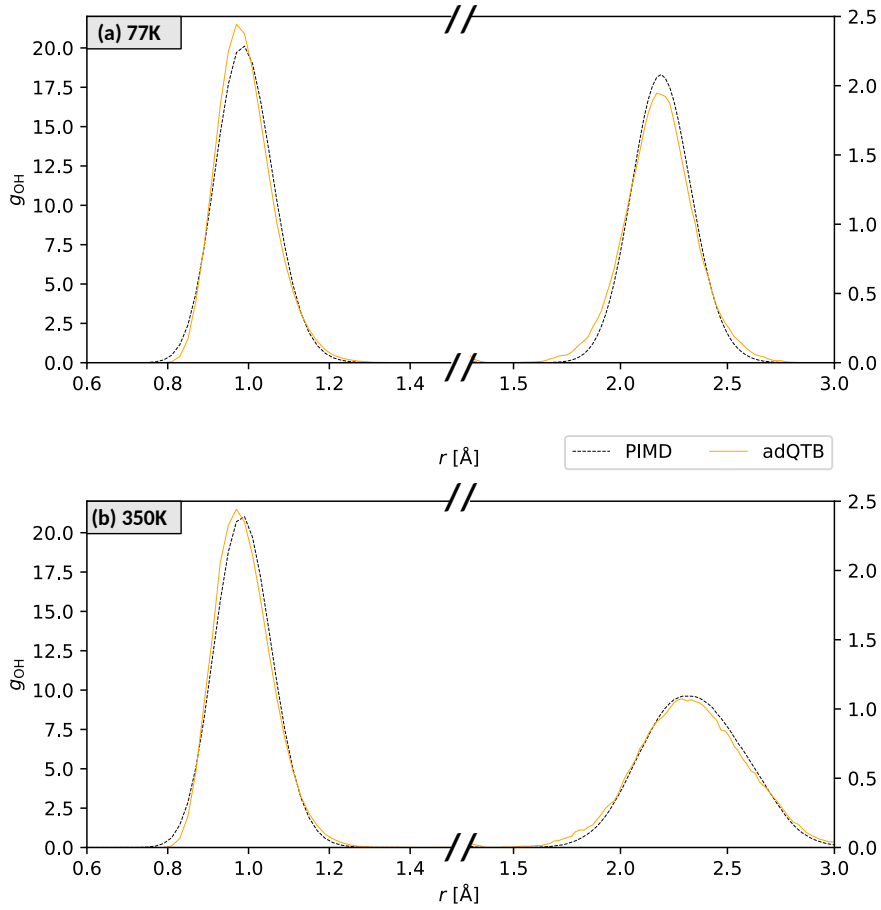



Figure 3.15: Radial distribution function for OH pairs from PIMD and adQTB simulations. Panel (a): low-T. Panel (b): high-T.

Concerning the order parameter of the phase transition $\theta$, we notice a significant difference between the adQTB and PIMD results at 77 K. The probability $P(\theta)$ for $\theta \to 0°$ is larger in the adQTB distribution. It is consistent with other observations that QTB tends to underestimate free energy barriers at low temperature [138]. The reason for this intrinsic feature of QTB is still under investigation, it might be related to an overestimation of quantum energy fluctuations by the method. As a result, the 2D distribution $P(\theta_x, \theta_{x+1/2})$ presents at 77 K both formation of AFE and FE states in the adQTB results, in contrast to what is predicted by PIMD (fig. 3.17).

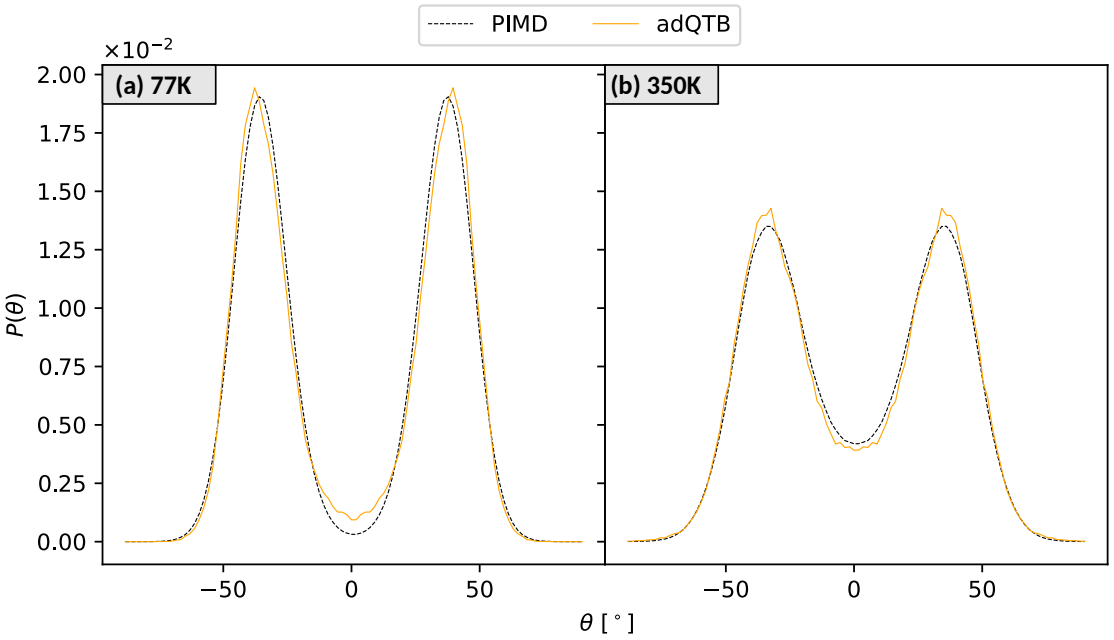



Figure 3.16: Probability distribution for $\theta$ from PIMD and adQTB simulations. Panel (a): low-T. Panel (b): high-T.

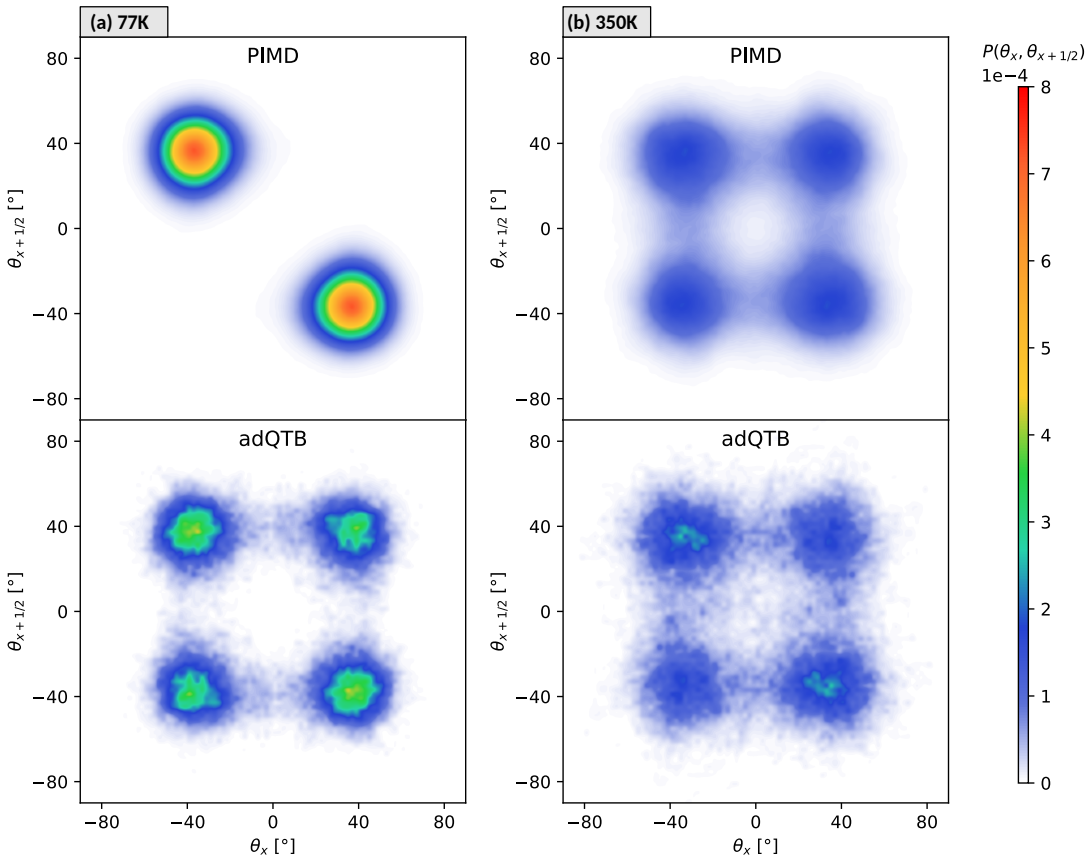



Figure 3.17: Joint probability distribution for $\theta_x$ and $\theta_{x+1/2}$ from PIMD and adQTB simulations. Panel (a): low-T. Panel (b): high-T.

## 3.7 Conclusion

The present study proposes an atomic-scale model for the low-T IVa $\rightarrow$ high-T II phase transition of KOH and KOD crystals and disentangles thermal from nuclear quantum effects, which are included via the path integral formalism. The interplay of (classical) thermal and quantum effects on the nuclei makes the understanding of the phase transition in potassium hydroxides delicate and complex despite the simplicity of its crystal structure. In the low-T IVa phase the positions of the hydrogen atoms are correlated in an AFE arrangement within a monoclinic crystal with space group $P2_1/a$. In contrast, the high-T II phase is characterized by an uncorrelated motion of the hydrogen atoms and, as a consequence, by a dynamical disorder of FE/AFE states, i.e. a dynamical paraelectric phase with half occupancy of H and D sites, which restores a mirror plane (symmetry $P2_1/m$). Therefore, we corroborate the order-disorder phase transition experimentally found by Bastow and coworkers [113], although we cannot discriminate the order of the phase transition from our simulations. The mechanism of the phase transition is the same for both KOH and KOD and it is ruled by the double-well energy profile as a function of the polar angle $\theta$, that is associated to the OH and OD libration modes. The flipping back and forth of the H and D atoms is driven by the thermal and zero-point fluctuations (larger in KOH than in KOD) and generates two possible orientations of the OH and OD dipoles along a zig-zag chain of weak hydrogen bonds parallel to the $b$-axis direction, which hold the KOH stack. Both types of fluctuations contribute to the phase transition, which a purely classical picture cannot capture correctly.

The presence of a network of weak hydrogen bonds is a specific feature that distinguishes KOH and KOD crystals and most alkali hydroxides from other hydrogen-bonded ferroelectrics such as KDP, where the hydrogen bonds are strong. The phase transition is governed by the weak hydrogen bonds. They are highly impacted by quantum fluctuations, as evidenced by the inverse Ubbelohde effect at low temperature: under deuteration the hydrogen bond is shorter and stronger. As a result, the lattice parameters $c$ and $\beta$ for KOH and KOD present a significant difference: in KOD the bilayers are closer, as highlighted by the smaller $c \sin \beta$. The experimental 24 K shift of $T_c$ upon deuteration can be mainly explained in terms of the larger zero-point energy motion of the OH libration mode with respect to OD.

The adQTB results compared to the PIMD give successfully an accurate description for the structural and geometrical properties, while they underestimate the free energy barrier of the proton flipping. Despite this main factor, the adQTB offers a valid model to study this kind of phase transitions. Due to its low computational cost, adQTB is very promising to study larger systems and to reduce the limit size effect, which is essential to study a phase transition in a quantitative way.

# APPENDIX B

## B.1 Computational details

DFT calculations have been conducted using the Quantum Espresso (QE) package [139] working with the PBE exchange-correlation functional [140]. Phonon calculations were performed within the harmonic approximation at T = 0 K via density functional perturbation theory (DFPT) [133]. We have employed ultra-soft pseudopotentials for the oxygen and hydrogen atoms and norm-conserving pseudopotentials for the potassium atom. The plane-wave expansion cutoff energies were 50 Ry for the Kohn-Sham states and 8 times as large for the charge density and the potential, ensuring the total energy convergence (see Figure B.1-(a)). The unit cell contained 2 or 4 molecular units depending on the required symmetry. The Brillouin zone was sampled using a $4 \times 4 \times 3$ ($Z = 2$) and $2 \times 4 \times 3$ ($Z = 4$) Monkhort–Pack k-point grid.

To locally reconstruct the potential energy surface at T = 0 K, we adopted the following procedure. We ran several climbing image-nudged elastic band (CI-NEB) [132] calculations at fixed lattice constants with different initial and final configurations. All the initial and final configurations given in the input were obtained via displacements of the crystal $y$-component of the OH vector, which is related to the coordinate of reaction of the phase transition, by using $y'_{\mathrm{OH}} = y_{\mathrm{OH,eq}} \times \alpha$. The term $y_{\mathrm{OH,eq}}$ is the equilibrium $y$-component of the OH vector and $\alpha$ corresponds to a coefficient in the range [-1.4, 1.4].

Molecular dynamics simulations in the NVT canonical ensemble were carried out within a generalized Langevin equation in a unit cell containing 4 molecular units. Nuclear quantum effects were taken into account using the path integral framework through the i-PI interface [141] combined with QE. The PIMD simulations were performed at the three following temperatures: 77 K, 215 K and 350 K. We have used the PILE-L thermostatting scheme [142] with a centroid friction coefficient of 10 THz. The number of beads in the PIMD simulations was set to 32 and checked to provide the kinetic and potential energy convergence (see Figure B.1-b). Furthermore, we have checked the convergence of the O-X and O$\cdots$X lengths in KOH and KOD at 77 K by looking the results at 32 beads vs 100 beads (see Figure B.1-c). The difference between the KOH and KOD

lengths is larger than the difference between the 32 beads and 100 beads lengths.

The optimized crystal structures have been obtained through systematic volume relaxation by gradually varying the lattice constants $a$, $b$, $c$ and the monoclinic angle $\beta$ until the hydrostatic pressure ($\sigma_{ij} \simeq 0$) was reached, in trajectories of 5 ps each, within an error on the average stress tensor components lower than 2 kbar. Finally, statistical averages were obtained from trajectories of duration time ranging from 20 to 40 ps.

The adQTB simulations were conducted in the NVT canonical ensemble using a local modified version of QE. The total length of the simulation was 60 ps, of which the first 30 ps were discarded as the time for the adaptation of the $\gamma_r$ coefficients for each atom species. The Langevin friction coefficient was set to 10 THz and the adaptation velocity to coefficient $A_\gamma = 10^{-1}$ ps$^{-1}$. The adQTB results were compared with PIMD simulations at 100 beads at 77 K and 32 beads at 350 K.

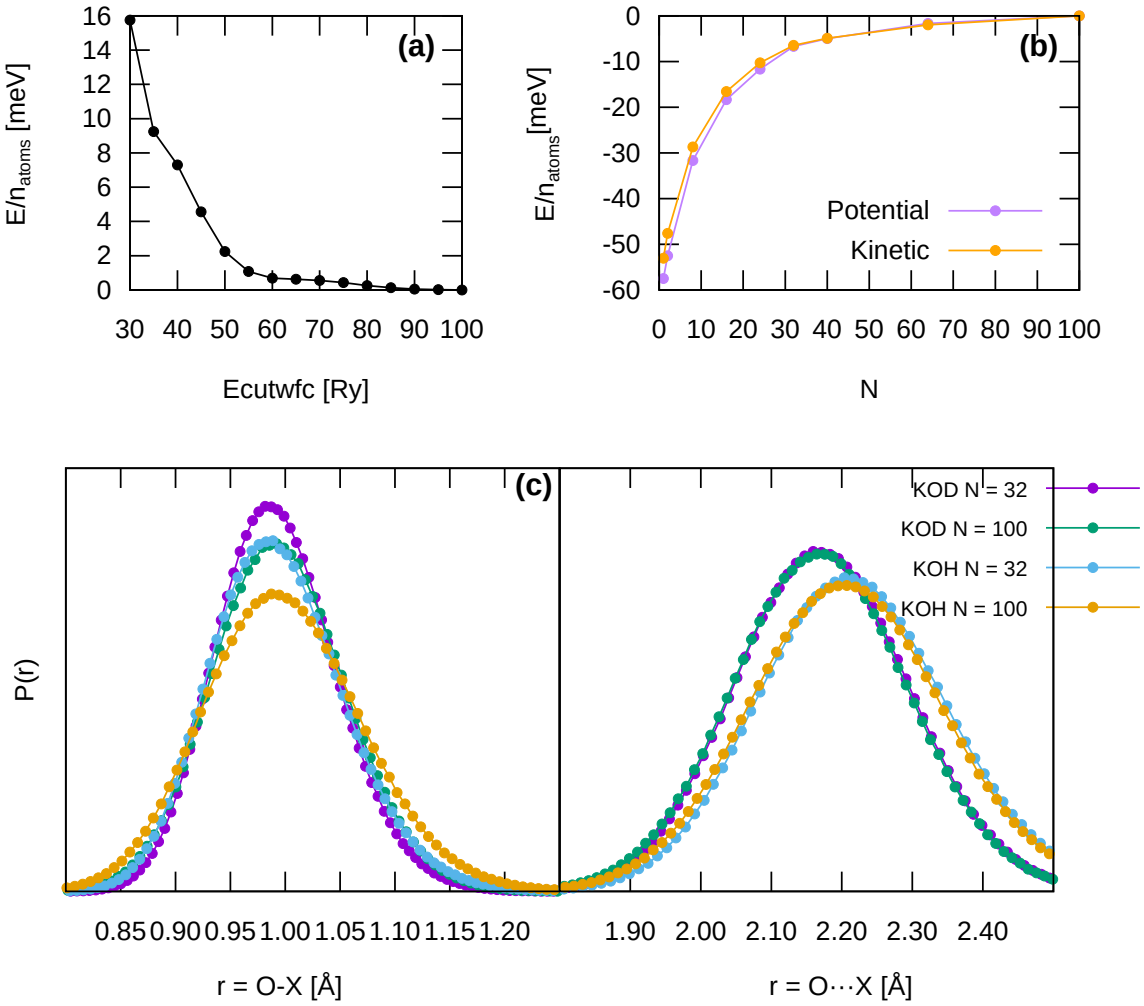



Figure B.1: (a) Convergence of the total energy respect to the energy cutoff for the wavefunctions for KOH. (b) Convergence of the potential and kinetic energies respect the number of beads for KOH at 77 K. (c) Convergence of the O-X and the O$\cdots$X lengths respect the number of beads for KOH and KOD at 77 K.

CHAPTER *4*

# THE PUZZLING ADSORPTION OF FORMIC ACID ON THE TITANIA ANATASE (101) SURFACE

## 4.1 Introduction

Titanium dioxide ($TiO_2$) is nowadays a well-established prototype material due to its optoelectronic and photocatalytic properties tunable for a wide variety of technological applications involving various surfaces of titania [143]. In particular, the interaction with organic molecules can play an important role in catalysis. Carboxylic acids represent a class of compounds of great importance for organic–$TiO_2$ systems. It has been shown that $TiO_2$ has an high affinity for adsorption of atmospheric carboxylic acids, even though these are present at very low concentrations [144]. R-COOH compounds are often used in solar-cell sensing $TiO_2$-based devices to link the dye molecule via the carboxylic group [145–151]. Other applications involve, for example, the catalytic solvent-free amidation of carboxylic groups with amines [152] and the catalytic oligomerization of amino-acid in prebiotic conditions [153–155]. Formic acid (FA) is the simplest carboxylic acid (R=H) and can be used to study the interactions of R-COOH groups and $TiO_2$ by means of both experimental and theoretical techniques. Furthermore FA represents a promising hydrogen carrier material, and a large interest has been devoted to its catalytic decomposition on acid–base pairs at $TiO_2$ surfaces [156–161]. FA can be a source of hydrogen via the dehydrogenation reaction $\mathrm{HCOOH} \rightarrow \mathrm{CO_2} + \mathrm{H_2}$ and eventually can follow the dehydration reaction $\mathrm{HCOOH} \rightarrow \mathrm{CO} + \mathrm{H_2O}$.

Given the wide academic interest in FA-$TiO_2$ systems, a deep understanding at the atomistic level of the interaction between the acid molecule and the $TiO_2$ surface is needed. More generally, the adsorption of FA on oxides surfaces can be described as a molecular (more frequent at low temperature) or dissociative process. In molecular adsorption the intact molecule adsorbs on the surface, while in the dissociation a basic $O^{2-}$ anion abstracts the acid proton of HCOOH forming a $O_{\mathrm{lattice}}H$ species and a formate species $HCOO^-$, which is bound to acidic metal cation sites. The

competition between these two mechanisms depends on the atomic-scale structure of the system.

| | |
|---|---|
| Molecular | no fragmentation |
| Dissociative | $O_{lattice} + HCOOH_{ads} \rightarrow O_{lattice}H + HCOO_{ads}$ |

Among the surfaces of titania, anatase (101), the most thermodynamic stable face of anatase phase, presents quite interesting properties. The $TiO_2$ anatase (101) was observed to be more photocatalytic active than rutile (110) [162]. It is characterized by the presence of coordinative unsaturated ions ($Ti_{5c}$ and $O_{2c}$) and fully coordinated ions ($Ti_{6c}$ and $O_{3c}$), resulting in a peculiar morphology due to the presence of oriented step edges, that strongly influence the reactivity towards adsorbing molecules (fig. 4.1).

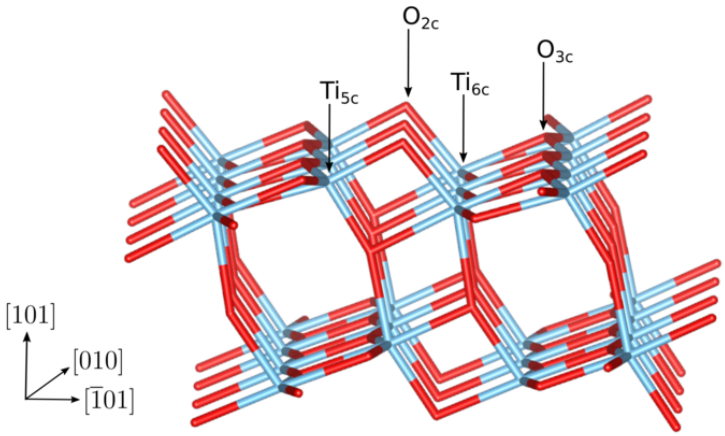



Figure 4.1: Stick representation of clean $TiO_2$ anatase (101) surface. Ti and O atoms are colored turquoise and red respectively.

A general consensus has been reached that FA adsorbs dissociatively on rutile (110) [156, 163–165] forming a bridging bidentate formate and a bridging OH group, while for the adsorption on anatase (101) (FA@A01), previous first-principle calculations [166–174] and experimental studies using FA [158, 173, 175, 176] and acetic acid ($CH_3COOH$) [177] have yielded a complex picture with the presence of competing adsorption configurations (see fig. 4.2). One is a monodentate mode through C=O–$Ti_{5c}$ bond, which can be either molecular (MH) with the acid proton still attached to the molecule, or dissociative (M(H)) with the formation of a monodentate formate. Alternatively FA can adsorb, similarly on what is found for rutile (110), in a bridging bidentate (BB(H)) configuration, where FA coordinates two titanium sites through $Ti_{5c}$–O–C–O–$Ti_{5c}$ bonds, resulting in a OCO plane parallel to the [010] direction.

The molecular representations for MH and BB(H) modes are illustrated in Figure 4.3. In particular, in the case of monodentate, there are two configurations namely MH-intra and MH-inter. They refer to the different orientation of the FA molecule on the surface due the formation of hydrogen bonds with a $O_{2c}$ site of the surface. This is located either on the same chain of the $Ti_{5c}$ adsorption site (MH-intra), either on the neighboring chain (MH-inter). In the precursor DFT work by

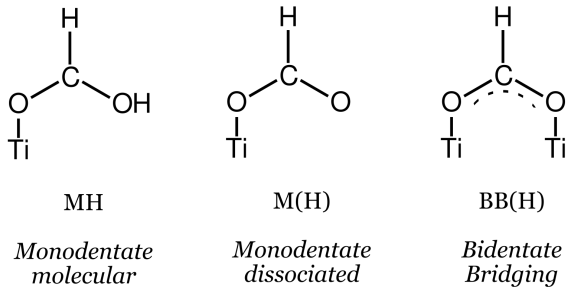



Figure 4.2: Sketch of the most probable adsorption modes for FA@A101. (H) indicates dissociation of FA with the acid proton located on the $TiO_2$ surface.

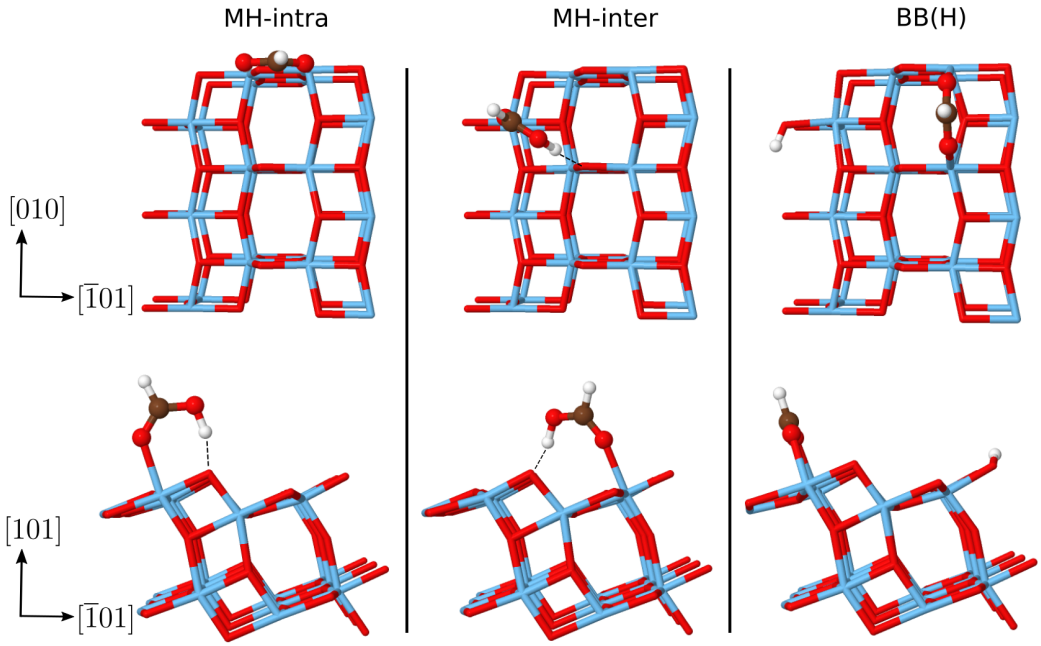



Figure 4.3: FA@A01: stick and ball-stick representation of FA adsorbed on the anatase $TiO_2$ (101) surface in molecular monodentate (intra, inter) and bridging bidentate modes. Ti, O, C and H atoms are colored turquoise, red, brown and white respectively.

Vittadini et al. [166], MH-inter is the most stable configuration. Later on, a study by Wang et al. [173] combined DFT calculations using the strongly constrained and appropriately normed (SCAN) meta-GGA functional [178] with a series of experimental techniques as scanning tunneling microscope (STM) and infrared reflection absorption spectroscopy (IRAS) on single-crystalline anatase (101). The authors proposed for the low temperature adsorption of FA on anatase (101) surface a coexistence, depending on the level of coverage, of both monodentate (preferentially deprotonated) and bridging bidentate configurations. The latter is the most stable on the basis of SCAN DFT calculations and its population decreases at high coverage due to the lower availability of $Ti_{5c}$ pair sites. STM images were not able to distinguish between a molecular or dissociative monodentate, while some features of IRAS spectra were assigned to MH species.

Vibrational spectroscopy represents a fundamental tool for better understanding molecule-surface interactions and, therefore, for assign the species on surfaces. Infrared experiments by themselves

frequently are not sufficient for this attribution and they often require a complementary theoretical study, which can be essential in the identification of the experimental peaks. In the most of the literature, the molecule-surface vibrations are calculated within the harmonic approximation at DFT level of theory, which is for such heterogeneous systems of many DOF the only handy way to provide accurate quantitative information about the PES. However, harmonic approximation is often not adequate and can lead into errors of dozens to hundreds of $cm^{-1}$ [179]. Sometimes these errors can be compensated by the limited accuracy of the exchange-correlation functional employed in the DFT approach, sometimes the harmonic frequencies are simply multiplied by a scale factor to reproduce the experimental data. Therefore, the inclusion of anharmonic effects is essential to properly describe these interactions. Among the different methods to include anharmonic contributions [179], one can rely to "on-the-fly" AIMD, through which it is possible to compute IR spectra through Fourier transform of the dipole moment autocorrelation function. Furthermore, in an anharmonic picture the coupling between different vibrational modes is introduced, which, on the contrary, is completely neglected in the harmonic approximation.

A recent study by Tabacchi et al. on the FA adsorption on the anatase $TiO_2$ (101) surface [172] is, to our knowledge, the only theoretical work, up to now, that takes into account the dynamical effects on the FA adsorption via AIMD and not simply the (meta) stable configurations. In this study, much stress is given to the presence of short and strong hydrogen bonds (SSSH) in the monodentate species, which are considered to be responsible of a strong coupling with the surface. Furthermore, the authors claimed that the proton is rapidly shuttling between the molecule and the surface, and guess that quantum effects at low temperature might be significant, although in their AIMD simulations the nuclei are treated classically.

The presence of H-bonding between the FA molecule and the anatase (101) surface suggests that anharmonic and quantum effects can play a crucial role in the dynamics of such a complex system. However, the inclusion of anharmonic of quantum effects in vibrational spectra of molecule-surface systems is very challenging due to multiple factors [179]. First of all, such systems are limited to electronic structure calculations at DFT level of theory, which can present more accuracy problems with respect to the free molecules calculations, in particularly in the case of stretched and dissociated bonds. Furthermore, despite the existence of different approaches to build accurate PES, DFT calculations cannot provide the estimation of a global potential energy surface function for such systems due to the enormous computational cost of producing *ab initio* datasets and the complexity in the fitting procedure of a quality PES function. Secondly, the degrees of freedom coming from the surface significantly increase the configuration space dimensionality and generate frustrated molecular translations and rotations.

In this chapter, we aim to gain a deeper insight into the FA adsorption on the $TiO_2$ anatase (101) surface through a theoretical investigation including both quantum and anharmonic effects. To describe the vibrational properties beyond the harmonic approximation, we will included the zero-point energy and anharmonic effects through QCT approach and quantum effects such as overtones and combination bands through "on-the-fly" DC-SCIVR, which has been recently applied to study the adsorption of water on the same surface [86]. A tentative assignment of the surface species

will be conducted accompanied by new IR experiments performed by S. Stankic and S. Chenot.[1] If not indicated differently, the experimental results presented in this chapter correspond to their measurements. We will stress the importance of accounting anharmonic and quantum effects into this adsorption phenomenon, in particularly we will focus our attention on the hydrogen-bonded region between the surface and the molecule in monodentate species in combination with PIMD simulations.

The chapter is organized as the following. Firstly, we will present a DFT analysis at T = 0 K of the meta (stable) configurations and provide a local description of the PES in the above-mentioned hydrogen-bonded region. Then, we will illustrate the IR experiments on anatase nanopowders using FA and deuterated FA at room and low temperature and clarify with respect to the literature, what it is still certain and not certain of the assignment. After that, we will present the calculated gas-phase spectra of isolated FA molecule and discuss the accuracy of the chosen DFT approximation. Next, we will proceed with the assignment of the different adsorption configurations by comparing our calculated spectra with the IR experiments and a description of the molecule-surface couplings. Finally, we will discuss the PIMD results for the equilibrium probability distributions of significant geometrical parameters and additionally make a comparison with other well-known systems.

## 4.2 Static adsorption configurations at zero temperature

In this section we report the results from static (T = 0 K) DFT/PBE [26] calculations performed with the Quantum Espresso (QE) package [180]. The calculations were conducted adsorbing a single molecule of FA on a non-defective surface of $TiO_2$ anatase (101) surface, modeled as a periodic slab of 4 $TiO_2$ layers with surface area 10.37 $\times$ 11.37 Å$^2$.

### 4.2.1 Binding energies and geometrical parameters

First of all, we have performed geometry optimizations of molecular MH (intra and inter), dissociated M(H) monodentate and bridging bidentate BB(H). Tables 4.1 and 4.2 report the corresponding binding energies and relevant geometrical parameters for the different adsorption modes. FA can adsorb molecularly on the $TiO_2$ surface in a monodentate mode, either intra or inter, at one $Ti_{5c}$ adsorption sites with very similar Ti-O bond-lenght (2.2 Å). MH-intra and MH-inter are very close in energy with MH-intra configuration being more stable of about 0.01 eV ($E_{\rm tot}^{\rm intra} - E_{\rm tot}^{\rm inter}$), the same trend is observed in the work by Tabacchi et al. [172], where PBE approximation is employed. As introduced previously, these configurations are quite interesting due to the formation of an H-bond between the hydroxyl oxygen of the FA molecule and a $O_{2c}$ site of the surface. The H-bond is relatively short having an inter-oxygen distance length of 2.52 Å and 2.57 Å for MH-intra and MH-inter minima respectively. In the work of Tabacchi et al. [172], the inter-oxygen separation for MH-intra minimum is shorter and corresponds to 2.479 Å. H-bonds are classified strong for inter-oxygen distances in the range 2.4–2.55 Å. MH-intra configuration has the strongest hydrogen bond, consistently with the slightly greater stability. The H-bond angle is close to 180°, when the molecule is rotated on the surface, i.e. the MH-inter configuration, while it deviates from linearity for the MH-intra mode.

[1] Experimental collaborators: **Slavica Stankic and Stéphane Chenot** – INSP–CNRS.

Table 4.1: Binding energy (BE) (eV) for MH, M(H) and BB(H) modes.

| | BE |
|---|---|
| MH-intra | 0.82 |
| MH-inter | 0.81 |
| M(H) | 0.37 |
| BB(H) | 0.67 |

Table 4.2: Relevant distances and angles (Å, °) for gas-phase FA and adsorbed MH, M(H) and BB(H) modes. The $s$ subscript stands for *surface* sites.

| | $Ti_s$-O | $Ti_s$O-C | C-OH | C-H | O-H | H$\cdots$$O_s$ | O$\cdots$$O_s$ | $\angle OHO_s$ |
|---|---|---|---|---|---|---|---|---|
| Gas-phase | | | 1.35 | 1.11 | 0.98 | | | |
| MH-intra | 2.21 | 1.24 | 1.30 | 1.10 | 1.04 | 1.51 | 2.52 | 163 |
| MH-inter | 2.20 | 1.23 | 1.30 | 1.10 | 1.03 | 1.54 | 2.57 | 172 |
| | $Ti_s$-O | $Ti_s$O-C | C-OH | C-H | $O_s$-H | O$\cdots$H | O$\cdots$$O_s$ | $\angle OHO_s$ |
| M(H) | 1.99 | 1.29 | 1.24 | 1.10 | 1.01 | 1.65 | 2.62 | 160 |
| | $Ti_s$-O | $Ti_s$-O | $Ti_s$O-C | $Ti_s$O-C | C-H | $O_s$H | | |
| BB(H) | 2.09 | 2.12 | 1.26 | 1.27 | 1.11 | 0.97 | | |

The dissociation of the FA can lead to a monodentate M(H) or bidentate BB(H) configuration. The BB(H) mode is characterized by shorter Ti–O bonds with respect to both molecular monodentate MH-intra and MH-inter, and the PBE approximation predicts this configuration to be less stable than MH-intra by about 0.15 eV. Other studies have predicted MH to be more stable than BB(H): 0.92 *vs* 0.68 eV using GGA[2] [166], 0.971 *vs* 0.929 eV using PBE [172], while another study using meta-GGA SCAN method predicts BB(H) binding energy at 1.41 eV and MH at 1.21 eV [173]. The geometry optimization of a M(H) formate, where the acid proton is located on the surface and not forming any hydrogen bonds with the molecule, evolves in the optimized BB(H) local minimum. The geometry optimization of M(H) mode without constraints is only possible for a "dissociated inter" configuration, which we will refer simply to as M(H). The monodentate dissociated is highly unfavoured and presents the lowest binding energy at 0.37 eV, which is in good accordance with the binding energy predicted by Tabacchi and coworkers [172].

### 4.2.2 Potential energy surface of monodentate species

The minimum energy path of proton transfer between the FA molecule and the surface was modeled via CI-NEB method [132] (see fig. 4.4). The parameter $\delta$, defined as the difference between $O_s$H and OH distances, is chosen as the coordinate of reaction. Dissociation corresponds to $\delta < 0$, while for $\delta > 0$ the acid proton is attached to the FA molecule. At $\delta = 0$ the hydrogen bond is symmetric. From the left panel in fig. 4.4, the potential is highly anharmonic and corresponds to a skewed single-well potential rather than an asymmetric double well. Therefore, the dissociated configuration is not even a metastable minimum according to DFT/PBE approximation.

[2]The subtype of functional was not specified in the work.

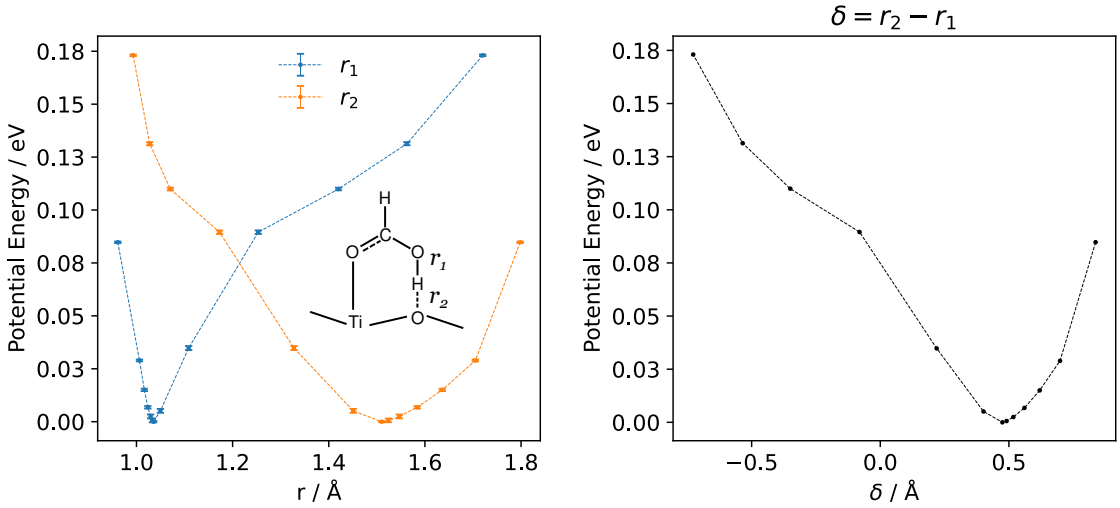



Figure 4.4: Potential energy respect $O_sH$ and OH distances and $\delta$ proton transfer coordinate.

**Electron charge density in OH$\cdots$O$_s$ region** At the equilibrium position of the MH-intra configuration, the maximum of the electron-density located on the oxygen and the O–H group (fig. 4.5) can be considered as a single identity. By fixing the hydrogen atom midway between the oxygen of the FA and the oxygen of the surface ($O_fH = O_sH$) a proton shared configuration is obtained. The O–H cannot be considered a single entity as in the equilibrium position, and two bond critical points in the Bader's terminology [181] occur: one between H and $O_s$, the other between H and $O_f$. This is analogous to the symmetrization of hydrogen bonds in ice-X [182]. However, at variance with the proton-symmetric phase X of ice under pressure, the proton at mid-distance between $O_s$ and $O_f$ is not a stable position. This result is also consistent with the lack of a mirror symmetry through the $O_s$-$O_f$ mid-point.

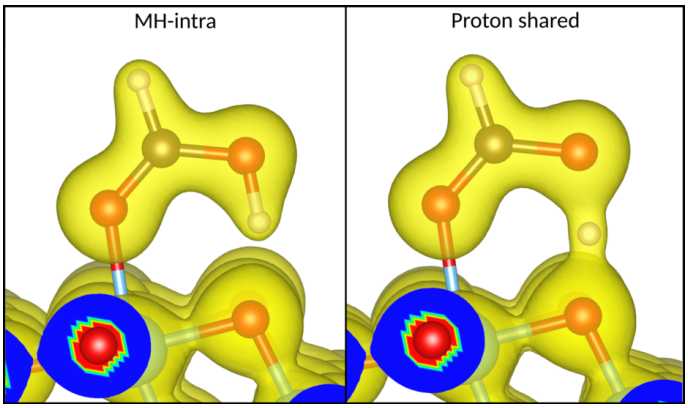



Figure 4.5: Charge electron density associated to the MH-intra mode (left) and the proton shared configuration (right). The isodensity used to generate this figure is equal to 0.1 e/Å$^3$.

## 4.3 IR experiments on anatase nanopowders

In this section we detail the IR experimental results. Fourier transform infrared spectroscopy (FTIR) experiments were performed on $TiO_2$ anatase nanopowders samples exposed to formic acid. As nanopowders have a very high surface-to-volume ratio, a large quantity of molecules can

be adsorbed and a more intense signal than single crystals can be obtained, at the expense of an increased level of complexity, as nanopowders could show various surface defects [183, 184]. The spectra were recorded as a function of the formic acid partial pressure, $P_{FA}$, at room temperature and at low temperature (13 K). To have a better understanding of the acid proton location, the same experiments were repeated using FA deuterated in the OH acid position, e.g. HCOOD. A tentative assignment of the IR peaks will be given on the basis of infrared data from single crystal [173, 175] and nanopowders [158, 185–187] measurements. Some new additional features with respect to the literature will be highlighted.

### 4.3.1 Room temperature

Figure 4.6 reports the FTIR spectra recorded for HCOOH at room temperature as a function of the partial pressure of FA.

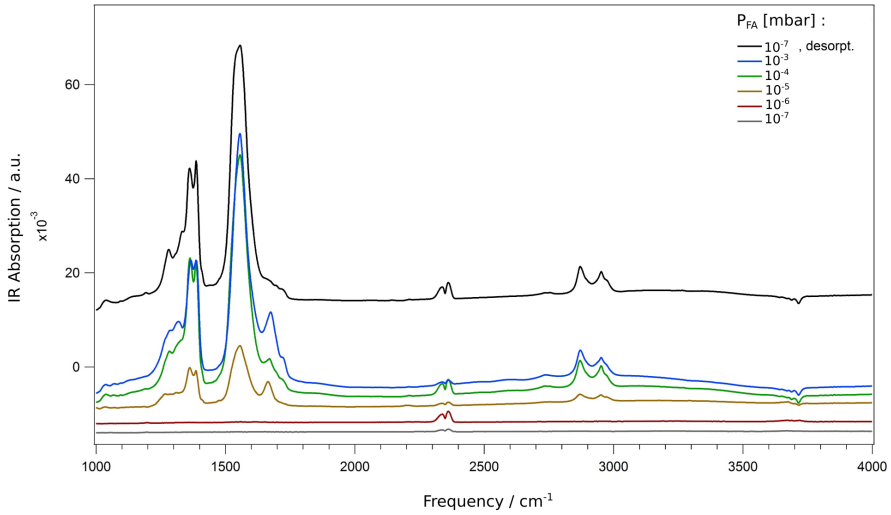



Figure 4.6: FTIR spectra at room temperature of anatase $TiO_2$ nanopowders exposed to $P_{FA}$ = $10^{-7}$–$10^{-3}$ mbar, compared to the spectra recorded after FA desorption at $P_{FA} = 10^{-7}$ mbar (black curve).

**High-frequency range** Around $\sim$ 3700 cm$^{-1}$ a negative pronounced signal appears, together with other secondary negative peaks of lower intensity. Similar signals were found in the work of Petrik et al. [158] and attributed to surface OH hydroxyl groups. The CH stretching region presents two main peaks at $\sim$ 2870 and 2955 cm$^{-1}$. Petrik and coworkers [158] assigned the first peak to the CH stretch fundamental, $\nu$(CH), and the second peak to a combination band. It is still need to be clarified the origin of the CH stretch splitting. The doublet at $\sim$ 2350 cm$^{-1}$ is the signal coming from the $CO_2$ of the chamber.

**Intermediate-frequency range** The OCO region around 1500-1750 cm$^{-1}$ is crucial for the identification of monodentate and bidentate species. The peak at 1725 cm$^{-1}$ and the one at 1675 cm$^{-1}$ were previously assigned to the C=O stretching, $\nu$(C=O), of molecular and dissociated monodentate mode, respectively [158]. Similar bands were found in the work by Nanayakkara and coworkers [187] and assigned to $\nu$(C=O) of FA chemisorbed and physisorbed, respectively. The peak at 1555

cm$^{-1}$ can be attributed, within a general consensus of the previous works, to the O–C–O asymmetric stretch, $\nu_a$(OCO), of a BB formate, and it is the main peak present after desorption, meaning that it is associated with the most stable configuration. The peak at 1385 cm$^{-1}$ was assigned in all the cited references to a CH bending mode, which might be compatible with both monodentate and bidentate configuration. The peak at 1360 cm$^{-1}$ was assigned to symmetric stretching of OCO, $\nu_s$(OCO), of BB(H) configuration. The peak at 1320 cm$^{-1}$ might be attributed to the C-O stretch, $\nu$(C-O), of a monodentate formate as suggested by Xu and coworkers [175]. Finally, we notice the presence of new additional peaks at 1280 and 1035 cm$^{-1}$.

### 4.3.2 Low temperature

Figure 4.7 shows the FTIR spectra recorded for HCOOH at low temperature as a function of the partial pressure of FA.

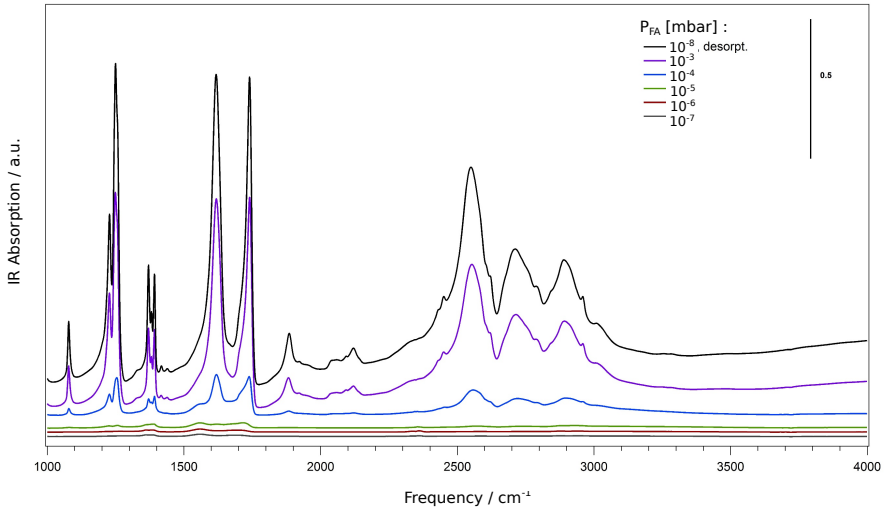



Figure 4.7: FTIR spectra at low temperature (13 K) of anatase $TiO_2$ nanopowders exposed to $P_{FA}$ = $10^{-7}$–$10^{-3}$ mbar, compared to the spectra recorded after FA desorption at $P_{FA} = 10^{-8}$ mbar (black curve).

With respect to the room temperature results, most of the IR bands are found at similar frequency but with a variation in the peak ratio, which stems from a different of level of population. Several new spectral features are also present, which in some cases superimpose some of the peaks present at room temperature. The combination of these factors leads to think that multiple adsorption configurations are present, some of which are more stable at low temperature rather than at room temperature. We point also out that the low temperature spectra could show also the adsorption of other impurities present in the chamber. Therefore, we limit our discussion to some of the features present in the spectrum, which is very complex. Notably, in the CH stretching region there are three intense large bands in the region 2500-3000 cm$^{-1}$, revealing a different behavior compared to the room temperature results. In the OCO range the room temperature peak around 1665-1675 cm$^{-1}$ is present at $\sim$ 1680 cm$^{-1}$ for FA partial pressure lower than $10^{-4}$ mbar, for higher pressure it is superimposed by the band at $\sim$ 1720-1740 cm$^{-1}$. A new high intensity band is present at 1620 cm$^{-1}$. A similar band has been attributed in literature to the OH bending of coadsorbed water

[186, 187], which could be formed, for example, by the reaction of adsorbed formate and surface hydroxyl groups. Finally, new peaks appear at $\sim$ 1250, 1230 and 1080 cm$^{-1}$.

### 4.3.3 The effect of deuteration

Figure C.3 (see the appendix) and fig. 4.8 show a comparison between protonated and deuterated FA adsorbed at room and low temperature, respectively. Firstly, we notice that the OH stretching negative signal around 3700 cm$^{-1}$ is also present in the room temperature spectrum of deuterated FA (Figure C.3), confirming that is not associated with formic acid molecule because, if it was coming from the acid proton, it would have shifted to lower frequency. At room temperature the two spectra are very similar, probably meaning that at high temperature the most stable species is deprotonated and, in particular, it corresponds to the bridging bidentate configuration. Instead, the low temperature spectra for the adsorption of HCOOH and HCOOD significantly differ, particularly in the region 1900-3000 cm$^{-1}$. As introduced before, some broad bands in the range 2500-3000 cm$^{-1}$ are present in the HCOOH spectrum, while the HCOOD spectrum is characterized by two main high intensity bands in the range 1900-2300 cm$^{-1}$. Therefore these bands are in some way involved with the OH and OD groups. Finally, it is clear that passing from room to low temperature for both HCOOH and HCOOD the characteristic bands for bridging bidentate adsorption $\nu_a$(OCO) and $\nu_s$(OCO) decreased significantly in intensity ratio, meaning that the bridging bidentate mode is less populated and leaves space for the adsorption of other species, most likely molecular monodentate FA.

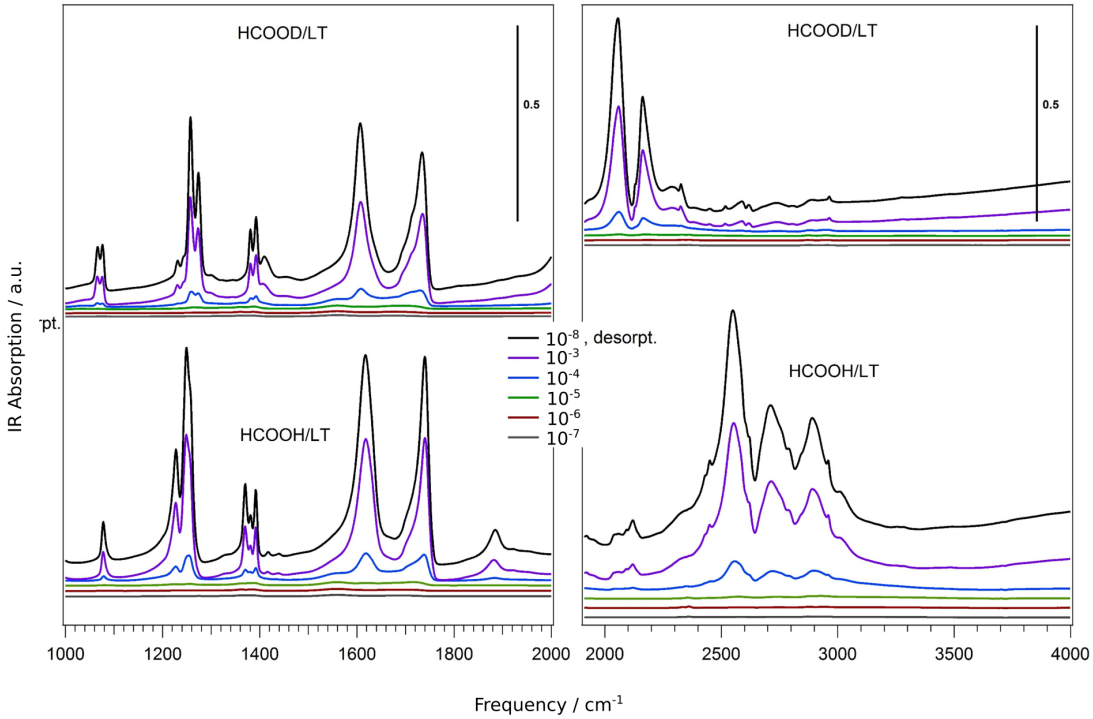



Figure 4.8: FTIR spectra at low temperature (13 K) of protonated FA *vs* deuterated FA adsorbed on anatase $TiO_2$ nanopowders exposed to $P_{FA}$ = $10^{-7}$–$10^{-3}$ mbar, compared to the spectra recorded after FA desorption at $P_{FA} = 10^{-8}$ mbar (black curve).

### 4.3.4 Summary

The spectroscopic signals in the OCO region are the key to distinguish bidentate and monodentate configurations. At room temperature the most likely configuration for both protonated and deuterated FA is bridging bidentate mode with $\nu_a$(OCO) and $\nu_s$(OCO) at 1550 and 1360 cm$^{-1}$, though some low intensity peaks characteristic for $\nu$(C=O) at $\sim$ 1675 and 1725 cm$^{-1}$ are clearly visible in the spectra. At very low temperature, four peaks are present in the OCO region for FA and deuterated FA for pressures below $10^{-4}$ mbar. One is the $\nu_a$(OCO) signal, whose intensity decreases significantly with the increase in FA concentration, i.e. BB(H) is less stable at high level of coverage. This might due to the fact that bridging bidentate mode actually needs two titanium sites to adsorb, which are less available at high coverage, thus the monodentate becomes thermodynamically more likely [173]. The other three instead are at $\sim$ 1600-1620, $\sim$ 1680 and $\sim$ 1720-1740 cm$^{-1}$. The first one was, as already mentioned, attributed to water adsorption in previous works, while the other two probably correspond to the $\nu$(C=O) peaks present at room temperature, but it is still to clarify which one (or both) is associated to molecular adsorption. Finally, we stress the importance of the H/D isotopic shift at low temperature of the bands in the region 2500-3000 cm$^{-1}$. This could be due to the OH stretch of a very strong hydrogen-bonded species that, under deuteration, shift to lower frequencies. To summarize the most relevant characteristics of IR experimental spectra at room temperature *vs* low temperature, we report in the following a short scheme.

| | |
|---|---|
| Room temperature | bridging bidentate mode most stable<br>low population of monodentate species |
| Low temperature (13 K) | at high coverage bridging bidentate (monodentate) is less (more) stable<br>new spectroscopic features<br>H/D substitution effect |

An exhaustive assignment of the different vibrational modes will be conducted in section 4.5, where we will compare these results to the theoretical calculated spectra.

## 4.4 Gas-phase formic acid and influence of the exchange-correlation approximation

Before, discussing the calculated spectra of FA adsorbed on the anatase (101) surface, we present here the spectra of isolated formic acid in the gas phase together with an analysis of the accuracy of the DFT computational set-up, that we employed in this study, DFT/PBE level of theory using plane waves basis set as implemented in the QE code. Firstly, we calculated the harmonic frequencies of formic acid molecule, then we employed QCT (eq. (2.70)) and "on-the-fly" DC-SCIVR[3] (eq. (2.58)) approaches. To estimate the accuracy of the employed DFT set-up, we have compared these results to the same calculations using DFT/PBE, DFT/PBE0 and DFT/B3LYP calculations on isolated formic acid molecule via a Gaussian basis set. The latter calculations were performed with the NWChem suite code [188] by M. Cazzaniga.[4] The B3LYP functional has been employed

[3] From now on, we will refer to DC-SCIVR simply as semiclassical (SC) method.

[4] Theoretical collaborator: **Marco Cazzaniga** – Dipartimento di Chimica, Università degli Studi di Milano

successfully in different semiclassical spectroscopy studies for gas-phase molecules and it serves here as a robust reference. The results are reported in fig. 4.9 and table 4.3. The following notations is used: $\nu$ (stretching), $\gamma$ (rocking), $\delta$ (bending/deformation), $\omega$ (wagging) and $\tau$ (torsion). A visualization of normal modes displacements is reported in appendix, fig. C.4.

**Harmonic vs QCT and SC** In order to appreciate the degree of anharmonicity and quantum effects, harmonic, QCT and SC spectra are shown in the same graphs. For simplicity, we limit to discuss the PBE spectra using QE (panel a of fig. 4.9), but similar conclusions can be extended to the other graphs. The QCT and SC spectrum for each normal mode are red-shifted respect the harmonic estimate. To identify the fundamental of SC calculations we compare SC spectra to the classical reference (QCT). We generally take as fundamental frequency the maximum of the QCT peak and the corresponding signal in the SC calculation. By comparing QCT and SC spectra, we see the increase in the level of complexity. On the one hand, QCT just includes the anharmonicity. The SC spectra instead present additional features as both even and odd overtones and combination bands. The black spectrum include the even overtones and the ZPE peak, which we have placed at 0 $cm^{-1}$. Instead the spectra for each vibrational mode include the corresponding fundamental and the odd overtones. For example at $\sim$ 2180 $cm^{-1}$ there is the overtone of $\omega$(CH), while at 2220 $cm^{-1}$ the one of $\nu$(C-O), which are completely neglected in the QCT spectrum.

**Periodic conditions vs isolated system** We firstly point out that the PBE calculations using plane waves basis set (QE) are consistent with the PBE calculations using a Gaussian basis set on a isolated system (NWChem) at the three levels of approximation (harmonic, quasi-classical and semiclassical). The largest difference between the two methods is over $\nu$(C=O) and $\tau$(COH). We remind that also other factors might be the source of these differences such as the cutoff for the energy wavefunctions and density in the QE calculations and the basis set employed for the NWChem calculations.

**Exchange-correlation functional** By using different exchange-correlation functionals, we found that QCT and SC frequencies are more accurate using the hybrid functional PBE0 and reach the best agreement using B3LYP functional, which presents a mean absolute error (MAE) of $\sim$ 30 $cm^{-1}$. However, the semiclassical PBE spectrum using QE reproduces quite well all the features of B3LYP semiclassical spectra, overtones and combination bands. Thus, we consider our computational set-up reliable to study the adsorption of HCOOH on the $TiO_2$ anatase (101) surface, keeping in mind that some of the QCT and SC frequencies will be largely shifted compared to the experiment.

## 4.5 Assignment of monodentate and bridging bidentate vibrational features

As previously discussed in section 4.3, there is evidence that both mono- and bidentate configurations contribute to the spectra. However, the monodentate/bidentate relative population is temperature-dependent. To disentangle purely anharmonic and temperature-dependent effects from those coming from the sum of signals from mono- and bidentate configurations, we compare

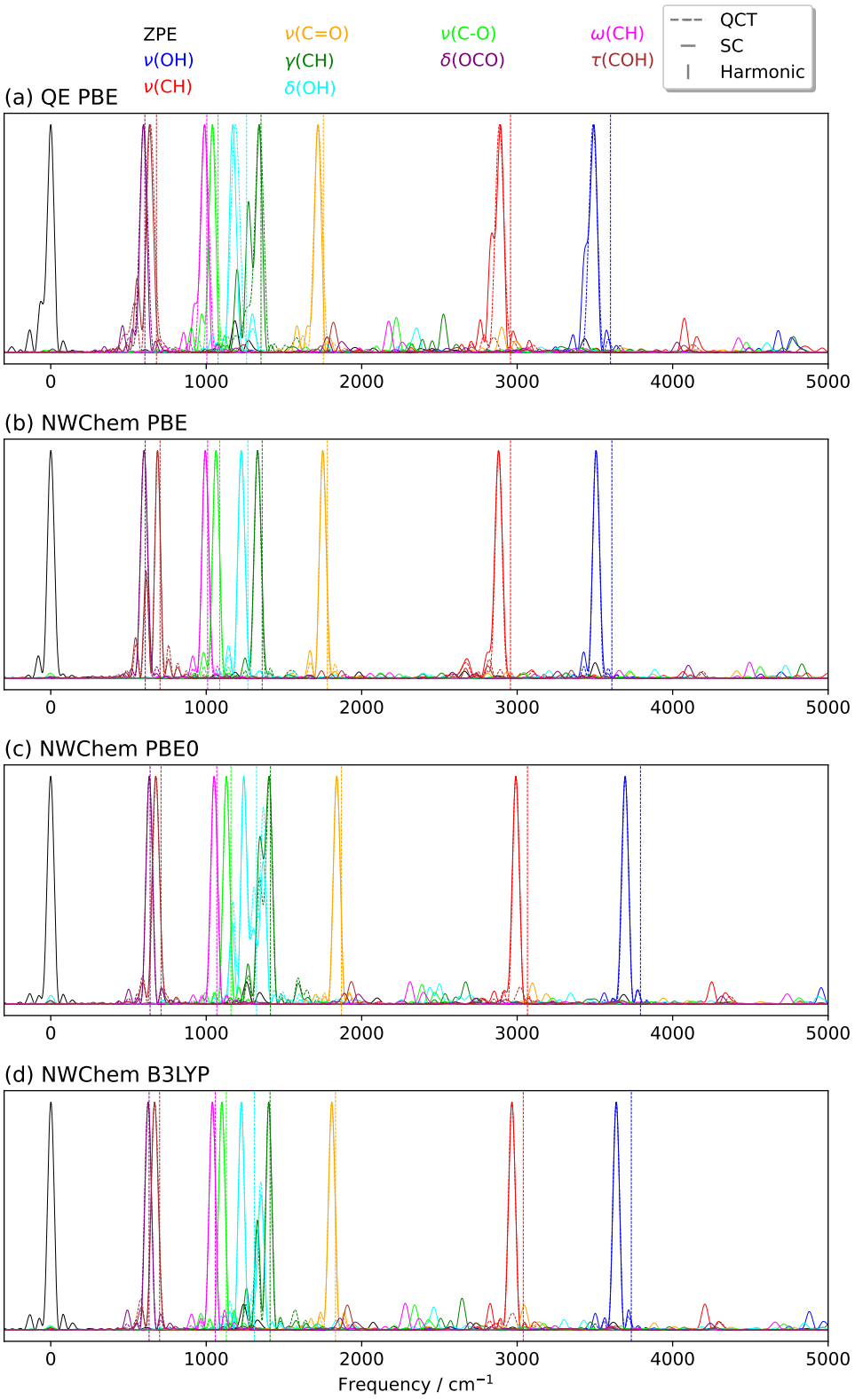



Figure 4.9: QCT and SC spectra for the gas-phase FA calculated at different theoretical approximations. The QCT and SC are reported in dashed and continuous lines respectively, harmonic values are represented with dashed vertical lines. The intensity of the each peak has been normalized to one and SC spectra have been shifted with ZPE peak to 0 energy value.

Table 4.3: Harmonic, QCT and SC frequencies for the gas-phase FA calculated at different theoretical approximations. All values are given in $cm^{-1}$.

| | Exp.[a] | Harmonic | QCT | SC | Harmonic | QCT | SC |
|---|---|---|---|---|---|---|---|
| | | QE PBE | | | NWChem PBE | | |
| $\nu$(OH) | 3570 | 3600 | 3489 | 3493 | 3609 | 3507 | 3508 |
| $\nu$(CH) | 2942 | 2957 | 2886 | 2893 | 2957 | 2881 | 2880 |
| $\nu$(C=O) | 1776 | 1753 | 1720 | 1718 | 1779 | 1749 | 1750 |
| $\gamma$(CH) | 1379 | 1352 | 1342 | 1340 | 1360 | 1330 | 1330 |
| $\delta$(OH) | 1306 | 1260 | 1187 | 1171 | 1268 | 1226 | 1226 |
| $\nu$(C-O) | 1104 | 1076 | 1042 | 1039 | 1087 | 1063 | 1062 |
| $\omega$(CH) | 1033 | 1004 | 990 | 989 | 1009 | 994 | 994 |
| $\tau$(COH) | 641 | 680 | 641 | 635 | 703 | 687 | 686 |
| $\delta$(OCO) | 636 | 605 | 597 | 596 | 608 | 600 | 600 |
| MAE | | 30 | 55 | 57 | 27 | 49 | 49 |
| | | NWChem PBE0 | | | NWChem B3LYP | | |
| $\nu$(OH) | 3570 | 3793 | 3695 | 3694 | 3734 | 3637 | 3638 |
| $\nu$(CH) | 2942 | 3066 | 2992 | 2992 | 3039 | 2966 | 2966 |
| $\nu$(C=O) | 1776 | 1869 | 1840 | 1840 | 1832 | 1807 | 1808 |
| $\gamma$(CH) | 1379 | 1413 | 1405 | 1404 | 1411 | 1402 | 1402 |
| $\delta$(OH) | 1306 | 1324 | 1242 | 1242 | 1310 | 1225 | 1226 |
| $\nu$(C-O) | 1104 | 1159 | 1130 | 1130 | 1129 | 1100 | 1100 |
| $\omega$(CH) | 1033 | 1068 | 1052 | 1052 | 1058 | 1039 | 1040 |
| $\tau$(COH) | 641 | 709 | 677 | 676 | 700 | 669 | 668 |
| $\delta$(OCO) | 636 | 639 | 634 | 634 | 633 | 626 | 626 |
| MAE | | 73 | 45 | 45 | 52 | 30 | 31 |

[a] From reference [189].

the frequencies of mono- and bidentate modes as obtained by means of *ab initio* QCT and SC calculations with the location of the measured peaks. To understand better the role of OH and OD groups, we have also calculated the spectra for deuterated HCOOD. Relying on the extensive tests on the gas-phase FA (see section 4.4), the assignment of the experimental spectra will be conducted by comparing the calculated fundamentals with the measured IR peaks, and by taking into account the intrinsic error due to the DFT/PBE approximation. We have divided the spectra into multiple regions: the fingerprint (900-1500 $cm^{-1}$), the carbonyl and the C-H stretching regions, whose spectroscopic features are essential to distinguish between the mononodentate and briding bidentate modes. Then, we will focus on the OH and OD stretching signals. Since the deprotonated monodentate is highly unfavoured according to our calculations, we will not include it in this analysis. We remind that the intensities of the computed QCT and SC power spectra cannot be compared to experimental IR absorption intensities. Therefore we will limit our study to the peak positions and shape of the signals.

### 4.5.1 Fingerprint region

Figure 4.10 presents the experimental and calculated spectra in the fingerprint region (900-1500 $cm^{-1}$).

**COH torsion** Starting from the lower extreme of fig. 4.10, we notice the presence of a peak at $\sim 970$ $cm^{-1}$ in the low-temperature spectrum of HCOOH (panel (a2)), which is not present in the spectrum of HCOOD (panel (b2)). A tentative assignment of this peak is the vibrational mode $\tau$(COH) of an adsorbed molecular monodentate geometry, compatible with both MH-intra (panel (a3)) and MH-inter ((panel (a4)). We have calculated the corresponding torsion for the deuterated system, $\tau$(COD), which for both harmonic, QCT and SC methods is below 800 $cm^{-1}$ and, therefore, it does not belong to the investigated frequency-region (900-1500 $cm^{-1}$).

Table 4.4: Experimental and calculated COH torsion frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\tau$(COH) HCOOH | | |
|---|---|---|---|
| | *gas-phase* | *molecularly adsorbed* | |
| Exp. | 641[a] | 969[b] | |
| | | MH-intra | MH-inter |
| Harmonic | 680 | 997 | 976 |
| QCT | 641 | 941 | 952 |
| SC | 635 | 935 | 947 |

[a] From reference [189].
[b] The data refer to the low temperature spectrum panel a2 of fig. 4.10.

**CH wagging** Next, we assign the peaks of the experimental spectra of HCOOH at 1038 $cm^{-1}$ (r.t.) and 1079 $cm^{-1}$ (l.t.) to the $\omega$(CH) mode of either MH-intra or MH-inter (or both) and to the $\omega$(CH) of BB(H) configuration (panel (a5)), respectively. Similar peaks occur in the IR spectra of HCOOD, with the low-temperature spectrum presenting a broad signal at 1035 $cm^{-1}$, which we assign to the $\omega$(CH) mode of BB(D) (panel (b5)), and a doublet at 1066 and 1076 $cm^{-1}$. Our calculated spectra of $\omega$(CH) do not present any doublet neither for MD-intra (panel (b3)) nor for MD-inter (panel (b4)). Therefore, we assign the peak at 1066 $cm^{-1}$ to the $\omega$(CH) mode of MD-intra or MD-inter (or both), for which our harmonic, QCT and SC calculations predict lowered wavenumbers with respect to the corresponding modes in HCOOH. Instead, the peak at 1076 $cm^{-1}$ might be due to the presence of HCOOH impurities.

**C-O stretching** Proceeding in the l.t. IR spectra of HCOOH (panel (a2)), we notice a doublet at 1228 and 1255 $cm^{-1}$, which can be resolved, according to our calculations, as the $\nu$(C-O) of MH-inter and MH-intra species. In the experimental deuterated spectra these peaks are shifted at 1259 and 1273 $cm^{-1}$ (panel (b2)).

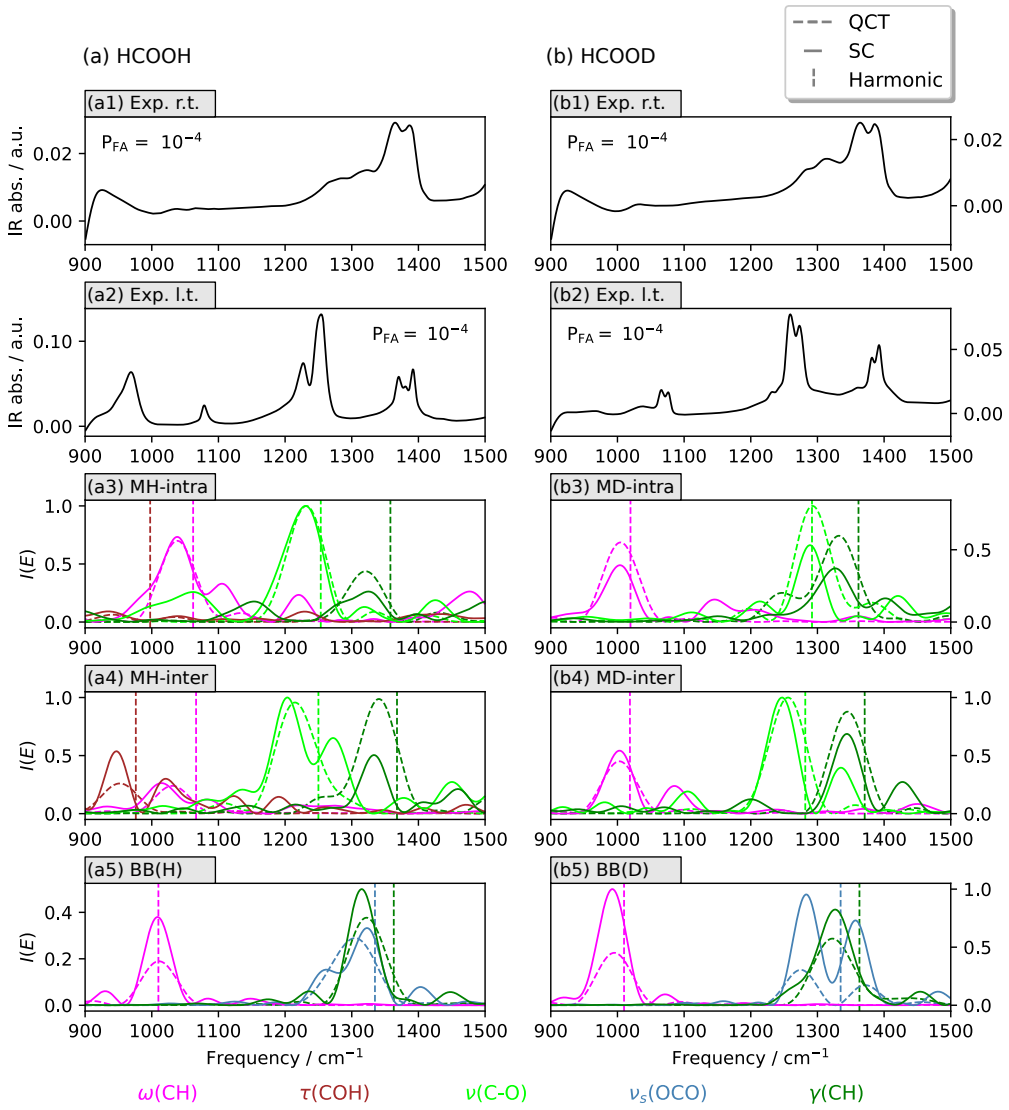



Figure 4.10: The QCT and SC power spectra of the fingerprint region of adsorbed FA compared with the room- (r.t.) and low-temperature (l.t.) FT-IR spectra at a selected partial pressure $P_{FA}$ (mbar). Left: HCOOH adsorption; right: HCOOD.

**OCO symmetric stretching** A peak at $\sim$ 1360 cm$^{-1}$ appears in both IR spectra of HCOOH and HCOOD at room and low temperatures. We attribute this peak to $\nu_s$(OCO) of BB(H) and BB(D) formates. In our calculated SC spectra of BB(H) and BB(D), $\nu_s$(OCO) is splitted into a doublet.

**CH rocking** Lastly, we attribute the signals at $\sim$ 1380-1390 cm$^{-1}$ in the HCOOH spectra to the $\gamma$(CH) vibrational mode of adsorbed HCOOH, which is also present in HCOOD spectra. We are not able to determine if this signal, which is splitted in the experimental low-temperature of about 10 cm$^{-1}$, is due to molecular adsorption or bridging bidentate formate.

### 4.5.2 Carbonyl region

This region (see fig. 4.11) is characterized by the presence of the antisymmetric stretching of OCO of the formate HCOO, $\nu_a$(OCO), and of the $\nu$(C=O) of monodentate configuration.

We confirm the attribution, made previously in the existing literature, of $\nu_a$(OCO) at $\sim$ 1560 cm$^{-1}$,

Table 4.5: Experimental and calculated CH wagging frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\omega$(CH) | | | |
|---|---|---|---|---|
| | HCOOH | | HCOOD | |
| | *gas-phase* | | | |
| Exp.[a] | 1033 | | | |
| Harmonic | 1004 | | 1001 | |
| QCT | 990 | | 983 | |
| SC | 989 | | 982 | |
| | *molecularly adsorbed* | | | |
| Exp. l.t. [b] | 1079 | | 1066, 1076 | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Harmonic | 1062 | 1067 | 1020 | 1019 |
| QCT | 1038 | 1031 | 1005 | 1003 |
| SC | 1038 | 1015 | 1004 | 1003 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.10.

Table 4.6: Experimental and calculated C-O stretching frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\nu$(C-O) | | | |
|---|---|---|---|---|
| | HCOOH | | HCOOD | |
| | *gas-phase* | | | |
| Exp.[a] | 1104 | | 1176 | |
| Harmonic | 1076 | | 1147 | |
| QCT | 1042 | | 1132 | |
| SC | 1039 | | 1130 | |
| | *molecularly adsorbed* | | | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Exp. l.t. [b] | 1255 | 1228 | 1273 | 1259 |
| Harmonic | 1253 | 1250 | 1292 | 1282 |
| QCT | 1232 | 1215 | 1293 | 1256 |
| SC | 1231 | 1203 | 1289 | 1247 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.10.

as seen both at room and low temperature in HCOOH (panels (a1) and (a2)) and HCOOD (panels (b1) and (b2)) adsorption. The harmonic estimates for $\nu_a$(OCO) of BB(H) and BB(D) are lower than the experimental frequencies by about 46 $cm^{-1}$. Such underestimate stays in the QCT and SC calculations, which are significantly red-shifted respect to the harmonic estimate leading to an

Table 4.7: Experimental and calculated CH rocking frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\gamma$(CH) | | | |
|---|---|---|---|---|
| | HCOOH | | HCOOD | |
| | *gas-phase* | | | |
| Exp.[a] | 1379 | | 1365 | |
| Harmonic | 1352 | | 1340 | |
| QCT | 1342 | | 1310 | |
| SC | 1340 | | 1310 | |
| | *molecularly adsorbed* | | | |
| Exp. r.t. (l.t.) [b] | 1387 (1381, 1391) | | 1386 (1382, 1392) | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Harmonic | 1358 | 1368 | 1362 | 1371 |
| QCT | 1321 | 1341 | 1332 | 1345 |
| SC | 1325 | 1333 | 1327 | 1344 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.10.

Table 4.8: Experimental and calculated fingerprint region vibrational frequencies ($cm^{-1}$) of FA adsorbed as bridging bidentate formate.

| | *bridging bidentate* | | | |
|---|---|---|---|---|
| | Exp. r.t. (l.t.)[a] | Harmonic | QCT | SC |
| | HCOOH BB(H) | | | |
| $\omega$(CH) | 1038 (n.o.) | 1010 | 1011 | 1009 |
| $\nu_s$(OCO) | 1365 (1370) | 1335 | 1305 | 1262, 1323 |
| $\gamma$(CH) | 1387 (1381, 1391) | 1363 | 1322 | 1315 |
| | HCOOD BB(D) | | | |
| $\omega$(CH) | 1034 (1039) | 1010 | 995 | 992 |
| $\nu_s$(OCO) | 1364 (1362) | 1335 | 1275,1371 | 1283,1357 |
| $\gamma$(CH) | 1386 (1382, 1392) | 1363 | 1323 | 1327 |

[a] The data refer to formic acid partial pressure reported in the panels of fig. 4.10.

error of almost 100 $cm^{-1}$ with respect to the experiment (table 4.9).

The attribution of $\nu$(C=O) is more controversial. At room temperature the IR spectra of both HCOOH and HCOOD are characterized by $\nu$(C=O) stretching signal at $\sim$ 1675 $cm^{-1}$. Taking into account the intrinsic error due to the PBE approximation, we attribute this peak to the $\nu$(C=O) of molecular monodentate MH- and MD-intra (or -inter) (table 4.10). In particular, the SC spectrum of $\nu$(C=O) in the intra and inter configurations of both HCOOH and HCOOD presents multiple peaks in contrast to the $\nu_a$(OCO) signal. This is mainly due to the strong coupling with the other molecule modes and the phonons of the surfaces. Furthermore, a peak at $\sim$ 1715 $cm^{-1}$ shows up at very

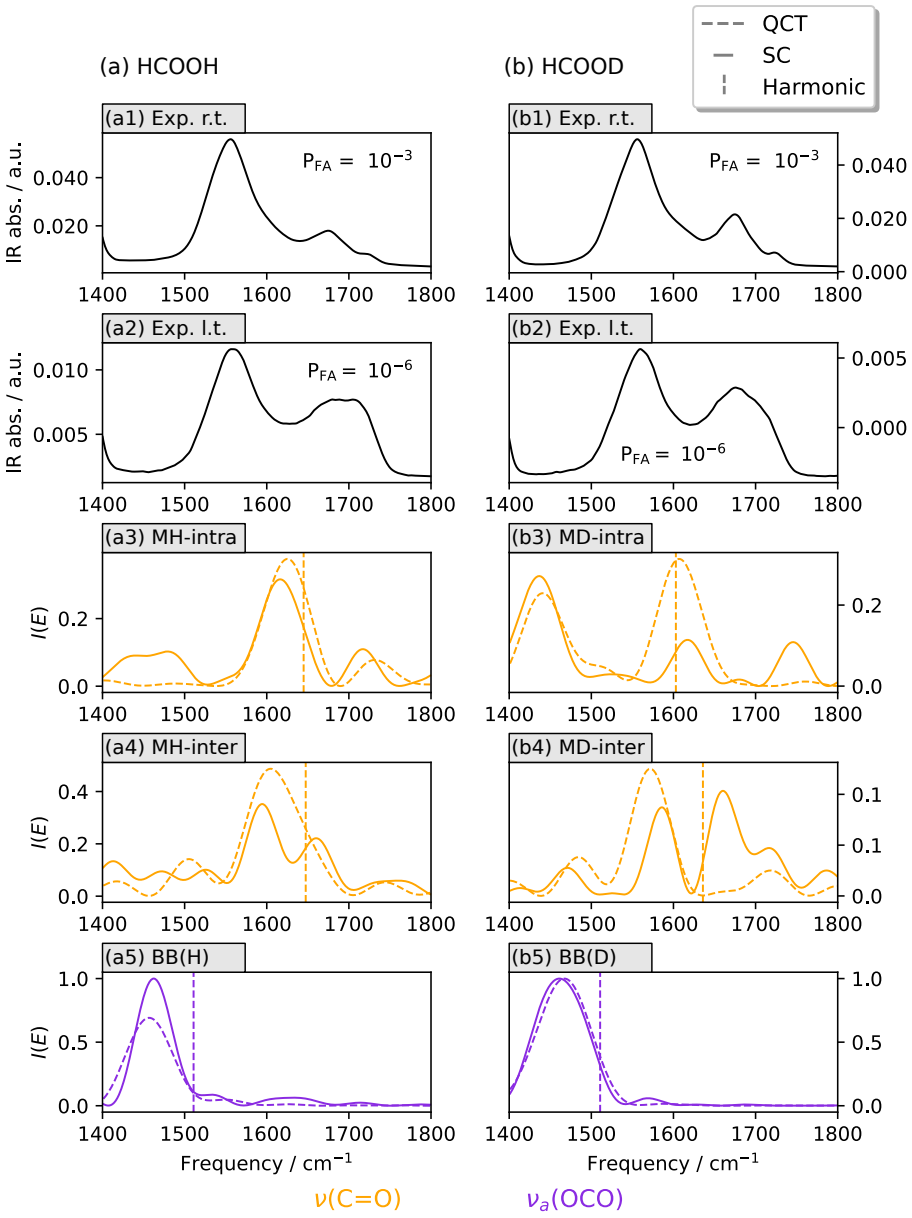



Figure 4.11: The QCT and SC power spectra of C=O and asymmetric OCO stretching vibration of adsorbed FA compared with the room- (r.t.) and low-temperature (l.t.) FTIR spectra at a selected partial pressure $P_{FA}$ (mbar). Left: HCOOH adsorption; right: HCOOD.

low intensity in the room-temperature spectrum of HCOOH and HCOOD and it is significantly enhanced at low temperature. On the basis of our calculations, which predicts the $\nu$(C=O) of molecular monodentate more compatible with the signal at 1675 cm$^{-1}$, we guess that the peak at 1720 cm$^{-1}$ is due to the physisorption of HCOOH and HCOOD on the $TiO_2$ nanoparticles, as suggested by the work of Nanayakkara et al. [187], although we did not study the physisorption of formic acid on anatase.

### 4.5.3 CH stretching

The room-temperature IR spectrum of adsorbed HCOOH shows, in the window 2600-3200 cm$^{-1}$ (fig. 4.12-(a1)), three main bands at $\sim$ 2740, 2870, 2950 cm$^{-1}$. An additional band is present in

Table 4.9: Experimental and calculated OCO asymmetric stretching frequencies ($cm^{-1}$) of FA adsorbed as bridging bidentate formate.

| | *bridging bidentate* | | | |
|---|---|---|---|---|
| | Exp. r.t. (l.t.)[a] | Harmonic | QCT | SC |
| | HCOOH BB(H) | | | |
| $\nu_a$(OCO) | 1556 (1557) | 1511 | 1457 | 1462 |
| | HCOOD BB(D) | | | |
| $\nu_a$(OCO) | 1556 (1559) | 1511 | 1467 | 1461 |

[a] The data refer to formic acid partial pressure reported in the panels of fig. 4.11.

Table 4.10: Experimental and calculated C=O stretching frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\nu$(C=O) | | | |
|---|---|---|---|---|
| | HCOOH | | HCOOD | |
| | *gas-phase* | | | |
| Exp.[a] | 1776 | | 1772 | |
| Harmonic | 1753 | | 1747 | |
| QCT | 1720 | | 1727 | |
| SC | 1718 | | 1725 | |
| | *molecularly adsorbed* | | | |
| Exp. r.t. (l.t.) [b] | 1675 (1680) | | 1675 (1676) | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Harmonic | 1645 | 1647 | 1602 | 1636 |
| QCT | 1625 | 1605 | 1607 | 1572 |
| SC | 1616 | 1594 | 1617 | 1585 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.11.

the low-temperature spectrum of adsorbed HCOOH at 2936 $cm^{-1}$ (fig. 4.12-(a2)). The IR spectra for deuterated FA present similar features (fig. 4.12-(b1), fig. 4.12-(b2)).

As discussed previously, the room-temperature adsorption is dominated by the presence of bridging bidentate species, while at low temperature the formation of molecular monodentate is more favored. Thus, we attribute the band at 2936 $cm^{-1}$ to the CH stretching, $\nu$(CH), of molecular monodentate, while the other bands are associated with bidentate species. The other three bands have been also observed by Nanayakkara et al. [187] at 2735, 2872 and 2951 $cm^{-1}$ and attributed, respectively, to the $\nu_s$(OCO) + $\delta$(CH)[5] combination band, the fundamental $\nu$(CH) and the $\nu_a$(OCO) + $\delta$(CH) combination band of bridging bidentate formate species. Our SC calculations do not show combination bands for BB(H) and BB(D) formates. We notice instead that in the SC spectrum of BB(H) (fig. 4.12-(a5)) the $\nu$(CH) stretching signal is splitted in three main peaks (see values in table 4.12). We interpret this splitting as a Fermi resonance, which can also be detected at the QCT

[5] In our notation $\delta$(CH) corresponds to $\gamma$(CH).

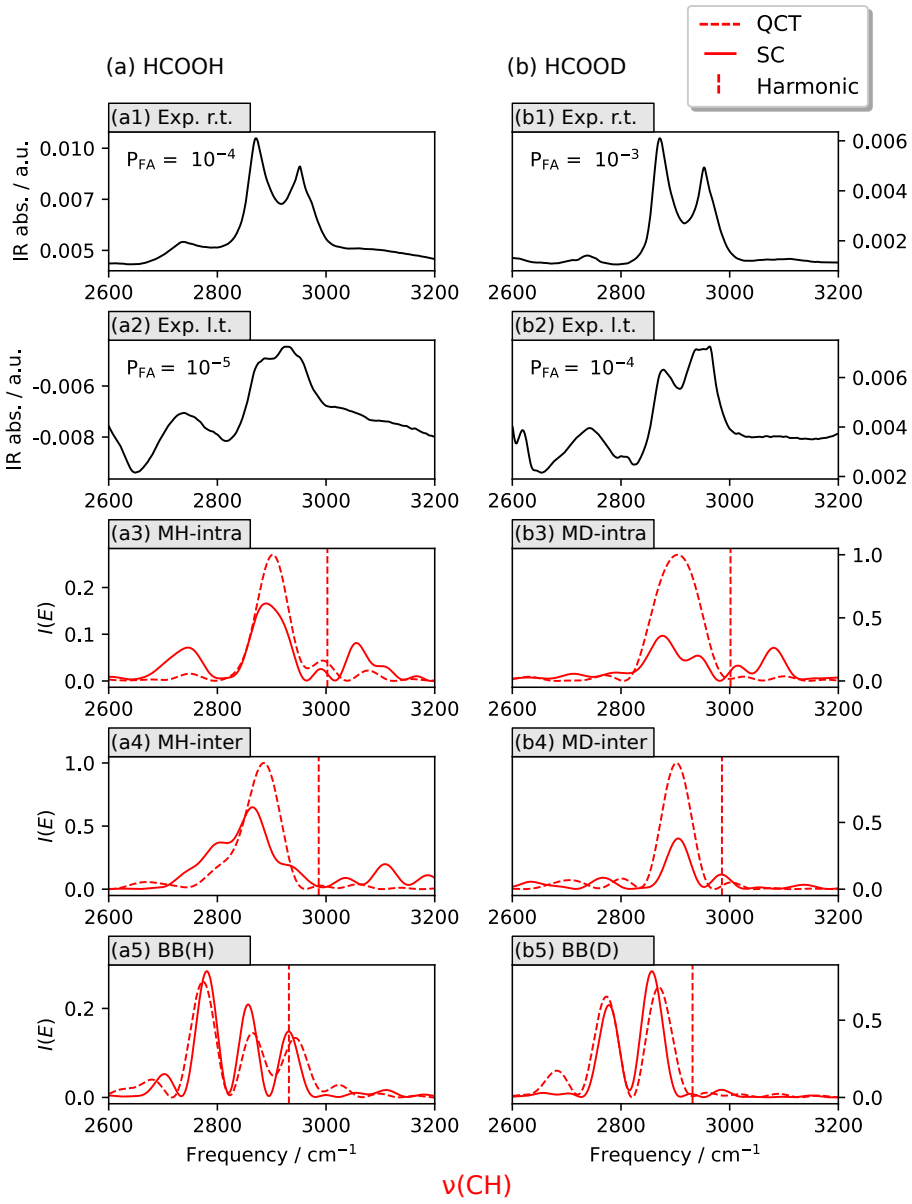



Figure 4.12: The QCT and SC power spectra of CH stretching vibration of adsorbed FA compared with the low- (LT) and room-temperature (RT) FT-IR spectra at a selected partial pressure $P_{FA}$ (mbar). Left: HCOOH adsorption; right: HCOOD.

level. As regarding HCOOD, $\nu$(CH) is splitted in two main components (see values in table 4.12).

### 4.5.4 OH and OD stretching

At the simple harmonic level, the adsorption of protonated formic acid in molecular monodentate species is characterized by a red-shift of the OH stretch on the order of 1000 $cm^{-1}$ with respect to the gas-phase (see table 4.13). This red-shift is caused by the formation of a molecule-surface hydrogen bond that elongates the OH covalent bond when compared to the FA gas-phase (see table 4.2). The resulting PES of the hydrogen and deuterium (see fig. 4.4) is highly anharmonic. Because of anharmonicity and dynamical effects, the hydrogen-bonded OH stretching signal is very complex with multiple peaks (see fig. 4.13). Indeed, the OH stretch couples with both high-

Table 4.11: Experimental and calculated CH stretching frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\nu$(CH) | | | |
|---|---|---|---|---|
| | HCOOH | | HCOOD | |
| | *gas-phase* | | | |
| Exp.[a] | 2942 | | 2954, 2938 | |
| Harmonic | 2957 | | 2958 | |
| QCT | 2886 | | 2848 | |
| SC | 2893 | | 2847 | |
| | *molecularly adsorbed* | | | |
| Exp. l.t.[b] | 2936 | | 2938 | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Harmonic | 3002 | 2987 | 3002 | 2896 |
| QCT | 2902 | 2885 | 2904 | 2902 |
| SC | 2889 | 2865 | 2877 | 2905 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.12.

Table 4.12: Experimental and calculated CH stretching frequencies ($cm^{-1}$) of FA adsorbed as bridging bidentate formate.

| | *bridging bidentate* | | | |
|---|---|---|---|---|
| | Exp. r.t. (l.t.)[a] | Harmonic | QCT | SC |
| | HCOOH BB(H) | | | |
| | 2738 (2739) | | 2773 | 2780 |
| $\nu$(CH) | 2871 (2875) | 2932 | 2865 | 2857 |
| | 2952 (2954) | | 2943 | 2931 |
| | HCOOD BB(D) | | | |
| | 2739 (2742) | | 2773 | 2778 |
| $\nu$(CH) | 2872 (2878) | 2932 | 2870 | 2857 |
| | 2953 (2953) | | | |

[a] The data refer to formic acid partial pressure reported in the panels of fig. 4.12.

and low-frequency modes and with both internal modes of the molecule and the phonon modes of the surface. This strong anharmonic modes mixing makes very difficult a pure assignation of the OH fundamental stretch.

In the procedure as employed in previous DC-SCIVR calculations on adsorbed molecules on the $TiO_2$ anatase (101) surface [85, 86], the semiclassical spectra were generated from a single classical trajectory at an energy equal to the harmonic estimate of the ZPE. In order to be able to provide a better resolution of the OH and OD stretching signals and increase the quality of the QCT and SC spectra, we have varied the initial momenta of the classical trajectory by decreasing the amount of the harmonic ZPE at which the system is initialized. As an example, we present in Figure 4.13 the

QCT spectrum of OH stretch for a selected configuration of HCOOH, MH-intra, with respect to the variable coefficient $\alpha$ ($1/4 \leq \alpha \leq 1$), which sets the initial condition of the nuclei momenta: $E_{kin,0} = \alpha \times E_{ZPE,Harmonic}$. The time evolution of the $O_fH$ ($f$ = formic) and $O_sH$ ($s$ = surface) distances from the corresponding classical NVE trajectories are reported in appendix, fig. C.5.

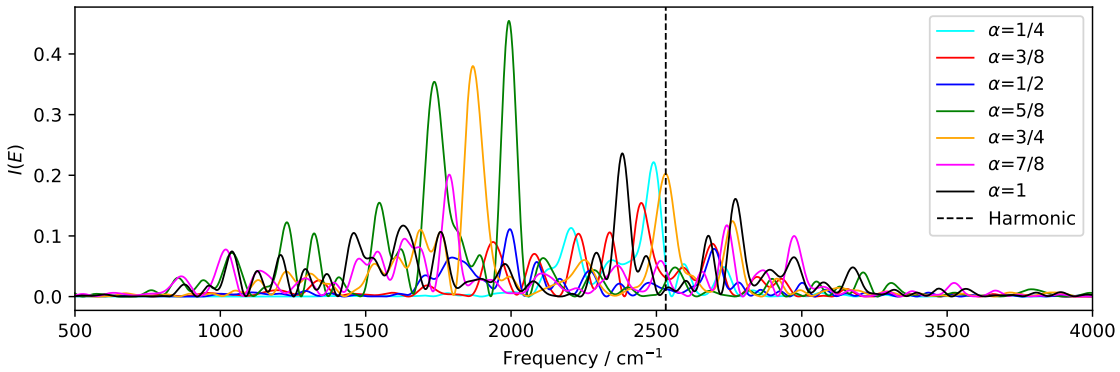



Figure 4.13: QCT spectra of OH stretching mode of FA adsorbed in MH-intra configuration respect to different initial conditions given by $\alpha$ ($1/4 \leq \alpha \leq 1$): $E_{kin,0} = \alpha \times E_{ZPE,Harmonic}$.

At $\alpha = 1$, the momenta are initialized to a temperature corresponding to the total harmonic estimate of the ZPE, causing an excitation of both low- and high-frequency modes. The acid proton jumps several times back and forth from the molecule to the surface. The OH stretch signal presents multiple peaks in a wide window of frequencies and yields the hydrogen-bonded monodentate species a complex floppy nature. We underline that the high-intensity peaks for frequencies below 2000 $cm^{-1}$ are due to the strong coupling with the rest of vibrational modes of the molecule. By decreasing $\alpha$, the system energy is closer to the bottom of the potential-well represented in fig. 4.4. As a consequence, the proton hopping between the molecule and the surface is generally reduced and the QCT spectra become more alike to the harmonic ones.

The SC calculations for the OH and the OD stretch have actually been performed at $\alpha = 1, \alpha = 3, 4$, $\alpha = 1/2$. Figure C.7 and fig. C.7 report all the SC spectra of OH and OD stretch at different values of $\alpha$. In the following, we present the QCT and SC spectra with the best defined signal and compare them with the IR experiments. As the OH and CH stretching signals are close in frequency, we remind the main conclusion of the previous section 4.5.3. The OH stretching for chemisorbed molecular FA on $TiO_2$ nanoparticles was attributed at frequency 2591 $cm^{-1}$ by Nanayakkara and coworkers [187]. By comparison of the room- (fig. 4.14-(a1)) and the low-temperature experiment (fig. 4.14-(a2)), we identify the new peak in the low-temperature spectrum at $\sim$ 2560 $cm^{-1}$ as a possible candidate for the OH stretching vibration.

The harmonic estimate of the OH stretch in MH-intra is in good agreement with this attribution (table 4.13). The QCT trajectory at $\alpha = 3/4$ of MH-intra (fig. 4.14)-(a3)) gives a peak at the harmonic estimate. When including quantum effects, the SC calculation shows the same peak at slightly lower wavenumbers. In the case of MH-inter (fig. 4.14)-(a3)), the SC OH stretch signal is blue-shifted with respect to the harmonic value and it is found at 2718 $cm^{-1}$. The OH stretching fundamental of MH-intra are in better agreement with the experiments than the one of the MH-inter at all the three levels of calculations (harmonic, QCT, SC). We will therefore, for simplicity, continue the discussion by referring to the MH-intra configuration.

Table 4.13: Experimental and calculated OH and OD stretching frequencies ($cm^{-1}$) of FA gas-phase and molecularly adsorbed.

| | $\nu$(OH) HCOOH | | $\nu$(OD) HCOOD | |
|---|---|---|---|---|
| | | *gas-phase* | | |
| Exp.[a] | 3570 | | 2631 | |
| Harmonic | 3600 | | 2616 | |
| QCT | 3489 | | 2562 | |
| SC | 3493 | | 2563 | |
| | | *molecularly adsorbed* | | |
| Exp. l.t.[b] | 2558 | | 2062, 2170 | |
| | MH-intra | MH-inter | MD-intra | MD-inter |
| Harmonic | 2532 | 2592 | 1872 | 1919 |
| QCT | 2532 | 2622 | 1951, 2043 | 1922, 2076 |
| SC | 2529 | 2718 | 1956, 2089 | 1850, 2185 |

[a] From reference [189].
[b] The data refer to formic acid partial pressure reported in the panels of fig. 4.14.

As introduced previously in section 4.3.3, the deuteration of FA causes a H/D isotope shift. It is clear from both the experiments and our calculated spectra, that the OH stretch band shifts to lower wavenumbers. The experimental low-temperature IR spectrum of HCOOD (fig. 4.14)-(b2)) is characterized by the presence of two large bands in the 1900-2200 $cm^{-1}$ region. The MD-intra SC spectrum (fig. 4.14)-(b3)) presents two peaks at 1956 and 2089 $cm^{-1}$, while the harmonic estimate is at 1872 $cm^{-1}$. To summarize, the OH and OD stretching spectra of the monodentate intra configuration are in well agreement with the experimental findings. On the one hand, the OH stretching vibration in HCOOH is red-shifted respect to the "free" vibration in the gas-phase by 1068 $cm^{-1}$ at the harmonic level. This huge red-shift is caused by the formation of a strong hydrogen bond with the surface and decreased of about 100 $cm^{-1}$ when anharmonic and quantum effects are included. On the contrary, the HCOOD shift respect to the gas-phase is smaller at the harmonic level (744 $cm^{-1}$) and it is even less when anharmonic and quantum effects are taken into account: it is 607 and 474 $cm^{-1}$ for the two OD stretching peaks, respectively. The previous considerations suggest that the hydrogen bond in adsorbed HCOOD is weaker than in adsorbed HCOOH, making the OD stretching to vibrate at frequencies closer to the "free" OD stretch in the gas-phase HCOOD system.

Concerning the bridging bidentate species, the OH stretch signal coming from the free hydrogen atom on the surface is a well defined single peak at 3664 $cm^{-1}$ at the harmonic level, which shifts down to 3563 $cm^{-1}$ in semiclassical calculations. The experimental band at $\sim$ 3700 $cm^{-1}$ in the room and low temperature spectrum of adsorbed HCOOH, which is also present in the room temperature spectrum of HCOOD (see fig. C.3), could be attributed to the OH stretching mode in the bidentate formate. This signal was previously attributed to surface OH hydroxyl groups coming from water dissociation, most likely on surface defects [158]. Indeed the negative absorption intensity due to the subtraction with the background refers to the chemical species that were already present in the chamber and, therefore, are not coming from the FA molecule and are most likely due

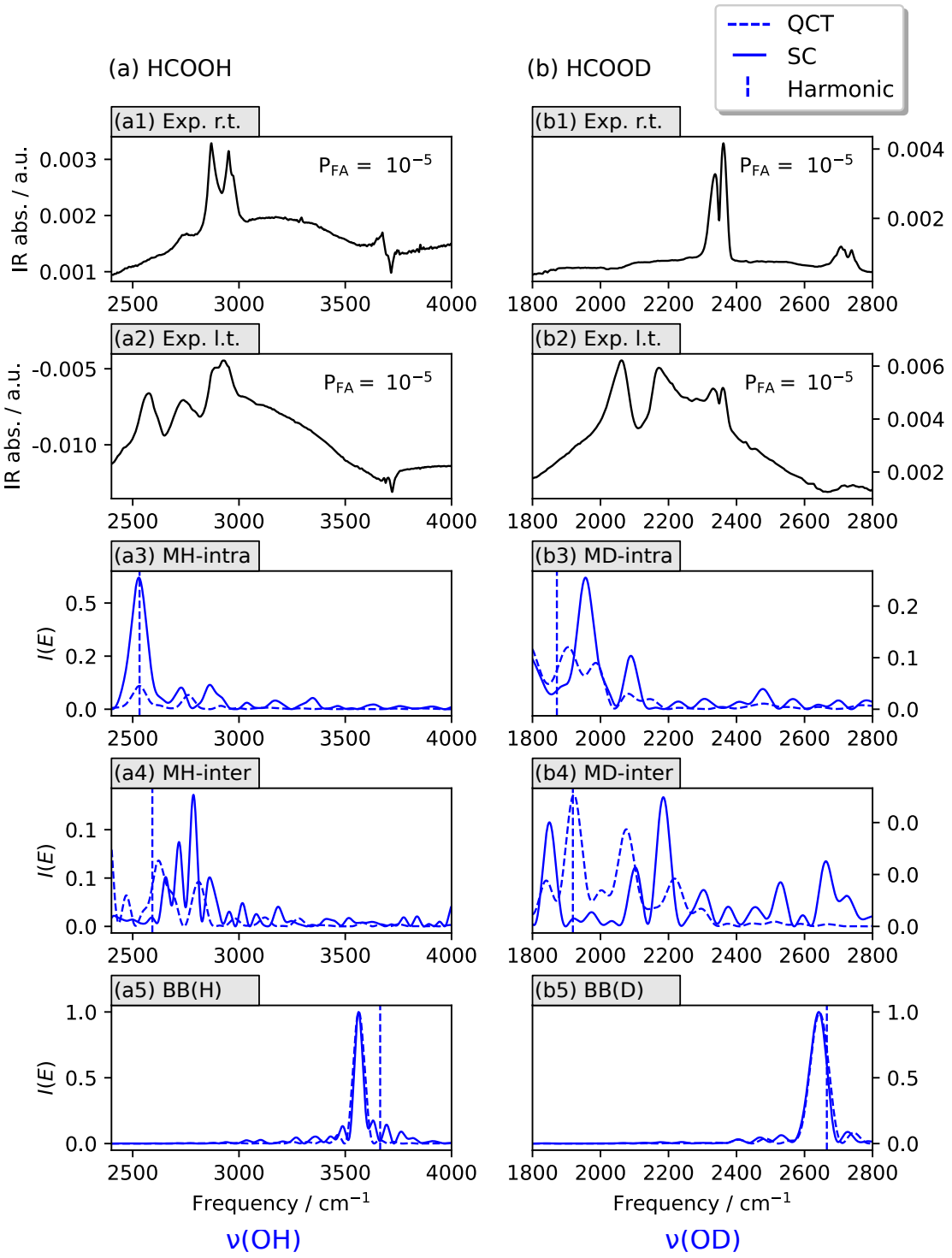



Figure 4.14: The QCT and SC power spectra of OH and OD stretching vibration of adsorbed FA compared with the room- (r.t.) and low-temperature (r.t.) FTIR spectra at a selected partial pressure $P_{FA}$ (mbar). We remind that the doublet at $\sim 2300$ cm$^{-1}$ is due to the $CO_2$ present in the chamber. Left: HCOOH adsorption; right: HCOOD. All the QCT and SC refer to a single classical trajectory with $\alpha = 1$, except MH-intra ($\alpha = 3/4$) and BB(D) ($\alpha = 1/2$).

to water vapor molecules. It is unclear whether the dissociation of FA via the bridging bidentate adsorption should contribute in some way to this signal.

On the contrary, for the deuterated system it is possible to assign a $\nu$(OD) stretching frequency

coming from HCOOD dissociation. At the harmonic level the OD stretch of BB(D) is 2666 $cm^{-1}$ and shifts at 2642 $cm^{-1}$ in SC calculation using $\alpha = 1/2$. These frequencies are compatible with the experimental signal at $\sim$ 2717 $cm^{-1}$, which is clearly visible in the room temperature spectrum of adsorbed HCOOD. A similar peak was found by Wang et al. [173] and attributed to a OD stretch. It is interesting to note that if $\alpha = 1$, it is very difficult to catch the fundamental of OD stretch in QCT and SC spectra (see in appendix, fig. C.7). This is due to the fact that the free OD on the surface is strongly coupled with low-frequency phonon modes of the surface, which are below 500 $cm^{-1}$. This results in a very broad signal for the SC calculation with $\alpha = 1$.

Table 4.14: Experimental and calculated OH and OD stretching frequencies ($cm^{-1}$) of FA adsorbed as bridging bidentate formate.

| | | *bridging bidentate* | | |
|---|---|---|---|---|
| | Exp. r.t. (l.t.)[a] | Harmonic | QCT | SC |
| | | HCOOH BB(H) | | |
| $\nu$(OH) | n.o. | 3664 | 3564 | 3563 |
| | | HCOOD BB(D) | | |
| $\nu$(OD) | 2717 (2718) | 2666 | 2644 | 2642 |

[a] The data refer to formic acid partial pressure reported in the panels of fig. 4.14.

## 4.6 Coupling with the surface

One advantage of using the DC-SCIVR method is the division in subspaces of the different vibrational modes according to the averaged Hessian criterion (eq. (2.60)), which helps to identify the relevant coupling between the different vibrational modes, which is completed neglected in the harmonic approximation. The modes with the strongest interaction will be contained in the same subspace. Tables C.18 and C.19 report for HCOOH and HCOOD, respectively, the dimension and the threshold $\epsilon$ of the different subspaces employed to compute the presented semiclassical spectra. For the molecular adsorption of HCOOH (intra/inter) the subspaces contain all the modes of the molecule and the modes coming from the interaction with the surface, while for the bridging bidentate adsorption the formate HCOO and the OH stretching were treated from two different subspaces. To exhaustively study the complex coupling between the molecule and the surface, one should consider and study all the molecule-surface modes contained in each subspace that we used to compute the SC spectra. In the following, we will limit to analyze some relevant examples for each adsorption configuration. To visualize these surface-molecule interactions, we use the displacement arrows that describe the harmonic eigendisplacements, which are mass-scaled so that a longer arrow corresponds to a larger atomic displacement (fig. 4.15). Furthermore, for simplicity sake we will give the harmonic frequencies for each mode.

The molecule-surface modes can be generally divided into two main categories: resonances between the adsorbant and the surface phonons, on the one hand, and localized modes on the adsorbant molecule, on the other hand.

For example, the mode #144 of MH-intra is a strong resonance between the adsorbant and the sur-

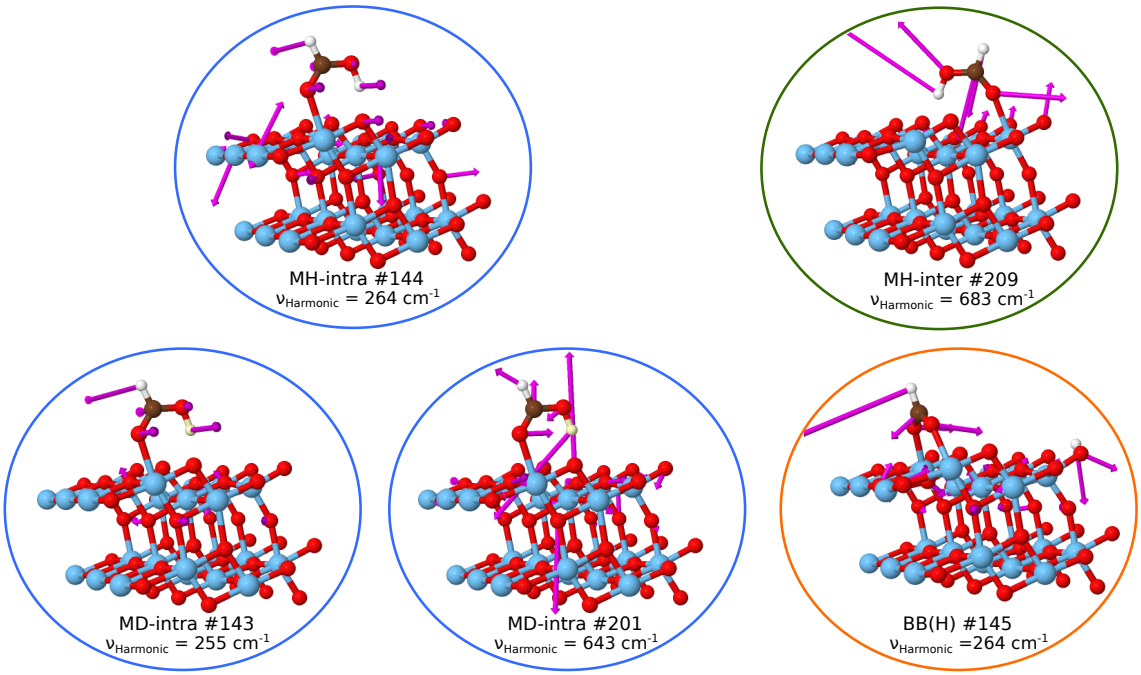



Figure 4.15: The eigendisplacements of some relevant molecule-surface vibrational modes that are coupled to the adsorbant modes in the three different adsorption configurations. The arrows illustrate the movement of each atom. Larger is the mass-scaled atomic displacement, longer is the arrow of the corresponding atom.

face phonons. In particular, this mode can be described as a bending of the OH$\cdots$O hydrogen bond accompanied by displacements of the atoms of the surface, mainly oxygen atoms, as underlined by the long arrow displacements. Such mode has a very low harmonic frequency, 264 cm$^{-1}$, compared to the frequencies of the adsorbant which are in the range 900-3000 cm$^{-1}$. Despite this large difference, mode #144 of MH-intra is contained in the same subspace of the adsorbant modes and therefore it is strongly coupled to them. When considering the corresponding deuterated configuration, MD-intra, we focus on the modes #143 and #201. The mode #143 of MD-intra has a similar motion of the adsorbant atoms of mode #144 of MH-intra, with the difference that the oxygen atoms of the surface participate very little to the overall displacements. This can be therefore characterized as a localized mode of the adsorbant. The mode #201 of MD-intra is a strong resonance that interest particularly the hydrogen-bonded region. Here we observe a general deformation of the OCO group of the adsorbant and a significant displacement of the $O_{2c}$ site that is hydrogen-bonded to the deuterium, which is counterbalanced by the nearest neighbor $O_{3c}$ site bond to the $Ti_{5c}$ site of adsorption. The harmonic frequency of this mode is 643 cm$^{-1}$, which is very close to the $\delta$(OCO) of gas-phase FA (605 cm$^{-1}$, table 4.3) and to phonon modes of the clean surface in the range 630-648 cm$^{-1}$. For the inter pair configuration the mode #209 is a weak localization mode of a O-C-O bending, which is accompanied by a small displacements of surface oxygen atoms; the one which contributes the most is the neighboring $O_{2c}$ site of the $Ti_{5c}$ adsorption site. Lastly, we report an example for the BB(H) configuration. As anticipated before, the OH stretching due to HCOOH dissociation in a BB formate is not contained in the same subspace of the HCOO formate vibrational space. However, this does not mean that the formate and the free OH of the surface are not affected by the presence of one another. Indeed, the mode #145 of BB(H)

is a strong resonance between a out-of-plane bending of the O-C-O group and a movement that affects the hydroxyl group on the surface. The frequency of mode #145 is 264 $cm^{-1}$, quite low if compared to the characteristic formate vibrations, which are in the 1000-3000 $cm^{-1}$ range. Note, again, despite the large differences of frequencies, that mode #145 is dynamically coupled with the formate vibrations.

## 4.7 Quantum equilibrium properties

In section 4.5, we have shown how much it is important to include both anharmonic and quantum effects to properly describe the dynamics of the investigated adsorption phenomenon. As most interesting properties, we have highlighted the importance of the hydrogen-bond in molecular monodentate, which highly affects the vibrational properties of both adsorbed HCOOH and HCOOD. In this last part of the chapter, we therefore focus on the molecular monodentate configuration and the impact of NQEs on the equilibrium properties of the system through PIMD simulations. In particular, we have used a path integral generalized Langevin equation thermostat [46, 47] to reduce the number of replicas in the PIMD simulations. In the following, we will illustrate the probability distributions of bond lengths and the temperature-dependent observables. Firstly, a comparison with the gas-phase molecule is presented. Then, we will discuss the thermal and isotope effects. Finally, we investigate the proton localization and the proton transfer between the molecule and the surface.

### 4.7.1 Formic acid before ad after adsorption

To measure the impact of quantum effects in mondentate configuration, we have firstly compared the main bond lengths distances of FA before and after adsorption at low temperature (Figure 4.16) through PIMD method. The presence of a H-bond with the surface makes the probability function of the O–H distance highly asymmetric, which causes the elongation of the O–H covalent bond length with respect the isolated gas-phase FA. Very tiny effects occur on the C–H bond length. These results are completely in accordance to what we have found for the vibrational spectra, which, with respect to the gas-phase, predict a red-shift for $\nu$(C=O) and $\nu$(OH), a blue-shift for $\nu$(C-O) and very similar frequency for $\nu$(CH).

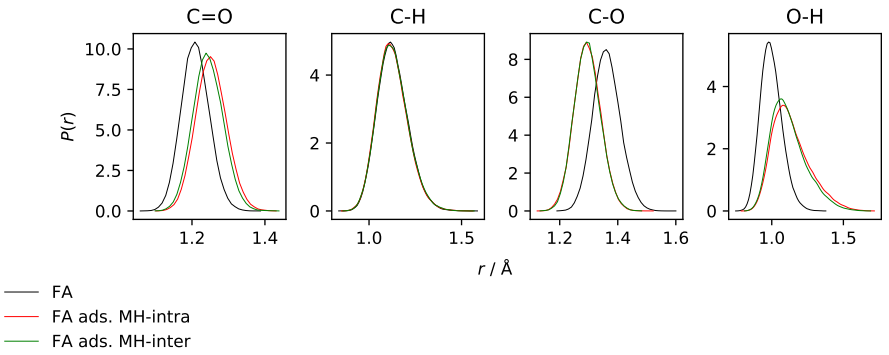



Figure 4.16: Probability distributions at 100 K of bond lengths of FA molecule before and after adsorption via monodentate mode, obtained from PIMD simulations.

### 4.7.2 Thermal and isotope effects

Figure 4.17 reports the probability functions for the inter-atomic distances close to the adsorption site at 100 K and 300 K. As a general trend, we point out that classical distributions are narrower than the quantum ones especially for O–H covalent bond and O$\cdots$H lengths. The largest difference between classical and quantum result is at low temperature, which is mainly due to the importance of the ZPE effects. However, we notice, that even at 300 K the classical distribution for the O–H covalent bond length is still quite different from the quantum case. The hydrogen bond distance O$\cdots$H is peaked at $\sim$ 1.40 Å ($\neq$ statistical average 1.38) for HCOOH, while O$\cdots$D is peaked at $\sim$1.45 Å ($\neq$ statistical average 1.43) in HCOOD. Therefore, the hydrogen bond is shorter in HCOOH than HCOOD. We remind that this is completely the opposite effect of what we have found in the crystal system KOH/KOD (section 3.5.3), where the hydrogen bonds are long and weak. Here, the hydrogen bonds are short and strong and the quantum fluctuations due to the stretching motion of OH and OD prevail on the bending ones and reinforce the hydrogen bond. Thus, the O$\cdots$H length predicted by the classical Langevin dynamics is the longest (weakest interaction), and the hydrogen bond is stronger in adsorbed HCOOH than adsorbed HCOOD, confirming what we suggested in section 4.5.4. Also $r_3$ ($Ti_{5c}$–$O_{2c}$) and $r_4$ ($Ti_{5c}$–OC) distances are affected by NQEs as evidenced by the broadening of the distributions and the shift of the peaks with respect the classical ones at low temperature, confirming that all the environment close to the hydrogen bond is involved.

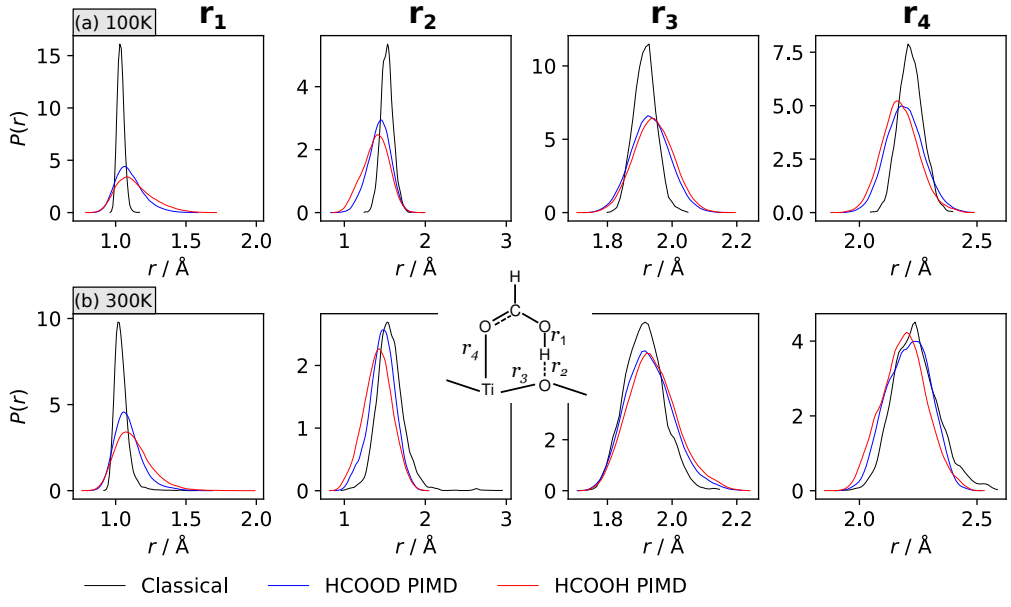



Figure 4.17: Probability distributions at (a) 100 K and (b) 300 K of inter-atomic distances $r_1$, $r_2$, $r_3$ and $r_4$ as sketched in the plot of protonated FA and deuterated FA adsorbed in the monodentate intra mode, obtained from PIMD simulations. The probabilities are compared to the corresponding classical results. The scales over the x-axis are the same for the two temperatures, but are different for $r_1$, $r_2$, $r_3$ and $r_4$.

### 4.7.3 Proton localization

In fig. 4.4 we have shown that the PES for monodentate mode is characterized by a well defined minimum corresponding to molecular monodentate MH-intra, while the dissociated configuration is highly unfavorable and does not represent a minimum. The potential is therefore characterized by a single well presenting a high level of anharmonicity. By looking at the fig. 4.18, the time evolution at 100 K of O–H covalent bond distance and O$\cdots$H hydrogen bond distance for the centroid is characterized by several hopping events in a short time frame of about just 1 ps, which are significantly less for HCOOD.

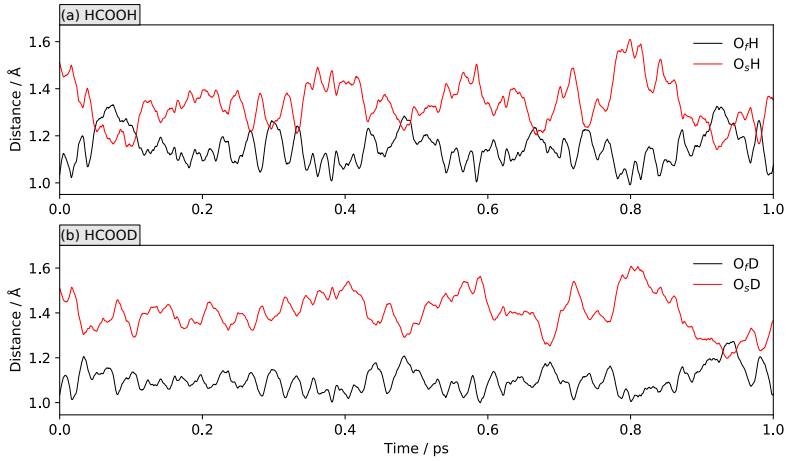
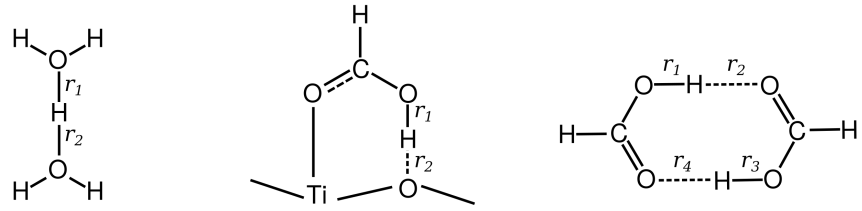



Figure 4.18: Time evolution at 100 K of O–X (X=H,D) covalent bond distance ($O_f$X) and O$\cdots$X hydrogen bond distance ($O_s$X) in a time frame of 1 ps. Panel (a) HCOOH, panel (b) HCOOD. The time evolution refers to the centroid trajectory starting from MX-intra mode.

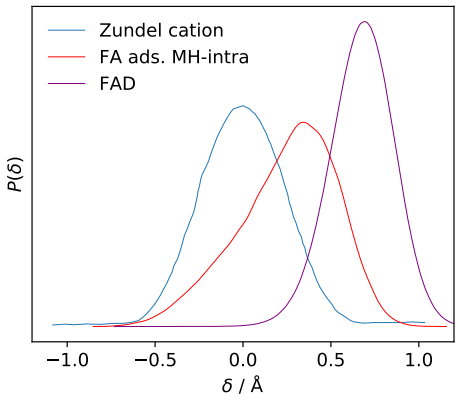
To appreciate the quantum properties of HCOOH/$TiO_2$ system, we have made a comparison of the distribution of the proton transfer coordinate at 100 K between MH-intra mode and other well-known systems, the formic acid dimer (FAD) and the zundel cation. The FAD was modeled using a fitted potential implemented by Qu et al. [190], while the zundel cation probability distributions were extracted from the work of Shran et al. [191]. The proton sharing coordinate $\delta$ is defined as the following: $\delta = (r_2 - r_1)$ for I and II and $\delta = (r_2 - r_1)$ and $\delta = (r_4 - r_3)$ for III. The zundel cation represents the extreme condition in which the inter-oxygen distance is just about 2.40 Å, the hydrogen bond is symmetrized and the proton is equally shared between the two oxygens. The formic acid dimer instead represents the opposite case, the hydrogen bonds are much more elongated corresponding to an equilibrium inter-oxygen distance of 2.68 Å. This results in low proton hopping events and the proton is localized on the donor sites. The $O_f$-$O_s$ distance of the FA adsorbed in the monodentate configuration is between the previous values. However, in contrast with both the zundel cation and the FA dimer, there is no mirror plane passing through the centered position. In the proton distribution for the adsorbed molecule, there is a single maximum for $\delta > 0$. The proton dissociation (corresponding to $\delta < 0$) is not an equilibrium position, even when NQEs are included. However, the short hydrogen bond (averaged $O_f$-$O_s$ distance = 2.47 Å) leads the hydrogen bond donor and acceptor sites to share much more the proton compared to the formic acid dimer. This results in highly anharmonic quantum properties which affect the OH stretching signal. The symmetric case $\delta = 0$ is not achieved due to the fact that, differently from the zundel

cation, this system is heterogeneous and the potential is not symmetric.




I: Zundel Cation II : FA@A101 MH-intra III : FAD

Figure 4.19: Sketch of the chemical structure of (I) zundel cation, (II) formic acid adsorbed on anatase (101) surface in monodentate mode (intra) and (III) formic acid dimer.




| | $\langle r_{OO} \rangle$ [Å] |
|---|---|
| I Zundel cation | 2.40 |
| II FA ads. MH-intra | 2.47 |
| III FAD | 2.68 |

Figure 4.20: Probability distribution at T = 100 K of proton sharing coordinate $\delta$ in (I) zundel cation from Shran et al. [191] (Figure 4, curve at 2.4 Å) , (II) formic acid adsorbed on anatase (101) surface in monodentate mode (intra) and (III) formic acid dimer using PES from Qu et al. [190].

## 4.8 Conclusion

In this chapter, we have addressed the complex phenomenon of adsorption of formic acid molecule on the $TiO_2$ anatase (101) surface by incorporating quantum and anharmonic effects in the molecular dynamics simulations. This study was conducted in close collaboration with some experimentalists, who performed new FTIR spectra on anatase nanopowders. Our results reveal the presence of competing adsorption configurations, as suggested in previous works. On the one hand, the formic acid molecule can adsorb in a molecular form; on the other hand, it can dissociate forming a formate species and free hydroxyl groups on the surface. According to our calculations, the most stable configuration is the molecular monodentate mode intra-pair, which presents an hydrogen bond with a $O_{2c}$ site of the surface. The dissociation via similar monodentate mode is highly unfavoured and does not represent a meta stable configuration. Instead, the dissociation presents a local minimum for a bridging bidentate configuration, where the OCO plane of the formate is parallel to the [010] direction.

The adsorption picture appears to be very complex, with experimental spectra at room and low temperature rich of different spectroscopic features. The room-temperature spectra is dominated by the presence of bridging bidentate species due to the presence of the characteristic $\nu_a$(OCO) and $\nu_s$(OCO) of the formate. However, at low temperature the adsorption of the FA molecule is different. Indeed, the low-temperature experiments revealed the formation of a more stable species, which, under specific deuteration of the acid proton position (HCOOD), presents an isotopic H/D shift of some bands.

To assign the surface species, we have compared the FTIR spectra with the calculated power spectra obtained through QCT and DC-SCIVR approaches. In this work, we have employed the PBE functional, which is very accurate to describe the hydrogen bonds but underestimates the harmonic frequencies, as we carefully benchmarked the gas-phase FA against higher levels of DFT. Despite the systematic error due to the PBE functional for some frequencies, we were able to assign some of the feature of the room temperature and low temperature spectra of both adsorbed HCOOH and HCOOD. Our calculations are compatible with the assignment of a molecular monodentate species at low temperature, although some bridging bidentate is also present. In particular, we focused on the OH and OD stretching vibrations of the hydrogen-bonded molecular monodentate species (intra-pair). The OH stretching frequencies for adsorbed molecular monodentate is close the CH stretching region at $\sim$ 2550 cm$^{-1}$. The harmonic estimate is in good accordance with this result, however completely fails to describe the behavior of molecular monodentate deuterated formic acid, which shows in the experiments two main bands at $\sim$ 2060 and 2170 cm$^{-1}$. Indeed, with the inclusion of anharmonic effects we revealed that the OH stretching is significantly more red-shifted with to respect the gas-phase than the OD stretch, suggesting that the hydrogen bonds are stronger in adsorbed HCOOH than adsorbed HCOOD, mainly due to the zero-point energy effects.

The adsorption of HCOOH and HCOOD presents a mixing of the modes coming from the molecule and the surface. Through investigation of the subspace used in the DC-SCIVR spectra, we have shown that both resonances and localized modes of the adsorbant with the surface atoms, mainly oxygen atoms, contribute to the complex picture of the adsorption, since they are dynamically

coupled with the vibrational modes coming from the adsorbant.

PIMD distributions confirmed that NQEs are essential to properly describe the monodentate hydrogen-bonded molecular adsorption, and predict a geometric isotope effect for the hydrogen-bonds length, which is shorter in HCOOH in accordance to what we found for the vibrational features. Furthermore, we have shown trough PIMD approach that the monodentate FA/$TiO_2$ system behaves quite differently with respect to other well-known systems, the zundel cation and the formic acid dimer. To summarize, in the FA/$TiO_2$ system the molecular adsorption is favored even when including nuclear quantum effects, although the proton is frequently shared between the surface and the molecule.

# APPENDIX C

## C.1 Computational details

***ab initio* level** First principles DFT calculations were performed using the QE package [180]. We employ the PBE functional [26] for the exchange and correlation potential and norm-conserving short-range pseudopotentials from Pseudo Dojo distribution [192] (v0.4). The energy cutoff for the wavefunction was expanded up to 72 Ry (see fig. C.1). The Brillouin zone was sampled with the only $\Gamma$ point in surface calculations. Bulk anatase properties with this computational set-up are reported in table C.15.

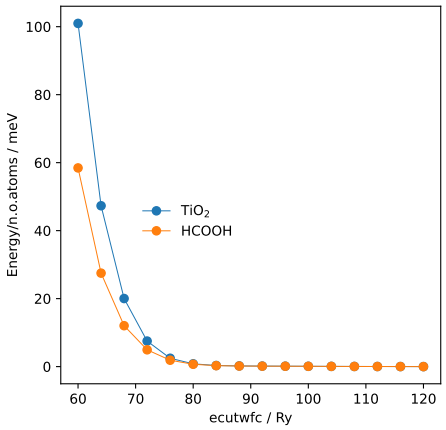



Figure C.1: Convergence respect to the energy cutoff for the wavefunctions.

Table C.15: Calculated lattice parameters (Å) and energy gap (eV) for bulk anatase $TiO_2$ compared with the available experimental data [143].

| | a | c | Gap |
|---|---|---|---|
| Calc. | 3.791 | 9.649 | 2.2 |
| Exp. | 3.782 | 9.502 | 3.2 |

**Relax calculations** All the structural optimization were carried out within a convergence threshold on forces of $\sim 0.001$ eV/Å. The anatase (101) surface was modeled using a periodic slab formed by 4 $Ti_6O_{12}$ layers (72 atoms total), which corresponds to a $(1 \times 3)$ surface unit cell and surface area 10.37 $\times$ 11.37 Å$^2$. A vacuum region $\sim 13$ Å wide was included to reduce the inter-slab interactions. The surface was build as the following. Bulk anatase containing 4 $TiO_2$ units (12 atoms total) and defined by the lattice vectors $(a_1, a_2, a_3)$ was optimized with a variable cell structural optimization. The primitive unit cell was then transformed in $(a'_1, a'_2, a'_3)$ by choosing $a'_1 = a_1 - a_3$ and $a'_2 = 3 \times a_2$ as the (101) plane vectors and $a'_3 = 3 \times a_1$ as the stacking vector. A slab of 6 $Ti_6O_{12}$ layers (108 atoms) was generated and the fully relaxation of the atoms with the exception of the 2 bottom layers was performed. The 6 layers slab was cut in order to obtain the final 4 $Ti_6O_{12}$ layers slab. A relax calculation with fixed bottom layers and free surface layers was performed.

Table C.16: Calculated inter-atomic distances (Å) for anatase $TiO_2$ (101) surface compared with the available experimental data [193].

| | Exp. | $Ti_6O_{12}$ (4 Layers) |
|---|---|---|
| Ti(1)-O(1) | 1.90 | 1.84 |
| Ti(1)-O(2) | 1.94 | 1.98 |
| Ti(1)-O(3) | 2.07 | 2.06 |
| Ti(1)-O(4) | 1.99 | 1.78 |
| Ti(2)-O(1) | 1.89 | 1.84 |
| Ti(2)-O(2) | 1.97 | 2.01 |
| Ti(2)-O(3) | 1.92 | 1.94 |
| Ti(2)-O(4) | 1.89 | 2.09 |
| Ti(2)-O(5) | 2.08 | 2.15 |

The adsorption of *trans*-FA molecule was modeled considering the adsorption of one molecule of FA in both monodentate (molecular/dissociated) and bidentate (dissociated) adsorption (limit weak coverage). The minimum distance between periodic copies is respectively $\sim 8$ and $\sim 11$ Å in $[\bar{1}01]$ in [010] direction for monodentate and $\sim 10$ and $\sim 9$ Å for bidentate. The binding energies were calculated as:

$$E_b = E_{surf.} + E_{mol.} - E_{sys.}$$

Furthermore, to assure the order of stability of the different adsorption configuration, we performed test for i) the BZ sampling : $\Gamma$ point vs $[2 \times 2 \times 1]$ $k$-points grid mesh ; ii) the surface unit cell : $(1 \times 3)$ vs $(1 \times 4)$.

Table C.17: Binding energy of monodentate and bidentate configuration minima of FA adsorbed on anatase (101) surface.

| | (1×3) ; Γ | (1×3) ; [2×2×1] | (1×4) ; Γ |
|---|---|---|---|
| MH-intra | 0.819 | 0.813 | 0.818 |
| MH-inter | 0.814 | 0.809 | 0.812 |
| M(H) | 0.374 | 0.374 | |
| BB(H) | 0.666 | 0.691 | 0.561 |

**PIMD and classical Langevin simulations** The NVT PIMD and classical Langevin simulations were performed using i-Pi interface [194] to QE. In particular, the method Path Integral Generalized Langevin Equation Thermostat (PIGLET) [46, 47] was used, allowing a faster convergence on the number of beads in the PIMD simulation. The convergence tests on the number of beads have been performed on FA molecule (Figure C.2). The number of beads was set $n = 12$ at 100 K and $n = 8$ at 300 K. The friction coefficient $\gamma$ was set to 10 THz and the time-step $dt$ to 0.5 fs. The total length of PIMD simulations was 5 ps for the gas-phase and between 5-10 ps for the molecular adsorption. Classical Langevin simulations were conducted instead over $\sim$ 23 ps. For the formic acid dimer, we have perfomed a PIMD calculation at T = 100 K using 64 beads within a simulation time of 50 ps, using the fitted potential energy surface from ref. [190]. The trajectories were analyzed using the python library MDTraj [195]. Probability distributions were computed using Gaussian kernel density estimation from python SciPy package [196].

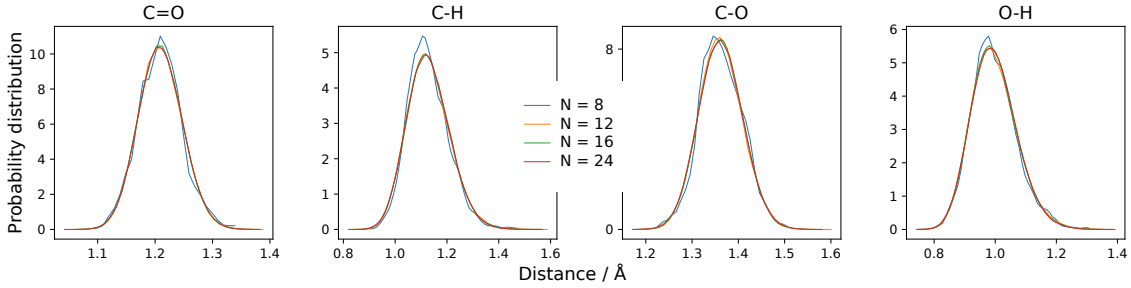



Figure C.2: Probability distribution at 100 K of bond lengths of FA molecule respect to the number of replicas N used in PIGLET calculations.

**Procedure for calculating QCT and SC power spectra** The QCT and DC-SCIVR power spectra for each adsorption configuration were computed according the following steps:

1. Hessian matrix at $R_{eq}$ using a finite difference scheme using the FD (PHonon QE) code;
2. Diagonalization of the hessian matrix to compute the harmonic normal modes;
3. NVE MD simulation (velocity-Verlet) with: i) initial positions $R_{eq}$; ii) momenta initialized to the harmonic estimate of ZPE; iii) time-step $dt \sim 0.24$ fs (10 Ry a.u.) and iv) total simulation length of 0.6 ps (2500 iterations);
4. Fast-Fourier Transform of velocity-velocity correlation function (QCT spectra);
5. Averaged hessian over 20 hessians along the trajectory $\rightarrow \langle H_{ij} \rangle$;
6. Division of the DOF in subspaces of reduced dimensionality via analysis of $\langle H_{ij} \rangle_{i \neq j}$;
7. Potential energy along the trajectory projected in the subspace of interest;
8. Hessian along the trajectory in the subspace of interest;
9. Semiclassical spectra (DC-SCIVR spectra).

**Subspaces subdivision for DC-SCIVR calculations** In the following the dimension of the selected subspaces are reported.

Table C.18: Selected subspaces for HCOOH ads. using a single trajectory initialized to a factor $\alpha$ of the ZPE harmonic energy ($E_{kin,0} = \alpha \times \mathrm{E_{ZPE,Harmonic}}$).

| | $\alpha = 1$ | | $\alpha = 3/4$ | | $\alpha = 1/2$ | |
|---|---|---|---|---|---|---|
| | $\epsilon$ ($10^{-6}$) | dim. | $\epsilon$ ($10^{-6}$) | dim. | $\epsilon$ ($10^{-6}$) | dim. |
| MH-intra | 2.0 | 17 | 1.5 | 18 | 1.5 | 13 |
| MH-inter | 1.2 | 15 | 1.6 | 13 | 1.5 | 13 |
| BB(H) [formate] | 5.0 | 6 | | | | |
| BB(H) [$O_s$H] | 15 | 3 | | | | |

Table C.19: Selected subspaces for HCOOD ads. using a single trajectory initialized to a factor $\alpha$ of the ZPE harmonic energy ($E_{kin,0} = \alpha \times \mathrm{E_{ZPE,Harmonic}}$).

| | $\alpha = 1$ | | $\alpha = 3/4$ | | $\alpha = 1/2$ | |
|---|---|---|---|---|---|---|
| | $\epsilon$ ($10^{-6}$) | dim. | $\epsilon$ ($10^{-6}$) | dim. | $\epsilon$ ($10^{-6}$) | dim. |
| MD-intra | 1.6 | 16 | 1.4 | 16 | 1.02 | 16 |
| MD-inter | 2.4 | 10 | 1.6 | 13 | 1.6 | 13 |
| BB(D) [formate] | 5.0 | 2,3 | | | 4.0 | 2,3 |
| BB(D) [$O_s$D] | 12.5 | 3 | | | 7.0 | 3 |

## C.2 Supplementary material

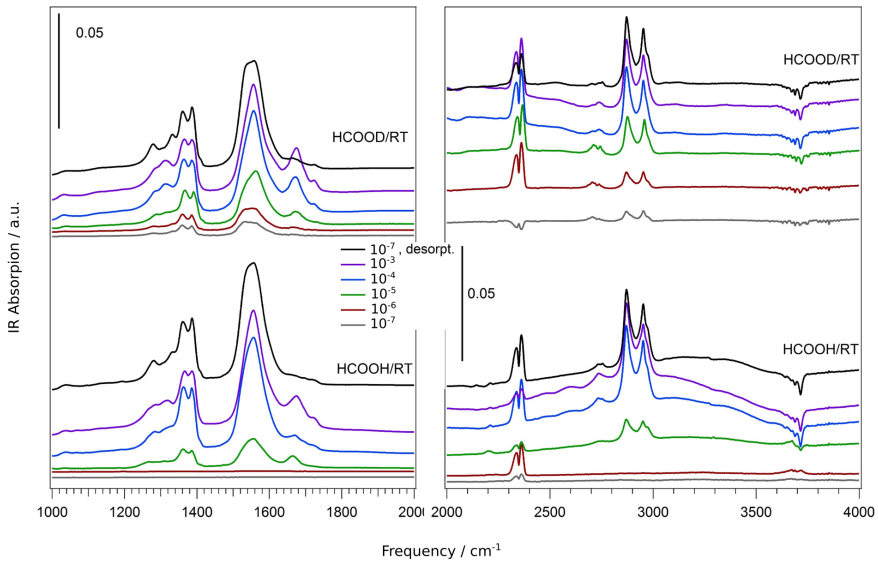



Figure C.3: FTIR spectra at room temperature of protonated FA *vs* deuterated FA adsorbed on anatase $TiO_2$ nanopowders exposed to $P_{FA}$ = $10^{-7}$–$10^{-3}$ mbar, compared to the spectra recorded after FA desorption at $P_{FA} = 10^{-7}$ mbar (black curve).

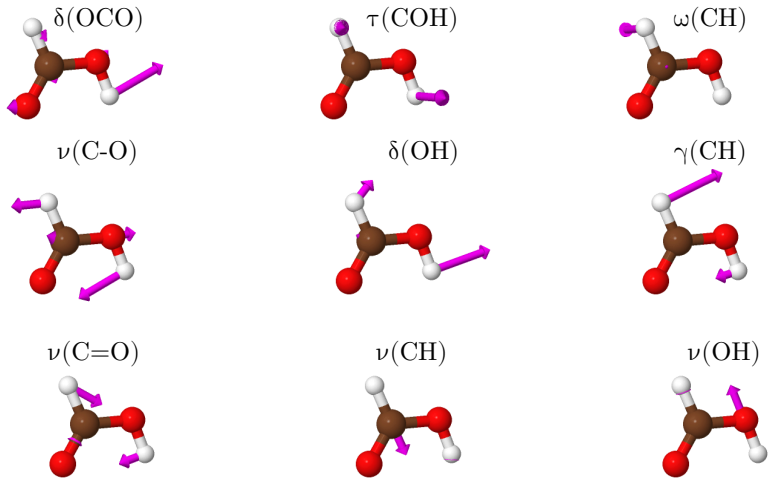



Figure C.4: The eigendisplacements of *trans* formic acid molecule.

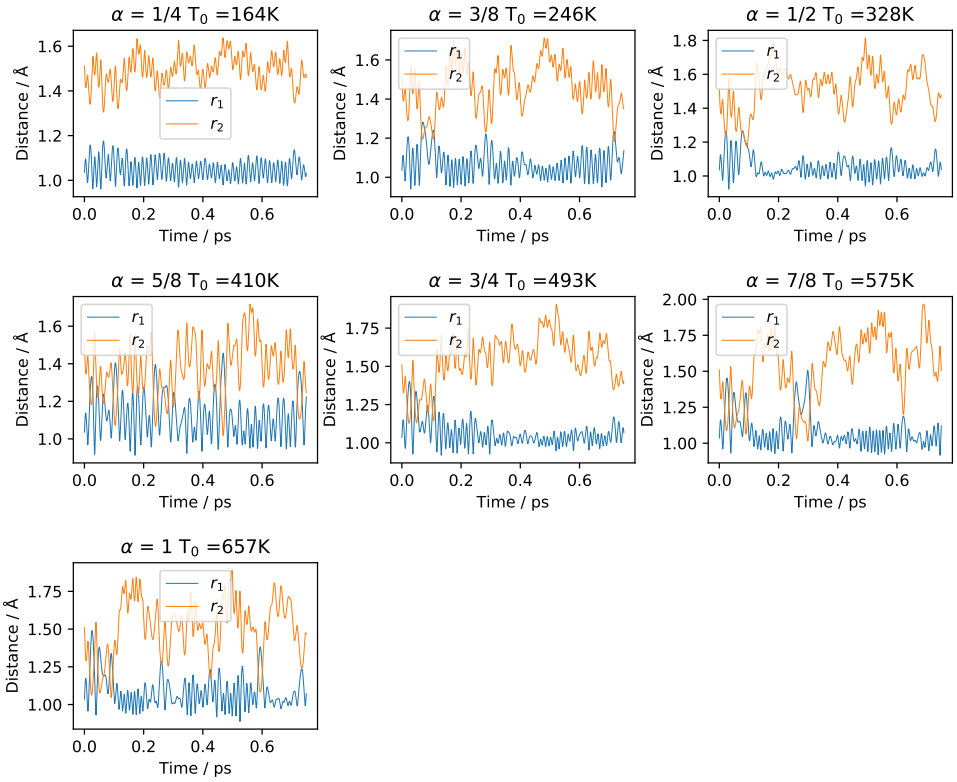



Figure C.5: Time evolution of $r_1$ = $O_f$H and $r_2$ = $O_s$H during the NVE dynamics of FA adsorbed in MH-intra configuration at different initial conditions given by $\alpha = [1/4, 1]$ ($E_{kin,0} = \alpha \times E_{ZPE,Harmonic}$).

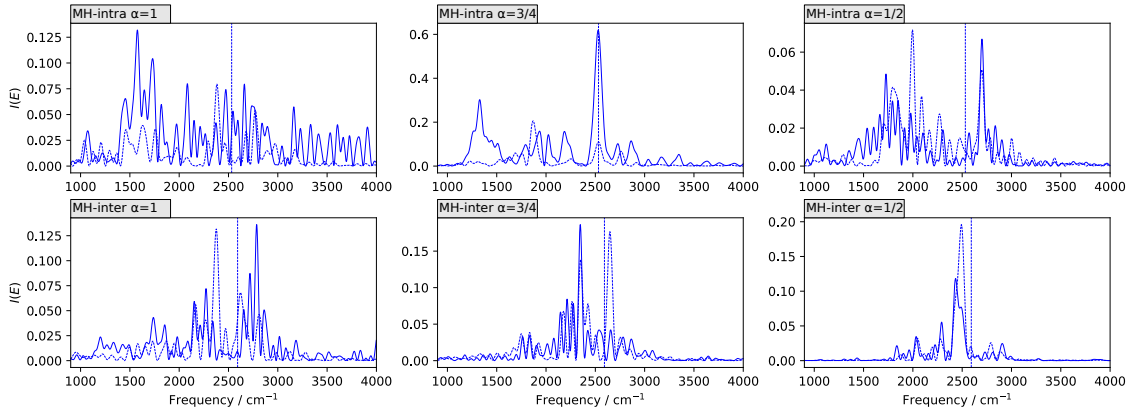



Figure C.6: SC and QCT spectra of OH stretching mode of FA adsorbed in MH-intra and MH-inter configurations at selected $\alpha$ (1, 3/4, 1/12).

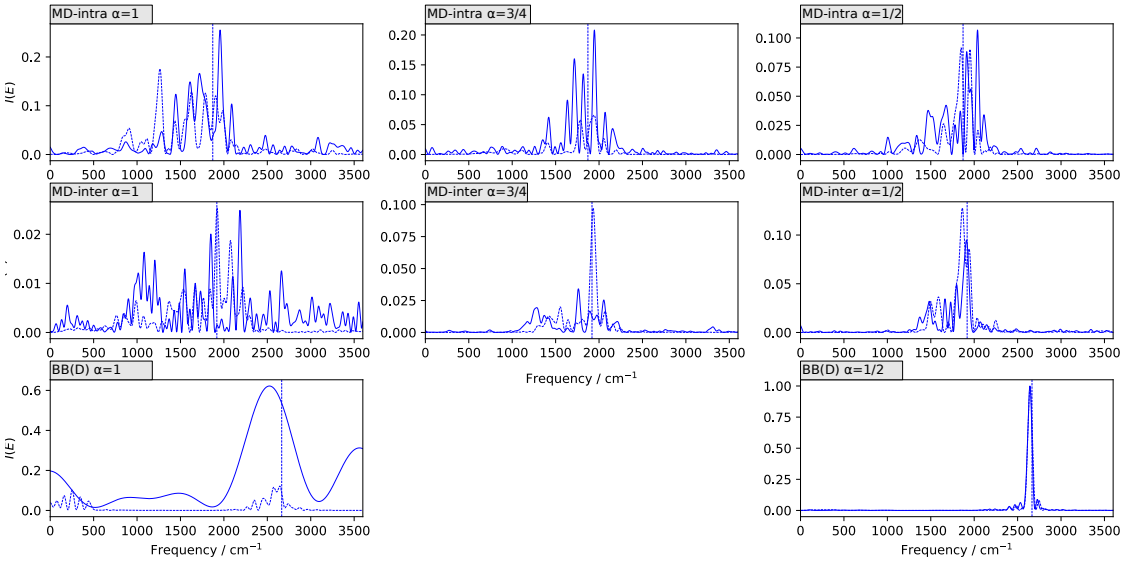



Figure C.7: SC and QCT spectra of OD stretching mode of FA adsorbed in MD-intra, MD-inter and BB(D) configurations at selected $\alpha$ (1, 3/4, 1/12).

CHAPTER 5

# A FIRST INVESTIGATION ON THE RELEVANCE OF QUANTUM AND ANHARMONIC EFFECTS IN GAS-PHASE AND SOLVATED GUANINE AND CYTOSINE BASE PAIR

## 5.1 Introduction

Hydrogen bonds play a crucial role in biomolecular systems. Here we consider the DNA molecule, the most important hydrogen-bonded system of all. The double helix of the DNA is a long polymer chain formed by three main components: the sugar ring deoxyribose, the phosphate groups and the purine (guanine (G), adenine (A)) and pyrimidine (cytosine (C), thymine (T)) bases. Two main factors contribute to the stability of the DNA: the base pairing between complementary bases (GC and AT) and the stacking interaction between adjacent bases. The base pairing essentially consists in the formation of three hydrogen bonds between guanine and cytosine and two between adenine and thymine, in their Watson and Crick (WC) conformation [197]. Despite being a very stable molecule, sometime the DNA can experience some damage, causing a modification of the original genetic material. The replication of damaged DNA may lead to gene mutations responsible for genetic disorders or illnesses. One of the possible sources of the DNA mutations is the base mispair. Already at the beginning of the elucidation of the DNA structure, Watson and Crick suggested that one of the sources of genetic mutations could be the tautomerisation of the base pairs [198]. Tautomerization is a form of isomerization, in which the protons of a molecular system are redistributed without changing the chemical formula (see, for example, fig. 5.1). Due to the presence of keto and amino groups, tautomerization is of relevant importance in nucleobases, for which an intense research field has been developed during the years [199].

In a study by Rejnek and Hobza, accurate calculations at MP2 level have shown that the WC-AT is much more stable than the noncanonical tautomers, while in GC the WC-GC, which is the

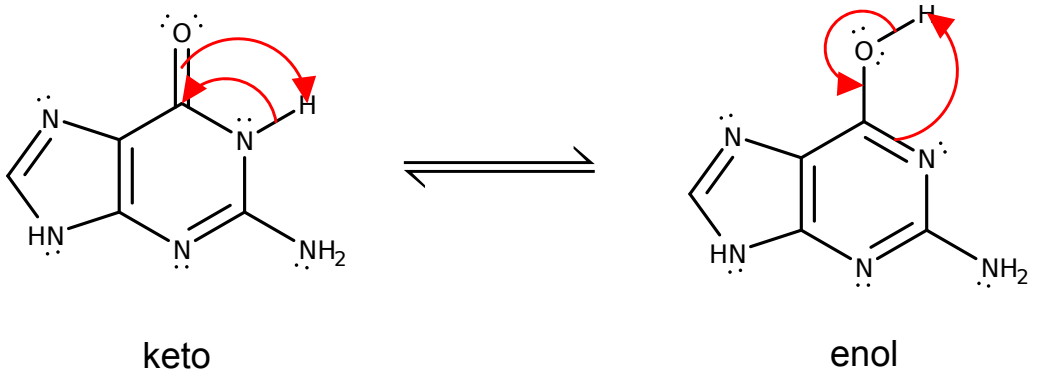



Figure 5.1: Sketch of keto-enol tautomerization in guanine nucleobase. Reproduction from reference [200].

most stable conformer, dominates but the population of some other noncanonical tautomers cannot be neglected, even when including a water environment via polarizable continuum model (PCM) [201]. Generally, the interconversion between tautomers can occur through a direct intramolecular proton transfer as shown in fig. 5.1 and/or assisted by a protic solvent. Another way in which tautomerism might occur in nucleobases is via a double proton trasfer (DPT) along the hydrogen bonds within the bases. This model was originally proposed by Löwdin [202, 203] as a possible source of the spontaneous mutation of DNA. The DPT transforms the canonical Watson and Crick forms AT and GC into rare tautomers A*T* and G*C* (see, as an example, the DPT mechanism in GC base pair fig. 5.2). Then, the A*T* and G*C* rare tautomers might eventually originate a base pair mismatch such as A*C, GT*, G*T and AC*.

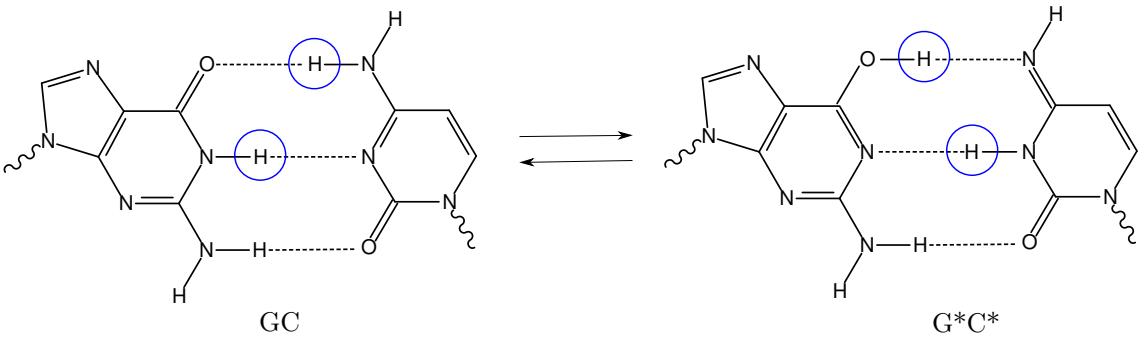



Figure 5.2: Sketch of the double proton transfer in GC Watson and Crick base pair. G*C* is a rare tautomer, where the Gua moiety is in enol form and the Cyt moiety is in imino form.

Furthermore, Löwdin hypothesized that the DPT might be associated with quantum tunneling due to the intrinsic quantum nature of the hydrogen atom. However, the DPT mechanism has not been evidenced yet under physiological conditions. Since the Löwdin hypothesis, several theoretical studies have been conducted [204]. Recent studies by Slocombe and coworkers proposed that a significant contribution to the DPT comes from the quantum tunnelling of the protons and calculated the tunneling probability in the G-C base pair [205, 206]. Soler et al. suggested in a QM/MM study with the inclusion of a realistic biomolecular environment that the G*C* tautomers are metastable in the gas phase and completely unstable in physiological conditions and dispel the DPT as source of spontaneous point mutations [207]. In support of that, a recent QM/MM study of aqueous DNA dodecamer [208] showed that the DPT reactions in GC have negligible contribution. The under-

standing of the equilibrium properties and dynamics of nucleobase pairs is, therefore, of crucial importance, but a very challenging problem due to the multiple factors that come into play, in particular the role of the solvent and the backbone and the tautomeric population.

In the present chapter we explore the equilibrium and vibrational properties of the GC base pair in the gas and condensed phases. Due to their computational complexity, accurate *ab initio* calculations are not flexible and fast enough to study the GC base pair in a realistic physiological environment and they are limited to the inclusion of the solvent water molecules in an implicit way as in the PCM approach. Thus, we employed a molecular mechanics approach to study the Watson and Crick conformation of GC using the AMOEBABIO18 force field [35] (see section 1.3.1), which has been recently applied to DC-SCIVR spectroscopy of isolated nucleosides [84] and solvated thymidine [209]. The accuracy of AMOEBABIO18 will be discussed in comparison with the available *ab initio* calculations performed by G. Botti.[1] Considering the exploratory nature of the present study, first we focus on the Watson and Crick conformer in the gas phase, where a careful comparison with the calculations by G. Botti is possible, and then we study the effect of the water solvent. We point out that this is a preliminary study, which is still work in progress.

## 5.2 Gas-phase Watson-Crick conformer

Our investigation started with the canonical Watson and Crick conformer of guanine and cytosine base pair using the polarizable force field AMOEBABIO18 [35] as implemented in the Tinker suite code [210]. The use of AMOEBABIO18 allowed us to investigate only the canonical Watson and Crick conformation, which is included in the force field parametrization. The exploration of other isomers such as relevant tautomers is only possible via a reparametrization of the force field parameters. For the same reason, the sugar moiety had to been included into the calculations. We have chosen the complex of deoxyguanosine and deoxycytosine (dG·dC) as illustrated in fig. 5.3. Since we are mostly interested in the GC interbase region, the complex dG·dC is a good reference of the isolated G·C base pair despite the steric effects caused by the sugar moiety. Our estimated binding energy of the dG · dC complex is 110.32 kJ/mol, which is in good agreement with the binding energy calculated at DFT/B3LYP level of theory in the gas phase, 109.66 kJ/mol, from reference [211].

The GC base pair presents three hydrogen bonds, of which one is formed between the keto C=O group of Gua and the amino $NH_2$ group of Cyt (N4H$\cdots$O6), one between the N1H of Gua and the N3 of Cyt (N1H$\cdots$N3) and one between the keto C=O group of Cyt and the amino $NH_2$ group of Gua (N2H$\cdots$O2) (see fig. 5.3). Table 5.1 reports the relevant hydrogen bond parameters on the optimized structure of the canonical dG·dC in the gas phase. The hydrogen bonds of the canonical structure in the gas phase are relatively long and weak, having a donor-acceptor distance $>$ 2.8-3 Å. The order of the hydrogen bond strength is N1H$\cdots$N3 $<$ N2H$\cdots$O2 $<$ N4H$\cdots$O6. To test the accuracy of the AMOEBABIO18 force field, we compare our results with DFT/B3LYP calculations performed on the same system [212]. The main discrepancy regarding the hydrogen bond lengths is associated with N4H$\cdots$O6 hydrogen bond, having a difference with the DFT reference over the distances AB (covalent) and HB (hydrogen bond) of 0.098 Å and 0.129 Å, respectively. It

[1] Theoretical collaborator: **Giacomo Botti** – Dipartimento di Chimica, Università degli Studi di Milano.

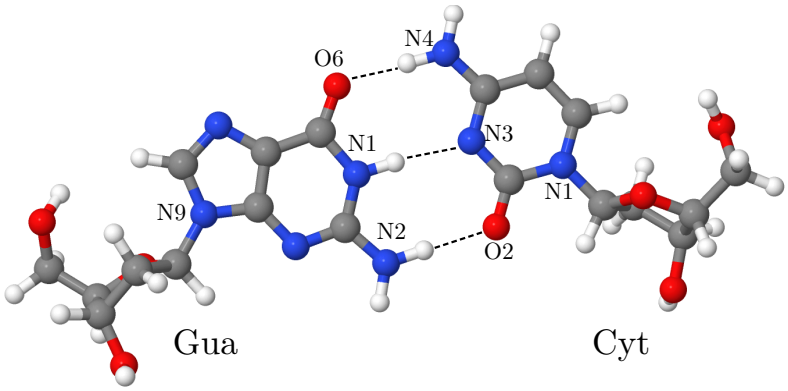



Figure 5.3: Representation of the canonical Watson-Crick deoxyguanosine and deoxycytosine nucleoside pair (dG·dC) employed in the current AMOEBABIO18 calculations. The oxygen, nitrogen, carbon and hydrogen atoms are colored red, blue, grey and white respectively.

is important to notice that AMOEBABIO18 was designed for the condensed phase, therefore the deviations from the DFT reference are reasonable. The hydrogen bond angles ∠AHB, which are very close to linearity, are well reproduced within the AMOEBABIO18 force field.

Table 5.1: GC interbasebase hydrogen bond and covalent bond distances (Å) and angles (°) in gas-phase canonical Watson-Crick dG·dC nucleoside pair using AMOEBABIO18 force field. The shifts of AMOEBABIO18 values with respect the ones from DFT/B3LYP, $\Delta_{\mathrm{AMOEBA/DFT}}$, are reported in parenthesis.

| | H-bond AH···B | $d_{AB}$ | $d_{HB}$ | $d_{AH}$ | ∠AHB |
|---|---|---|---|---|---|
| AMOEBABIO18 | N4H···O6 | 2.890 (0.098) | 1.883 (0.129) | 1.011 | 176.6 (-2.5) |
| (dG·dC) | N1H···N3 | 3.029 (0.091) | 2.018 (0.113) | 1.011 | 178.5 (0.7) |
| | N2H···O2 | 2.938 (0.028) | 1.927 (0.039) | 1.008 | 179.1 (0.6) |
| DFT/B3LYP[a] | N4H···O6 | 2.792 | 1.754 | | 179.1 |
| (dG·dC) | N1H···N3 | 2.938 | 1.905 | | 177.8 |
| | N2H···O2 | 2.910 | 1.888 | | 178.5 |

[a] From ref. [212] with 6-31G(d,p) basis set.

### 5.2.1 Thermal and quantum effects on structural properties

As a first investigation of the role of nuclear quantum effects on the structure of the GC base pair, we have studied the equilibrium properties of the dG·dC complex at 300 K. We have used the adQTB approach, PIMD simulations, as well as classical Langevin molecular dynamics simulations. Figure 5.4 shows the radial distribution function of NH and OH pairs in the GC interbase region calculated with the three different methods in the gas phase. The peak at 1 Å is related to the covalent bonds and the one between 1.8-2.2 Å to the hydrogen bonds. The $g_{\mathrm{NH}}(r)$ is the only one presenting a peak in the covalent bond region since there are no OH covalent bond in the interbase region. We recall that the use of conventional force fields does not allow any proton transfer, so the conformer is well confined in the canonical structure along the whole dynamics. We firstly point

out that the adQTB distribution is in good accordance with the PIMD one, meaning that the adQTB can describe faithfully the equilibrium properties of the dG·dC nucleoside pair. In all cases, the quantum distributions are wider than the classical ones due to the significant zero-point energy of the N-H covalent bond. Concerning the hydrogen bonds, there are no strong effects, there is no significant shift of the hydrogen bond distances.

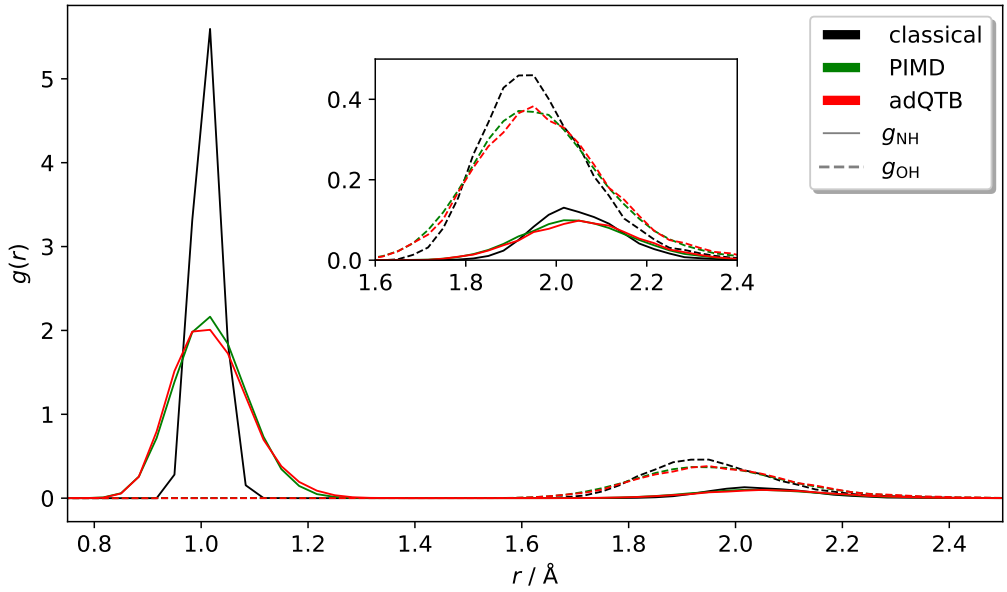



Figure 5.4: Radial distribution function $g(r)$ of NH and OH interbase pairs of gas-phase dG·dC at 300 K from adQTB, PIMD and classical Langevin simulations using AMOEBABIO18.

### 5.2.2 NH stretching vibrations

We performed a vibrational spectroscopy investigation to obtain more insight about the dynamical properties of dG·dC. The calculation of the vibrational properties of the nucleobase pairs is relevant to identify the role of the vibrational modes that might be involved in the proton transfer reactions and the factors that come into play in the stability of the double helix of DNA. Furthermore, vibrational spectroscopy is fundamental to identify the presence of tautomers in the experiments (see in section D.2 the description of tautomerism in the ANI model). In particular, we have adopted the QCT approach using a single trajectory initialized to the harmonic estimate of the ZPE in the isolated gas-phase dG·dC and we have also calculated the vibrational spectra using the adQTB method. The results are shown in table 5.2 and fig. 5.5.

To our knowledge, there is currently no experimental IR spectrum of the Watson-Crick conformation of dG·dC neither in the gas phase or in solution. Instead, some experimental vibrational studies have been conducted on G·C clusters [213–217]. However, in the available gas phase spectroscopic data the attribution to the canonical Watson-Crick GC was dubious due to the fact that it is not clear if the isolated species is either in the canonical or in a tautomeric form. Therefore, we restrict our discussion to the comparison between the different methods and accurate *ab initio* calculations performed on the equivalent G·C conformation as a benchmark of the force field accuracy. It has been shown by Bende et al. that anharmonic corrections on the GC base pair are significant, in particular in the NH and CH stretching and intermolecular normal modes [218]. To

Table 5.2: GC interbase NH stretching vibrational frequencies in gas-phase dG·dC using AMOEBABIO18 at the harmonic, QCT and adQTB levels. The shifts of AMOEBABIO18 values with respect the ones from DFT-D/B3LYP, $\Delta_{\mathrm{AMOEBA/DFT}}$, are reported in round brackets.

| | $NH_2^a$ | | $NH_2^s$ | | NH |
|---|---|---|---|---|---|
| | Gua | Cyt | Gua | Cyt | Gua |
| **AMOEBABIO18 (dG·dC)** | | | | | |
| Harmonic | 3718 (2) | 3727 (35) | 3577 (188) | 3621 | 3610 |
| QCT | 3553 (-42) | 3495 (-90) | 3448 (74) | 3456 | 3451 |
| adQTB | 3539 | 3558 | 3447 | 3465 | 3465 |
| **DFT-D/B3LYP[a] (G·C)** | | | | | |
| Harmonic | 3716 | 3692 | 3389 | | |
| QCT | 3595 | 3585 | 3374 | | |

[a] With basis set 6-31G**. The calculation was performed by G. Botti.

characterize the vibrational properties of the canonical Watson-Crick conformation of dG·dC, we focused on the high-frequency region in the 3000-4000 $cm^{-1}$ range, which is related to the NH stretching vibrations of the GC interbase region.

Concerning the force field accuracy, we compare the harmonic and QCT results using AMOEBABIO18 with DFT-D/B3LYP data, by considering only the normal modes that are equivalent in the two different computational approaches. Firstly, we point out that the largest difference between the force field and the DFT reference is at the harmonic level over the $NH_2$ symmetric stretching ($NH_2^s$) of Gua, ($\Delta_{\mathrm{AMOEBA/DFT}} = 188$ $cm^{-1}$). This shift is reduced to 74 $cm^{-1}$ when including anharmonicity at the QCT level, but the agreement with the DFT-D/B3LYP data worsens for the other two frequencies, the $NH_2$ asymmetric stretching ($NH_2^a$) of Gua and Cyt, which present a $\Delta_{\mathrm{AMOEBA/DFT}}$ of -42 and -90 $cm^{-1}$, respectively.

The adQTB spectra were calculated for each atom type[2] directly during the adQTB simulation and were deconvoluted in order to eliminate the broadening associated with the Langevin friction force and to obtain sharper peaks. The spectrum of atom type H2, corresponding to the two hydrogen atoms of the $NH_2$ group of Gua, is resolved in two peaks, the symmetric and antisymmetric stretching of $NH_2$ of Gua. Similarly, the spectrum of atom type H4 presents two peaks, which are the symmetric and antisymmetric stretching of $NH_2$ of Cyt. Lastly, the atom type H1, which corresponds to the hydrogen atom of the NH group in Gua, presents a single peak. The adQTB frequencies are very similar to their QCT counterparts in the gas phase, notwithstanding a lower resolution of the former ones. The only exception is $NH_2^s$ of Cyt, which is at a significantly higher frequency compared to the QCT result (3495 $cm^{-1}$ in QCT *vs* 3558 $cm^{-1}$ in adQTB). The similarity between the adQTB and QCT spectra are due to the fact that both methods employ a classical trajectory which includes the zero-point energy. Their relatively good agreement is expected.

[2] We recall the definition of atom types in a force field. Each atom of the system is associated to a specific atom type, which is related to its chemical environment.

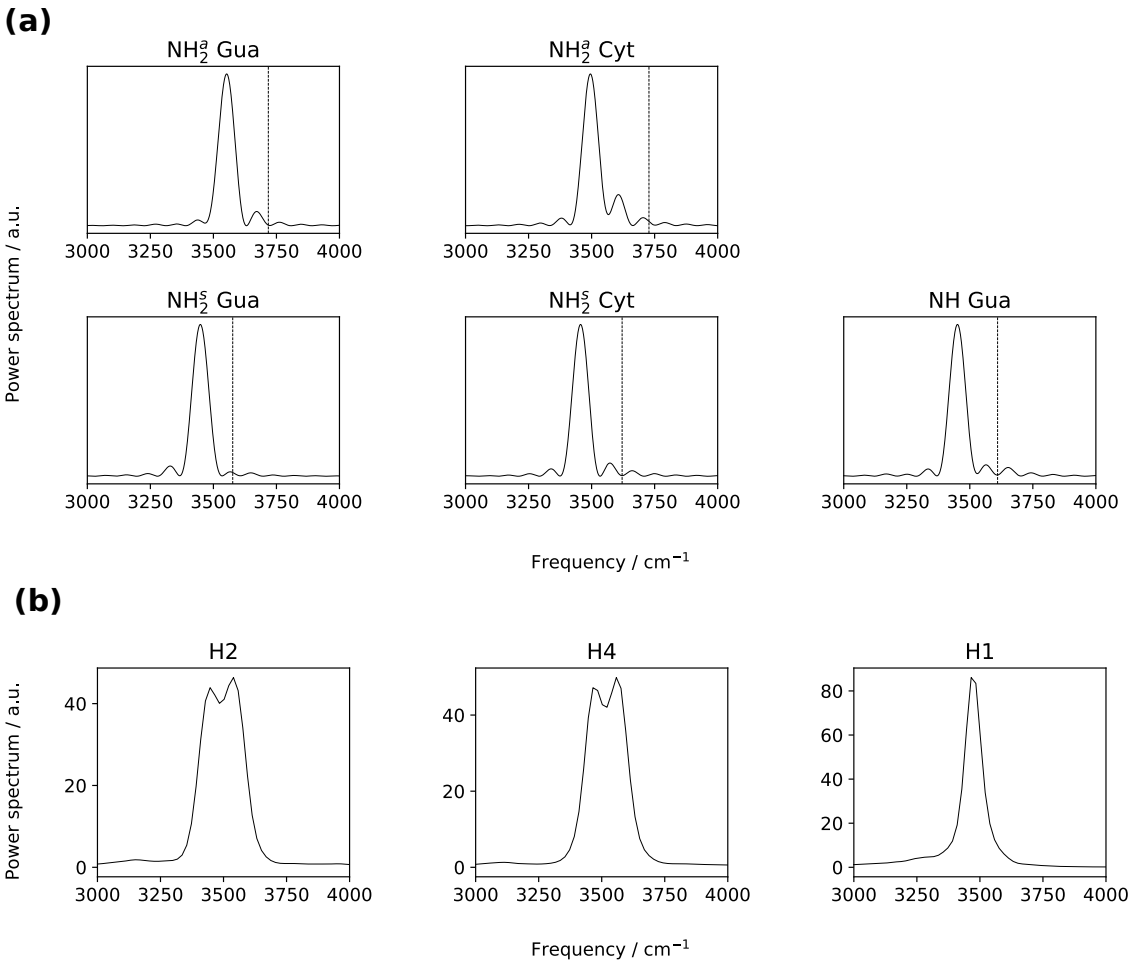



Figure 5.5: Interbase NH stretching vibrational spectra of gas-phase dG·dC using AMOEBABIO18. (a) QCT spectra. The corresponding harmonic values are reported as vertical dashed lines. (b) adQTB spectra. For the atomic sites, see numbering in fig. 5.3.

## 5.3 Biomolecular environment: preliminary results

### 5.3.1 Solvent effect: water-box

As previously stated, the advantage of employing a force field over *ab initio* calculations lies in the possibility of including the solvent explicitly, without having to use implicit and dubious solvent models. Thus, to study the solvent effect, we have placed the dG·dC nucleoside pair in a cubic box of water molecules of length 30 Å as shown in fig. 5.6. We have used periodic boundary conditions and the particle mesh Ewald (PME) scheme for the long-range interactions [219, 220]. The resulting total number of atoms is 2644 including both the dG·dC and the water molecules.

The geometry obtained after minimization of the solvated dG·dC complex is different with respect to the gas phase (table 5.3). In the gas phase, the Gua and Cyt moieties are approximately contained in the same plane. Instead, after the solvatation, the planarity is lost with one of the nucleoside being staggered from the other one. This significantly affects the interbase hydrogen bonds, which largely deviate from linearity (see ∠AHB in table 5.1). We notice that the solvent effect is to increase the $\mathrm{d_{AB}}$ and $\mathrm{d_{HB}}$ distances in the most external interbase hydrogen bonds N4H$\cdots$O6 and N2H$\cdots$O2, while it decreases the $\mathrm{d_{AB}}$ and $\mathrm{d_{HB}}$ distances in the internal interbase hydrogen bond, N1H$\cdots$N3.

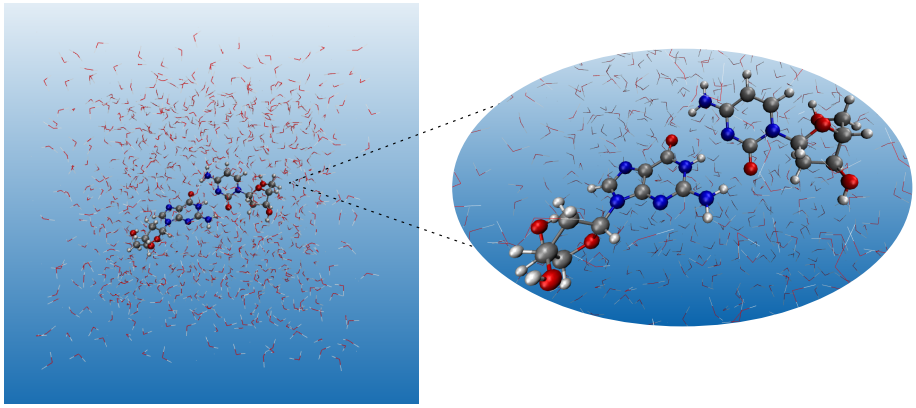


Figure 5.6: Representation of the Watson-Crick dG·dC nucleoside pair in a cubic box of water molecules of length 30 Å.

Table 5.3: GC interbasebase hydrogen bond and covalent bond distances (Å) and angles (°) in solvated-phase with respect gas-phase canonical Watson-Crick dG·dC nucleoside pair using AMOEBABIO18 force field.

| H-bond AH···B | $d_{AB}$ | $d_{HB}$ | $d_{AH}$ | ∠AHB |
|---|---|---|---|---|
| *Gas-phase* | | | | |
| N4H···O6 | 2.890 | 1.883 | 1.011 | 176.6 |
| N1H···N3 | 3.029 | 2.018 | 1.011 | 178.5 |
| N2H···O2 | 2.938 | 1.927 | 1.008 | 179.1 |
| *Water-box* | | | | |
| N4H···O6 | 3.133 | 2.147 | 1.008 | 165.7 |
| N1H···N3 | 2.963 | 1.989 | 1.013 | 160.4 |
| N2H···O2 | 2.963 | 1.960 | 1.010 | 171.6 |

This can also be read as a decrease of the hydrogen bonds strength in the interbase area, which is in contact with the surrounding water molecules, while instead the internal hydrogen bond is strengthened.

Table 5.4: GC interbase NH stretching vibrational frequencies in solvated-phase with respect gas-phase canonical Watson-Crick dG·dC nucleoside pair using AMOEBABIO18 at the three levels of theory from harmonic and QCT simulations. $\Delta_{gas/condensed}$, are reported in square brackets.

| | $NH_2^a$ | | $NH_2^s$ | | NH |
|---|---|---|---|---|---|
| | Gua | Cyt | Gua | Cyt | Gua |
| *Gas-phase* | | | | | |
| Harmonic | 3718 | 3727 | 3577 | 3621 | 3610 |
| QCT | 3553 | 3495 | 3448 | 3456 | 3451 |
| *Water-box* | | | | | |
| Harmonic | 3719 [1] | 3691 [-36] | 3592 [-15] | 3579 [-42] | 3609 [-1] |
| QCT | 3604 [51] | 3605 [110] | 3484 [36] | 3466 [10] | 3513 [62] |

In order to point out the differences in the vibrational properties that are due to the solvent, we

have calculated the harmonic and QCT vibrational spectra in the solvated dG·dC complex (table 5.4,fig. 5.7). In our simulations, we observe that NH stretching vibrations of dG·dC are very sensitive to the environment, both at the harmonic and QCT levels. The harmonic frequencies are all red-shifted in the condensed phase, with the exception of $NH_2^a$ and NH of Gua, which present almost the same frequency with respect to the gas phase. In contrast, in the QCT approach all the investigated modes are blue-shifted with respect to the gas phase (see table 5.4 and fig. 5.7). The largest contribution is over $NH_2^s$ of Cyt, which is blue-shifted of 110 $cm^{-1}$. As a general rule, one should expect that the effect of solvation in a polar medium is to red-shift the frequencies. Instead, the QCT spectra of dG·dC in the water-box reveal an "improper" blue shift for all the selected modes, a very peculiar property, whose origins we have not investigated yet but are probably associated with the effect of the surrounding water on the interbase hydrogen bonds. Usually, weak hydrogen bonds yield higher stretching frequencies when including anharmonicity due to the strengthening of the AH covalent bond.

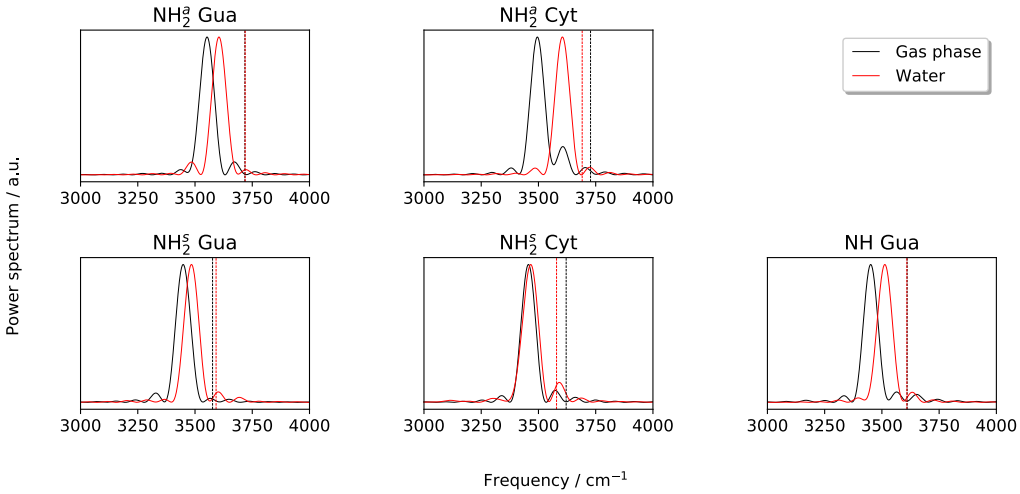



Figure 5.7: QCT spectra of interbase NH stretching vibrations in gas phase and solvated dG·dC using AMOEBABIO18. The corresponding harmonic values are reported as vertical dashed lines.

Concerning the solvated dG·dC nucleoside pair, we could not evaluate any vibrational and thermodynamic properties from adQTB and PIMD simulations at 300 K due to the instability of the complex in solution. Indeed, in less than 1 ns of simulation, the Gua and Cyt moieties separate (see fig. D.1 in appendix). This is due to the absence of the DNA backbone, which ensures the stability of the nucleoside pairs in solution on longer time scales.

### 5.3.2 Inclusion of DNA backbone: solvated dodecamer

The presence of the DNA-like backbone clearly appears to be essential in the study of solvated nucleobases. To model a solvated DNA-like system and study the effects on the hydrogen bonds equilibrium and dynamics of dG·dC base pair, we have adapted a new model, a DNA double helix dodecamer of sequence d(TTAGGGTTAGGG). The latter was neutralized with 22 $Na^+$ counterions and solvated in a rectangular water-box of dimensions 70 Å × 45 Å × 45 Å. We used periodic boundary conditions and the PME scheme. The relevant hydrogen bonds parameters for the optimized structure are reported in table 5.5. Also in this system, the G·C base pair presents with

respect to the gas-phase a deviation from planarity, particularly evidenced by the ∠AHB for the N4H$\cdots$O6 hydrogen bond, which is 155.8°. We note that N1H$\cdots$N3 and N2H$\cdots$O2 hydrogen bonds in the solvated dodecamer are longer and shorter than the ones in solvated dG·dC, respectively. It is well to specify, the role of the counterions and the periodic box lengths has to be further investigated and their influence on the geometry of the dodecamer.

Table 5.5: GC interbasebase hydrogen bond and covalent bond distances (Å) and angles (°) in the solvated dodecamer with respect to the gas-phase canonical Watson-Crick dG·dC nucleoside pair using AMOEBABIO18 force field.

| H-bond AH$\cdots$B | $d_{AB}$ | $d_{HB}$ | $d_{AH}$ | ∠AHB |
|---|---|---|---|---|
| *Gas-phase* | | | | |
| N4H$\cdots$O6 | 2.890 | 1.883 | 1.011 | 176.6 |
| N1H$\cdots$N3 | 3.029 | 2.018 | 1.011 | 178.5 |
| N2H$\cdots$O2 | 2.938 | 1.927 | 1.008 | 179.1 |
| *Solvated-dodecamer* | | | | |
| N4H$\cdots$O6 | 3.118 | 2.172 | 1.010 | 155.8 |
| N1H$\cdots$N3 | 3.033 | 2.025 | 1.013 | 172.9 |
| N2H$\cdots$O2 | 2.812 | 1.809 | 1.009 | 172.0 |

To investigate both the structural and vibrational properties of the solvated dodecamer, we have performed adQTB and classical Langevin simulations at 300 K. In contrast to the water-box simulations, where the Gua and Cyt moiety were splitting apart from each other, in the adQTB simulation at 300 K of the solvated dodecamer along a trajectory length of 12 ns, the double helix is maintained intact. We calculated from this simulation the radial distribution function for AH pairs, as defined in table 5.5 and compared with the results from the gas phase. The differences between the $g_{\mathrm{AH}}(r)$ obtained from classical and adQTB simulations are significant in both the gas-phase and the solvated dodecamer, as evidenced by the broadening of the distributions with the inclusion of NQEs (fig. 5.9). This is mainly due to the larger delocalization of the proton due to the zero-point energy motion in adQTB simulations. The $g_{\mathrm{AH}}(r)$ are very similar for the two phases, suggesting that there is no significant difference in the AH covalent bonds. The spectroscopic features in the solvated dodecamer are similar to the ones in the gas phase (fig. 5.8), consistently with the radial distribution functions. This result is in strong disagreement with the results obtained for dG·dC in the water-box simulated with the QCT approach. In order to understand if the origin of this discrepancy is due to the backbone effect or to the lower accuracy and resolution of the adQTB method, we plan to apply the QCT method to a DNA-like system with more than a couple of bases.

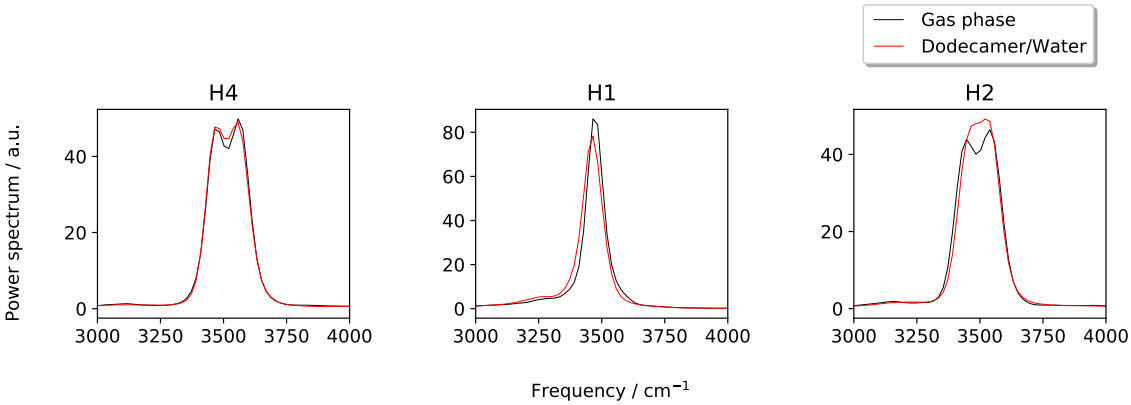



Figure 5.8: adQTB spectra of hydrogen atom types H2, H4, H1 in gas phase dG·dC using AMOEBABIO18. For the atomic sites, see numbering in fig. 5.3.

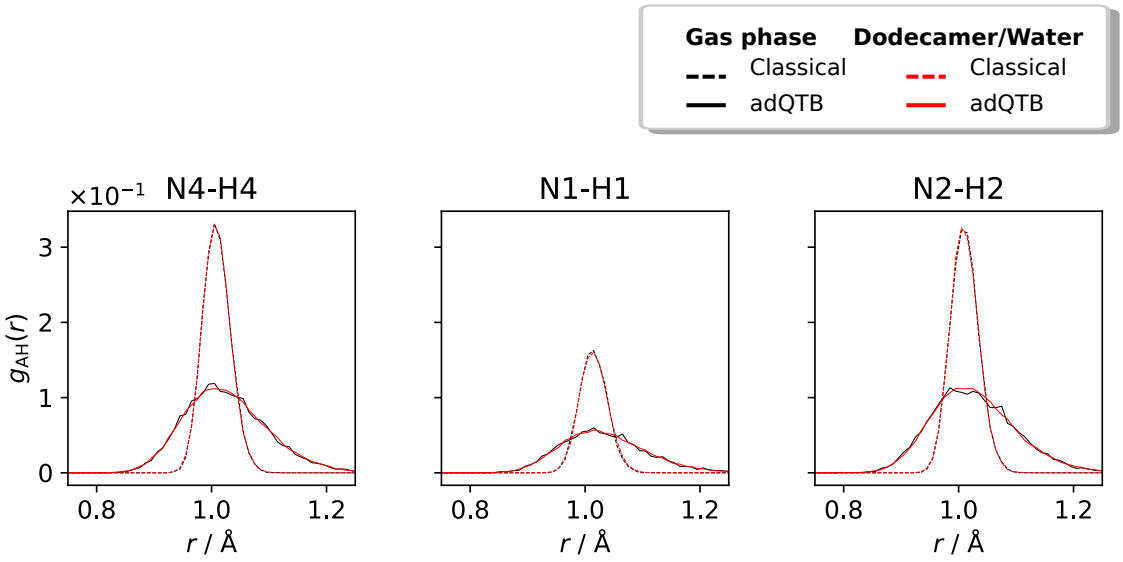



Figure 5.9: Radial distribution function $g(r)$ at 300 K of NH covalent bonds in the GC interbase region of dG·dC in the gas-phase and solvated dodecamer from adQTB simulations using AMOEBABIO18.

## 5.4 Conclusion

In this chapter, we have explored the equilibrium and vibrational properties of guanine and cytosine base pair, using different complementary methods. As a DNA building blocks, extensive theoretical studies on nucleobase pairs have been performed to understand multiple aspects of their behavior. To explain single point mutations in DNA, one proposed mechanism is the double proton transfer along the interbase hydrogen bonds, acting as a source of mispair of the bases through the formation of rare tautomers. The dynamics of nucleobase pairs is strongly affected by the environment in which they are simulated. To properly describe the physiological conditions, one fundamental step is the solvatation. This remains a very challenging problem since accurate *ab initio* calculations are impossible to obtain.

Via the AMOEBABIO18, we have investigated the canonical Watson and Crick conformation of dG·dC. Our results via adQTB suggest that the proton delocalization due to the zero-point energy motion is relevant even at 300 K both in the gas phase and in solvated DNA dodecamer, which includes both the solvent and the backbone effect. Concerning the vibrational properties, we have shown that the solvent effect is not straightforward to anticipate. The anharmonic spectra of solvated dG·dC obtained through QCT present an improper blue-shift of all the high-frequency modes related to the GC interbase region with respect to the gas-phase spectra. This contrasts with the harmonic approximation which provides mostly a red-shift. Within adQTB, which is less accurate for the study of vibrations, the high-frequency GC interbase spectroscopic features of solvated DNA dodecamer are similar to the ones in the gas phase. The more accurate QCT method will be applied to a DNA-like system in the near future.

# APPENDIX D

## D.1 Computational details

**Harmonic and QCT** We employed AMOEBABIO18 [35] force field as implemented in Tinker version 8.10.1 [221]. The optimized geometries where obtained using a RMS gradient of 0.01 kcal/mol/Å. The parameters in the Tinker input key file are indicated in the following.

For the gas phase dG·dC we used the following key file.

```
parameters amoebabio18
digits 8
verbose
remove-inertia 1
save-velocity
integrator beeman
save-force
```

For the condensed phase we build a cubic box of water molecules of length 30 Å and inserted the dG·dC inside, using the following key file.

```
parameters amoebabio18
digits 8
verbose
a-axis                  30.000
ewald
vdw-cutoff              12.0
ewald-cutoff            7.0
polar-eps               0.00001
polar-predict
save-velocity
integrator beeman
```

```
save-force
neighbor-list
```

The QCT spectra were computed using the following procedure.

1. Geometry optimization.
2. Hessian diagonalization at the equilibrium geometry and harmonic frequencies.
3. Preparation of the initial velocities of the dynamics.
4. Short time NVE dynamics using Beeman integrator of total length 0.6 ps with $dt$ = 0.2 fs.
5. Calculation of the QCT spectra with FT transform of classical $C_{vv}$.

**adQTB and PIMD** The adQTB and PIMD simulations were performed using the Tinker-HP software [222] on GPUs. The adQTB simulations at 300 K were long 12 ns, of which we discarded the first 6 ns as time of adaptation for the $\gamma_r$ of each atom species. We used the multi-time step integrator BAOAB-RESPA1 using a $dt = 2$ fs and a $dt_{\text{short}} = 0.00025$ ps. The adaptation velocity was set to coefficient $A_\gamma = 10^{-2}$ ps$^{-1}$. For the adQTB simulation of the gas-phase, we used the following input key file.

```
parameters amoebabio18
digits 8
verbose
a-axis 40.0
ewald
ewald-cutoff 13.0
vdw-cutoff 9.0
thermostat adqtb
integrator baoabrespa
dshort 0.00025
archive
ir_spectra
startsavespec 1000
friction 20
a_gamma 0.01
tseg 0.2
skipseg 5
omegacut 2000
spectra_cm1
REGISTER_SPECTRA
GAMMA_HISTORY
corr_pot
run-mode legacy
```

For the adQTB simulation at 300 K of the solvated dodecamer, we build a rectangular water box of dimension 70 Å $\times$ 45 Å $\times$ 45 Å and used the following input key file.

```
parameters amoebabio18
digits 8
verbose
a-axis 70
b-axis 45
c-axis 45
ewald
vdw-cutoff              12.0
ewald-cutoff            7.0
polar-eps               0.00001
polar-predict
thermostat adqtb
integrator baoabrespa
dshort 0.00025
archive
ir_spectra
startsavespec 1000
friction 20
a_gamma 0.01
tseg 0.2
skipseg 5
omegacut 2000
spectra_cm1
REGISTER_SPECTRA
GAMMA_HISTORY
corr_pot
run-mode legacy
```

The PIMD simulation at 300 K for the gas-phase dG·dC was long 4 ns using the following input key file.

```
parameters amoebabio18
digits 8
verbose
a-axis 40.0
ewald
ewald-cutoff 13.0
vdw-cutoff 9.0
integrator baoabrespa
dshort 0.00025
nbeads 16
archive
ir_spectra
friction 20
```

```
tseg 0.2
omegacut 2000
centroid_longrange
polar_centroid
run-mode legacy
```

## D.2 Supplementary material

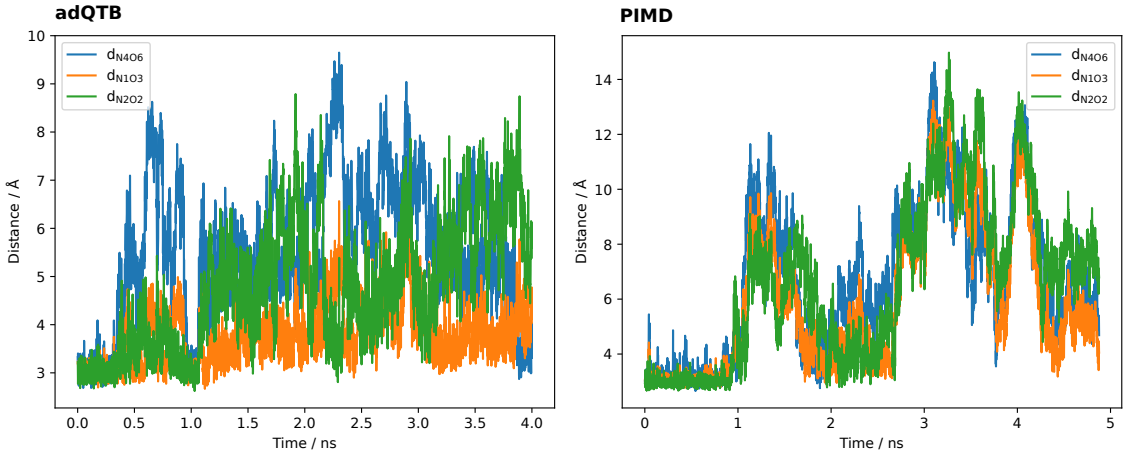


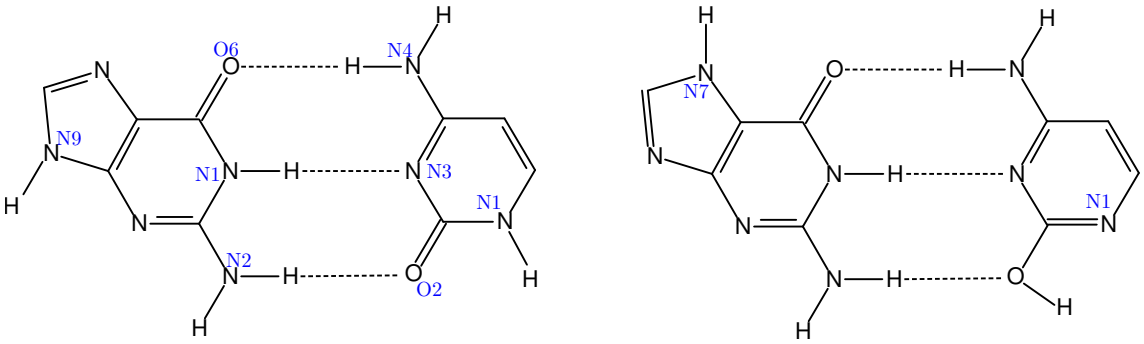

Figure D.1: Time evolution of the hydrogen bonds donor-acceptor distances with atom-numbering defined in fig. 5.6 for the adQTB and PIMD simulations at 300 K of dG·dC in a cubic box of water molecules of length 30 Å.

### Description of tautomerism in the ANI model

The tautomerism is of crucial importance in nucleobases and can affect both the equilibrium and the dynamics of the base pairs. To simplify the discussion, we now introduce the notation used by Nir et al. [215] to distinguish between the different tautomers of GC. In this notation 'K' stands for keto group, and 'E' for enol group. The labels '9' and '7' indicate where the substituent is attached to the Gua moiety, respectively the N9 and the N7 position. Lastly the '-1' indicates a WC type conformation. The K9K-1, for example, corresponds to the canonical WC conformer, while the K7E-1 to a WC form, in which the H substituent is attached to the N7 position of the Gua and the Cyt is enolized (see fig. D.2).

Nir et al. made a tentative assignment of the observed spectrum on the basis of harmonic calculations at the Hartree-Fock (HF) level performed on several isomers GC [215, 216] in the gas phase. They identified as the best candidates two isomers, in which the Cyt moiety is in enolic form, namely K9E-1 and K7E-1. The K7E-1 was later proposed to be the most compatible conformer [213, 214]. However, in all the calculations anharmonic effects were not included. Furthermore, it is not clear why the canonical WC form, K9K-1, which is expected to be the most stable conformation [215] is not appearing in the experimental spectrum. To elucidate the role of tautomerism in the gas phase GC base pair we have investigated the two isomer K9K-1 and K7E-1. The AMOEBABIO18 force field cannot provide a way to study different conformations besides the canonical

WC and to remove the sugar moiety, which is essential in order to investigate the tautomer K7E-1. Therefore, we have decided to use a different approach, the neural network ANI model. We have used the ANI family potentials as implemented in the Torchani [223] python interface integrated with the Atomic Simulation Environment [224] package. The classical NVE simulation was propagated using a velocity Verlet integration.




K9K-1 K7E-1

Figure D.2: Sketch of G·C isomers K9K-1 and K7E-1, as per notation by Nir et al. from reference [215].

**The ANI model** In the recent years a promising machine learning (ML) interatomic potential has been developed by the Roitberg group, namely ANI. ANI uses a modified version of the original Behler-Parrinello model for symmetry functions [225, 226] to construct single-atom atomic environment vectors. The first ANI potential, ANI-1x [227, 228], which was trained on about 20 million conformations derived from 57 thousand different molecular configurations computed with DFT, was developed with the goal of obtain a transferable ML potential that could accurately describe organic systems containing H, C, N, and O chemical elements. The ANI-1ccx potential [228] was built on a intelligently selected 10% sub-sample of the ANI-1x data set, but recomputed with an accurate coupled cluster (approximately CCSD(T)/CBS) level of theory [229]. A recent version of ANI, ANI-2x [230] includes three new chemical elements (S, F, and Cl).

We performed firstly a geometry minimization of the canonical WC, K9K-1 structure, and the tautomer K7E-1. The K9K-1 is predicted to be more stable than the K7E-1 by 0.45 eV in the ANI-2x model. Table D.6 illustrates the hydrogen bond parameters for both the isomers using the ANI-2x model and compared with DFT-D/B3LYP. As a general trend, we highlight that the ANI-2x potential gives WC GC base pair isomers in a distorted geometry. This is evidenced by the hydrogen bond angles ∠AHB, which present a deviation of 11-17° with respect the reference values. The hydrogen bonds distances are in good accordance with DFT-D/B3LYP, except for the N4H$\cdots$O6 hydrogen bond, which is overestimated by 0.212 Å and 0.235 Å in K9K-1 and K7E-1, respectively.

Similarly as we presented in section 5.3, we investigate here the high-frequency stretching vibrations related to the GC interbase region and the NH substituent groups in position N9 of Gua, N1 of Cyt (K9K-1) and N7 of Gua (KE-1). In particular, the presence of the enol group in K7E-1 with respect to the K9K-1 adds the OH stretch of the Cyt moiety and removes the N1H of Cyt stretching. We have calculated the QCT spectra for both K9K-1 and K7E-1 isomers, as well the harmonic

Table D.6: GC interbasebase hydrogen bond and covalent bond distances (Å) and angles (°) in the K9K-1 and K7E-1 isomers of G·C computed with ANI-2x model. The shifts of ANI-2x values with respect the ones from DFT-D/B3LYP, $\Delta_{\mathrm{ANI-2x/DFT}}$, are reported in round brackets.

| H-bond AH$\cdots$B | $\mathrm{d_{AB}}$ | $\mathrm{d_{HB}}$ | $\mathrm{d_{AH}}$ | $\angle$AHB |
|---|---|---|---|---|
| *Canonical structure:* K9K-1 | | | | |
| ANI-2x | | | | |
| N4H$\cdots$O6 | 2.967 (0.213) | 1.981 (0.269) | 1.015 (-0.026) | 163.0 (-16.6) |
| N1H$\cdots$N3 | 2.895 (-0.008) | 1.887 (0.020) | 1.033 (-0.003) | 164.0 (-13.7) |
| N2H$\cdots$O2 | 2.883 (-0.009) | 1.895 (0.027) | 1.015 (-0.009) | 163.8 (-15.8) |
| DFT-D/B3LYP[a] | | | | |
| N4H$\cdots$O6 | 2.754 | 1.712 | 1.041 | 179.6 |
| N1H$\cdots$N3 | 2.903 | 1.867 | 1.036 | 177.7 |
| N2H$\cdots$O2 | 2.892 | 1.868 | 1.024 | 179.6 |
| *Enol Cyt structure:* K7K-1 | | | | |
| ANI-2x | | | | |
| N4H$\cdots$O6 | 3.028 (0.235) | 2.033 (0.273) | 1.013 (-0.02) | 166.7 (-11.8) |
| N1H$\cdots$N3 | 3.090 (0.111) | 2.108 (0.162) | 1.023 (-0.011) | 160.1 (-16.8) |
| N2H$\cdots$O2 | 3.056 (0.034) | 2.074 (0.063) | 1.014 (-0.006) | 162.4 (-14.6) |
| DFT-D/B3LYP[a] | | | | |
| N4H$\cdots$O6 | 2.793 | 1.760 | 1.033 | 178.5 |
| N1H$\cdots$N3 | 2.979 | 1.946 | 1.034 | 176.9 |
| N2H$\cdots$O2 | 3.022 | 2.011 | 1.020 | 177.0 |

[a] With basis set aVDZ. The calculation was performed by G. Botti.

fundamental frequencies. Our results are compared with DFT-D/B3LYP data, by considering only the normal modes that are equivalent in the two different computational approaches.

**Accuracy of the ANI-2x model** The ANI-2x model presents at the harmonic level large $\Delta_{\mathrm{ANI-2x/DFT}}$ for K9K-1, in particular for the $NH_2^s$ of Gua and N1H of Cyt. The situation is improved for $NH_2^s$ when anharmonicity is included but becomes worse for all the other modes, suggesting that the ANI-2x model is not accurate enough to describe the vibrational properties of K9K-1. For K7E-1 the agreement with the DFT data is better, in particular for the OH stretching vibration of the Cytosine, which presents the lowest $\Delta_{\mathrm{ANI-2x/DFT}}$ compared to the other vibrational modes.

**QCT spectra** Figure D.3 presents the QCT spectra of both the isomers calculated with ANI-2x model and compared with the corresponding harmonic and experimental values. In K9K-1 and K7E-1 all the QCT peaks are red-shifted with respect to the harmonic results, with the exception of N1H Gua in K9K-1. For K7E-1, the average absolute error respect to the experiment is 107 cm

Table D.7: NH and OH stretching frequencies characterizing the hydrogen bonds region of G·C complex in the canonical and tautomer structures. The shifts of ANI-2x values with respect the ones from DFT-D/B3LYP, $\Delta_{\mathrm{ANI-2x/DFT}}$, are reported in round brackets. The MAE with respect to the experiments is reported for the form K7E-1.

| *Canonical structure:* K9K-1 | | | | | |
|---|---|---|---|---|---|
| | ANI-2x | | DFT-D/B3LYP[a] | | |
| | Harmonic | QCT | Harmonic | QCT | |
| $NH_2^a$ Gua | 3734 (18) | 3637 (42) | 3716 | 3595 | |
| $NH_2^a$ Cyt | 3743 (51) | 3660 (75) | 3692 | 3585 | |
| $NH_2^s$ Gua | 3567 (178) | 3514 (140) | 3389 | 3374 | |
| $NH_2^s$ Cyt | 3587 | 3491 | | | |
| N1H Gua | 3315 | 3342 | | | |
| N1H Cyt | 3741 (99) | 3716 (170) | 3642 | 3546 | |
| N9H Gua | 3725 (56) | 3642 (73) | 3669 | 3569 | |
| *Enol Cyt structure:* K7E-1 | | | | | |
| | ANI-2x | | DFT-D/B3LYP[a] | | Exp.[b] |
| $NH_2^a$ Gua | 3767 (58) | 3665 (56) | 3709 | 3609 | 3552 |
| $NH_2^a$ Cyt | 3709 (9) | 3628 (47) | 3700 | 3581 | 3532 |
| $NH_2^s$ Gua | 3605 (49) | 3549 (93) | 3556 | 3456 | 3426 |
| $NH_2^s$ Cyt + N1H Gua | 3550 | 3459 | | | |
| N1H Gua + $NH_2^s$ Cyt | 3589 | 3529 | | | |
| OH Cyt | 3793 (9) | 3653 (-28) | 3784 | 3681 | 3603 |
| N7H Gua | 3767 (76) | 3665 (74) | 3675 | 3558 | 3510 |
| MAE | 204 | 107 | 160 | 52 | |

[a] With basis set 6-31G**. The calculation was performed by G. Botti.
[b] From ref. [215].

$^{-1}$, which is more accurate about one hundred value than the harmonic values (MAE 204 cm $^{-1}$).

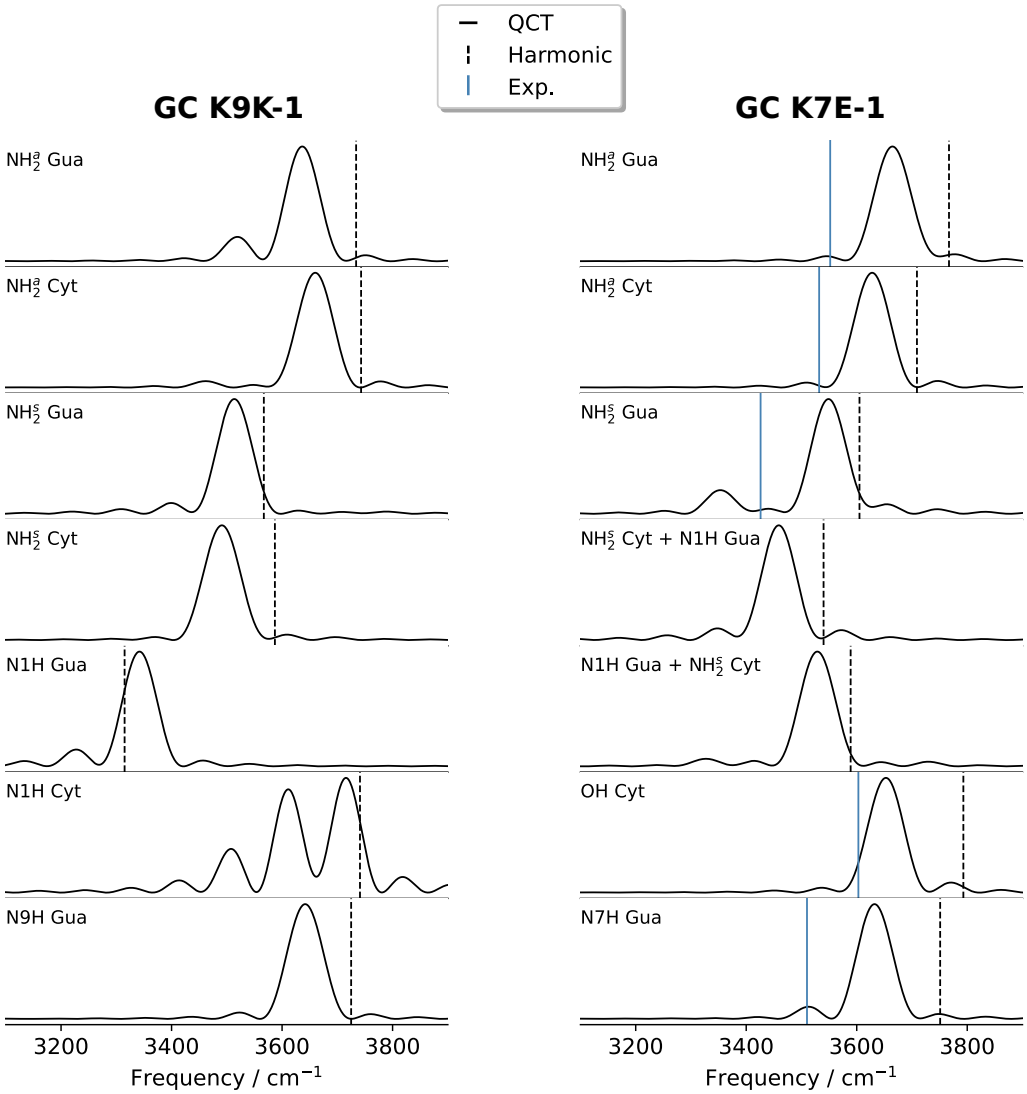



Figure D.3: QCT spectra of NH and OH stretching modes of K9K-1 and K7E-1 isomers of G·C base pair calculated with ANI-2x potential. The experimental values are taken from ref. [215].

# CONCLUSION

Hydrogen bond is present in nature in the most disparate forms, being a fundamental interaction in both inorganic and organic matter. Its features in different chemical environments make it a very fascinating object to study. Hydrogen-bonded systems are intrinsically affected by an anharmonic contribution, which can be more or less significant, to the shape of the potential energy surface, which is usually rather shallow. Although some similarities can be generally drawn, it is very difficult to predict the behavior of the proton in a hydrogen-bonded system, where nuclear quantum effects (NQEs) come into play. NQEs can have a strong impact on the sampling of the configurational space as the quantum (Bose-Einstein) and classical (Boltzmann) distributions are pretty different at low temperatures. As a general trend, NQEs enhance anharmonic effects, thus making the interpretation of vibrational spectra rather involved. The dynamics of light atoms (such as H and D) can be also impacted by tunneling phenomena, in sharp contrast to the classical dynamics. The complex interplay between NQEs and the anharmonicity of the PES greatly complicates the issue. Therefore, the phenomena that are related to hydrogen bonds cannot be easily predicted in such a variegated scenario. In this thesis, we have focused on the study of the proton behavior in selected hydrogen-bonded systems, beyond the classical and harmonic pictures.

NQEs can have an essential role in the phase transitions involving hydrogen atoms. The antiferroelectric to paraelectric phase transition in potassium hydroxide is governed by the hydrogen bonds. Here, weak hydrogen bonds (O-O $>$ 3 Å) keep together the crystal structure of KOH via formation of a zig-zag chain along the *b*-axis direction between the K-O bilayers. By means of *ab initio* PIMD, we have shown that the phase transition can be rationalized in terms of a flipping back and forth of the protons, which is associated to a OH bending motion. This results in a dynamically disordered paraelectric phase at high temperature. Under deuteration, a geometric H/D isotope effect is found in the low-temperature antiferroelectric phase. In KOD the hydrogen bonds are shorter than in KOH (inverse Ubbelohde effect), causing a structural contraction of the K-O layers, as evidenced by the lattice parameters.

When a molecule complex forms hydrogen bonds with a substrate, the delocalization of the quantum proton can also play a crucial role in the adsorption. For instance, the adsorption of formic acid

(HCOOH) on the $TiO_2$ anatase (101) is characterized by a special type of molecular adsorption, in which HCOOH in a monodentate configuration interacts with the surface via a strong hydrogen bond. This is the most stable type of adsorption at T = 0 K at the DFT/PBE level of theory, while the dissociation along the same type of monodentate mode is unfavorable. The molecular adsorption is the most favored one even when including quantum effects through PIMD but, in this scenario, the proton frequently shuttles between the molecule and the surface. The potential energy surface experienced by the proton is highly anharmonic, making very difficult to catch the OH stretching vibration. With the use of DC-SCIVR, we were able to identify the OH stretch, which is found at abnormally low frequencies ($< 2700$ cm$^{-1}$) in the nearby of the CH stretching region. When the acid is deuterated (HCOOD), the molecule-surface hydrogen bond becomes weaker, thus the red-shift of the OD stretch with respect to the gas-phase HCOOD is less significant than the corresponding one in HCOOH.

The role of the solvent in the equilibrium and vibrational properties of guanine and cytosine base pair is fundamental to understand the interactions in the DNA building blocks. A description of the hydrogen bonds in the guanine and cytosine dimer (dG·dC) in physiological conditions can be done by employing the polarizable force field AMOEBA. The guanine and cytosine dimer presents three hydrogen bonds. In the gas-phase dimer the proton delocalization due to the zero-point energy motion is significant also at room temperature, as suggested by adQTB and PIMD simulations. The effect of the solvent is not easy to anticipate. The vibrational spectra of solvated dG·dC simulated with quasi-classical trajectory method reveal an improper blue-shift of the NH stretching frequencies related to the interbase region. However, solvated dG·dC results to be unstable in the ns length scale due to the absence of the backbone, which ensures the stability of the dimer. In the near future, we plan to analyze more in detail the role of the solvent on the equilibrium and vibrational properties of the dG·dC dimer, by simulating a DNA like system.

The investigation on the selected hydrogen-bonded systems has shown how unpredictable can be the proton and deuteron behavior in different atomic environments. On the one hand, the KOH and KOD hydroxides are a manifestation of the importance of hydrogen bonds in inorganic solids. Here, the hydrogen bonds, although being weak, are fundamental to keep together the KOH structure. Since the phase transition is mostly governed by low-frequency libration modes, NQEs contribute to weaken even more the hydrogen bonds as evidenced by the differences with respect to the classical and the deuterated crystal. On the contrary, the opposite effect is found in the adsorbed formic acid on the $TiO_2$ anatase (101) surface, where the strong hydrogen bond between the molecule and the surface is reinforced when NQEs are included. This results in a largely enhanced anharmonicity affecting the vibrational spectra. Lastly, the guanine and cytosine base pair presents three hydrogen bonds that are normal to rather weak. The latter are strongly influenced by the presence of the solvent, which can strengthen/weaken the hydrogen bonds. At the present of the study, we are still investigating the impact of anharmonicity and NQEs on the hydrogen bond strength, which is essential to understand better the stability of DNA double helix.

All through this study, we have observed that quantum and anharmonic effects are fundamental to understand both the thermodynamic equilibrium and vibrational properties of hydrogen-bonded systems. At the same time, an accurate description of NQEs requires an accurate PES. At the present day, this is still very challenging. In this thesis we have employed both DFT methods, based

on the GGA-PBE approximation, and force field approaches. On the one hand, distinct approximations for the exchange-correlation energy in the DFT may considerably modify the hydrogen bond energy, the barriers and the overall PES profile, resulting in appreciable shifts of the vibrational frequencies, even at the simple harmonic level. On the other hand, force-field based methods need to be used very carefully; a benchmark with *ab initio* data, as we did in this thesis, is preferable when possible. One possible solution to treat high-dimensional systems, and, more generally, to include NQEs in molecular simulations with an affordable cost is to explore the PES via machine learning (ML) interatomic potentials. A lot of attention has been devoted in the recents years to this subject and several different frameworks have been developed [19, 20]. A big challenge is the transferability of the potential. In this respect, the ANI model [227, 230], which we benchmarked for the gas-phase of guanine and cytosine base pair, can be a promising solution. Very recently, they ANI-2x potential has been interfaced in a ML/MM framework with the AMOEBA force field [231].

# GLOSSARY

**adQTB** Adaptive Quantum Thermal Bath.

**AIMD** *ab initio* Molecular Dynamics.

**B3LYP** Becke-3-parameter-Lee-Yang-Parr.

**BO** Born-Oppenheimer.

**DC-SCIVR** Divide and Conquer Semiclassical Initial Value Representation.

**DFT** Density Functional Theory.

**FDT** Fluctuation-Dissipation Theorem.

**GGA** Generalized Gradient Approximation.

**MC-SCIVR** Multiple Coherent States Semiclassical Initial Value Representation.

**MD** Molecular Dynamics.

**NEB** Nudged Elastic Band.

**NQEs** Nuclear Quantum Effects.

**PBE** Perdew-Burke-Ernzerhof.

**PES** Potential Energy Surface.

**PIMD** Path Integral Molecular Dynamics.

**QCT** Quasi-classical trajectory.

**QE** Quantum Espresso.

**QTB** Quantum Thermal Bath.

**SC** Semiclassical.

**SCIVR** Semiclassical Initial Value Representation.

**SPA** Stationary Phase Approximation.

**TA-SCIVR** Time Averaging Semiclassical Initial Value Representation.

**TDSE** Time Dependent Schrödinger equation.

**ZPE** Zero Point Energy.

# BIBLIOGRAPHY


1. Nibbering, E. T., Dreyer, J., Kühn, O., Bredenbeck, J., Hamm, P. & Elsaesser, T. in *Analysis and control of ultrafast photoinduced reactions* 619–687 (Springer, 2007) (cit. on pp. 8, 25).
2. Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2,** 1–14 (2018) (cit. on p. 9).
3. Pereyaslavets, L. *et al.* On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proceedings of the National Academy of Sciences* **115,** 8878–8882 (2018) (cit. on p. 9).
4. Ceriotti, M., Fang, W., Kusalik, P. G., McKenzie, R. H., Michaelides, A., Morales, M. A. & Markland, T. E. Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chemical reviews* **116,** 7529–7550 (2016) (cit. on p. 9).
5. Vardi-Kilshtain, A., Nitoker, N. & Major, D. T. Nuclear quantum effects and kinetic isotope effects in enzyme reactions. *Archives of biochemistry and biophysics* **582,** 18–27 (2015) (cit. on p. 9).
6. Castro, C. & Karney, W. L. Heavy-Atom Tunneling in Organic Reactions. *Angewandte Chemie* **132,** 8431–8442 (2020) (cit. on p. 9).
7. Morrone, J. A. & Car, R. Nuclear quantum effects in water. *Physical review letters* **101,** 017801 (2008) (cit. on p. 9).
8. Meier, T., Petitgirard, S., Khandarkhaeva, S. & Dubrovinsky, L. Observation of nuclear quantum effects and hydrogen bond symmetrisation in high pressure ice. *Nature communications* **9,** 1–7 (2018) (cit. on p. 9).
9. McBride, C., Noya, E. G., Aragones, J. L., Conde, M. M. & Vega, C. The phase diagram of water from quantum simulations. *Physical Chemistry Chemical Physics* **14,** 10140–10146 (2012) (cit. on p. 9).

10. Bronstein, Y., Depondt, P., Finocchi, F. & Saitta, A. M. Quantum-driven phase transition in ice described via an efficient Langevin approach. *Physical Review B* **89,** 214101 (2014) (cit. on pp. 9, 48, 67).

11. Earnshaw, A. & Greenwood, N. N. *Chemistry of the Elements* (Butterworth-Heinemann Oxford, 1997) (cit. on p. 9).

12. Marx, D. & Parrinello, M. Ab initio path integral molecular dynamics: Basic ideas. *The Journal of chemical physics* **104,** 4077–4082 (1996) (cit. on pp. 9, 34, i).

13. Dammak, H., Chalopin, Y., Laroche, M., Hayoun, M. & Greffet, J.-J. Quantum thermal bath for molecular dynamics simulation. *Physical Review Letters* **103,** 190601 (2009) (cit. on pp. 9, 47, i).

14. Mangaud, E., Huppert, S., Plé, T., Depondt, P., Bonella, S. & Finocchi, F. The fluctuation–dissipation theorem as a diagnosis and cure for zero-point energy leakage in quantum thermal bath simulations. *Journal of chemical theory and computation* **15,** 2863–2880 (2019) (cit. on pp. 9, 47, 48, 49, i).

15. Miller, W. H. Classical S matrix: Numerical application to inelastic collisions. *The Journal of Chemical Physics* **53,** 3578–3587 (1970) (cit. on pp. 9, 39, i).

16. Born, M. & Oppenheimer, R. Zur Quantentheorie der Molekeln. *Annalen der Physik* **389,** 457–484 (1927) (cit. on p. 13).

17. Marx, D. & Hutter, J. *Ab initio molecular dynamics: basic theory and advanced methods* (Cambridge University Press, 2009) (cit. on pp. 14, 21).

18. Lower, S. & Neils, T. The Potential-Energy Surface Can Be Calculated Using Quantum Mechanics. *Chemistry LibreTexts* (2014) (cit. on p. 15).

19. Behler, J. Four generations of high-dimensional neural network potentials. *Chemical Reviews* **121,** 10037–10072 (2021) (cit. on pp. 15, 136).

20. Deringer, V. L., Caro, M. A. & Csányi, G. Machine learning interatomic potentials as emerging tools for materials science. *Advanced Materials* **31,** 1902765 (2019) (cit. on pp. 15, 136).

21. Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Physical review* **136,** B864 (1964) (cit. on p. 15).

22. Finocchi, F. Density Functional Theory for Beginners: Basic Principles and Practical Approaches. *Institut des NanoSciences de Paris (INSP) CNRS and University Pierre et Marie Curie, Paris* (2011) (cit. on p. 16).

23. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review* **140,** A1133 (1965) (cit. on p. 16).

24. Perdew, J. P. & Schmidt, K. *Jacob's ladder of density functional approximations for the exchange-correlation energy* in *AIP Conference Proceedings* **577** (2001), 1–20 (cit. on p. 17).

25. Toulouse, J. Review of approximations for the exchange-correlation energy in density-functional theory. *arXiv preprint arXiv:2103.02645* (2021) (cit. on p. 17).

26. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Physical review letters* **77,** 3865 (1996) (cit. on pp. 18, 79, 108).

27. Hertwig, R. H. & Koch, W. On the parameterization of the local correlation functional. What is Becke-3-LYP? *Chemical Physics Letters* **268,** 345–351 (1997) (cit. on p. 18).

28. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **25,** 1157–1174 (2004) (cit. on p. 19).

29. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of chemical theory and computation* **11,** 3696–3713 (2015) (cit. on p. 19).

30. Robertson, M. J., Tirado-Rives, J. & Jorgensen, W. L. Improved peptide and protein torsional energetics with the OPLS-AA force field. *Journal of chemical theory and computation* **11,** 3499–3509 (2015) (cit. on p. 19).

31. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., Grubmüller, H. & MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14,** 71–73 (2017) (cit. on p. 19).

32. Ponder, J. W. *et al.* Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **114,** 2549–2564 (2010) (cit. on p. 19).

33. Ren, P. & Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B* **107,** 5933–5947 (2003) (cit. on p. 19).

34. Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W. & Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *Journal of chemical theory and computation* **9,** 4046–4063 (2013) (cit. on p. 19).

35. Zhang, C., Lu, C., Jing, Z., Wu, C., Piquemal, J.-P., Ponder, J. W. & Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. *Journal of chemical theory and computation* **14,** 2084–2108 (2018) (cit. on pp. 19, 116, 126).

36. Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society* **114,** 7827–7843 (1992) (cit. on p. 20).

37. Tuckerman, M. *Statistical mechanics: theory and molecular simulation* (Oxford university press, 2010) (cit. on pp. 27, 31, 32, 33, 34).

38. Conte, R. & Ceotto, M. Semiclassical molecular dynamics for spectroscopic calculations. *Quantum Chemistry and Dynamics of Excited States: Methods and Applications,* 595–628 (2020) (cit. on pp. 27, 37).

39. Dirac, P. A. in *Feynman's Thesis—A New Approach To Quantum Theory* 111–119 (World Scientific, 2005) (cit. on p. 30).

40. Feynman, R. P. Space-time approach to non-relativistic quantum mechanics. *Reviews of modern physics* **20,** 367 (1948) (cit. on p. 30).

41. Feynman, R. & Hibbs, A. Quantum Mechanics and Path Integrals. McGraw-Hill, New-York (1965) (cit. on p. 30).

42. Chandler, D. & Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *The Journal of Chemical Physics* **74,** 4078–4095 (1981) (cit. on p. 34).

43. Parrinello, M. & Rahman, A. Study of an F center in molten KCl. *The Journal of chemical physics* **80,** 860–867 (1984) (cit. on p. 34).

44. Berne, B. J. & Thirumalai, D. On the simulation of quantum systems: path integral methods. *Annual Review of Physical Chemistry* **37,** 401–424 (1986) (cit. on p. 34).

45. Plé, T. *Nuclear Quantum Dynamics: exploration and comparison of trajectory-based methods* PhD thesis (Sorbonne université, 2020) (cit. on p. 35).

46. Ceriotti, M., Manolopoulos, D. E. & Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *The Journal of chemical physics* **134,** 084104 (2011) (cit. on pp. 36, 102, 110).

47. Ceriotti, M. & Manolopoulos, D. E. Efficient first-principles calculation of the quantum kinetic energy and momentum distribution of nuclei. *Physical review letters* **109,** 100604 (2012) (cit. on pp. 36, 102, 110).

48. Craig, I. R. & Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *The Journal of chemical physics* **121,** 3368–3373 (2004) (cit. on p. 36).

49. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics. *The Journal of chemical physics* **101,** 6168–6183 (1994) (cit. on p. 36).

50. Lawrence, J. E. & Manolopoulos, D. E. Path integral methods for reaction rates in complex systems. *Faraday Discussions* **221,** 9–29 (2020) (cit. on p. 36).

51. Habershon, S., Fanourgakis, G. S. & Manolopoulos, D. E. Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. *The Journal of chemical physics* **129,** 074501 (2008) (cit. on p. 36).

52. Witt, A., Ivanov, S. D., Shiga, M., Forbert, H. & Marx, D. On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *The Journal of chemical physics* **130,** 194510 (2009) (cit. on p. 36).

53. Rossi, M., Ceriotti, M. & Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *The Journal of chemical physics* **140,** 234116 (2014) (cit. on p. 36).

54. Berry, M. V. & Mount, K. Semiclassical approximations in wave mechanics. *Reports on Progress in Physics* **35,** 315 (1972) (cit. on p. 37).

55. Van Vleck, J. H. The correspondence principle in the statistical interpretation of quantum mechanics. *Proceedings of the National Academy of Sciences* **14,** 178–188 (1928) (cit. on p. 38).

56. Maslov, V. & Fedoriuk, M. Semiclassical approximation in quantum mechanics, volume 7 of. *Mathematical physics and applied mathematics* (1981) (cit. on p. 39).

57. Gutzwiller, M. C. Phase-Integral Approximation in Momentum Space and the Bound States of an Atom. *Journal of mathematical Physics* **8,** 1979–2000 (1967) (cit. on p. 39).

58. Miller, W. H. The semiclassical initial value representation: A potentially practical way for adding quantum effects to classical molecular dynamics simulations. *The Journal of Physical Chemistry A* **105,** 2942–2955 (2001) (cit. on p. 40).

59. Miller, W. H. Comment on: Semiclassical time evolution without root searches. *The Journal of chemical physics* **95,** 9428–9430 (1991) (cit. on p. 40).

60. Heller, E. J. Frozen Gaussians: A very simple semiclassical approximation. *The Journal of Chemical Physics* **75,** 2923–2931 (1981) (cit. on p. 40).

61. Herman, M. F. & Kluk, E. A semiclasical justification for the use of non-spreading wavepackets in dynamics calculations. *Chemical Physics* **91,** 27–34 (1984) (cit. on p. 40).

62. Kay, K. G. Integral expressions for the semiclassical time-dependent propagator. *The Journal of chemical physics* **100,** 4377–4392 (1994) (cit. on pp. 40, 41).

63. Kay, K. G. Numerical study of semiclassical initial value methods for dynamics. *The Journal of chemical physics* **100,** 4432–4445 (1994) (cit. on p. 41).

64. Kaledin, A. L. & Miller, W. H. Time averaging the semiclassical initial value representation for the calculation of vibrational energy levels. *The Journal of chemical physics* **118,** 7174–7182 (2003) (cit. on p. 41).

65. Ceotto, M., Atahan, S., Tantardini, G. F. & Aspuru-Guzik, A. Multiple coherent states for first-principles semiclassical initial value representation molecular dynamics. *The Journal of chemical physics* **130,** 234113 (2009) (cit. on pp. 41, 42).

66. Gabas, F., Conte, R. & Ceotto, M. On-the-fly ab initio semiclassical calculation of glycine vibrational spectrum. *Journal of chemical theory and computation* **13,** 2378–2388 (2017) (cit. on p. 42).

67. Zhuang, Y., Siebert, M. R., Hase, W. L., Kay, K. G. & Ceotto, M. Evaluating the accuracy of Hessian approximations for direct dynamics simulations. *Journal of chemical theory and computation* **9,** 54–64 (2013) (cit. on p. 42).

68. Ceotto, M., Zhuang, Y. & Hase, W. L. Accelerated direct semiclassical molecular dynamics using a compact finite difference Hessian scheme. *The Journal of Chemical Physics* **138,** 054116 (2013) (cit. on p. 42).

69. Aieta, C., Micciarelli, M., Bertaina, G. & Ceotto, M. Anharmonic quantum nuclear densities from full dimensional vibrational eigenfunctions with application to protonated glycine. *Nature communications* **11,** 1–9 (2020) (cit. on p. 42).

70. Ceotto, M., Dell'Angelo, D. & Tantardini, G. F. Multiple coherent states semiclassical initial value representation spectra calculations of lateral interactions for CO on Cu (100). *The Journal of chemical physics* **133,** 054701 (2010) (cit. on p. 42).

71. Conte, R., Aspuru-Guzik, A. & Ceotto, M. Reproducing deep tunneling splittings, resonances, and quantum frequencies in vibrational spectra from a handful of direct ab initio semiclassical trajectories. *The journal of physical chemistry letters* **4,** 3407–3412 (2013) (cit. on p. 42).

72. Micciarelli, M., Gabas, F., Conte, R. & Ceotto, M. An effective semiclassical approach to IR spectroscopy. *The Journal of chemical physics* **150,** 184113 (2019) (cit. on p. 42).

73. Micciarelli, M., Conte, R., Suarez, J. & Ceotto, M. Anharmonic vibrational eigenfunctions and infrared spectra from semiclassical molecular dynamics. *The Journal of chemical physics* **149,** 064115 (2018) (cit. on p. 42).

74. Tamascelli, D., Dambrosio, F. S., Conte, R. & Ceotto, M. Graphics processing units accelerated semiclassical initial value representation molecular dynamics. *The Journal of chemical physics* **140,** 174109 (2014) (cit. on p. 42).

75. Buchholz, M., Grossmann, F. & Ceotto, M. Mixed semiclassical initial value representation time-averaging propagator for spectroscopic calculations. *The Journal of Chemical Physics* **144,** 094102 (2016) (cit. on p. 42).

76. Buchholz, M., Grossmann, F. & Ceotto, M. Application of the mixed time-averaging semiclassical initial value representation method to complex molecular spectra. *The Journal of Chemical Physics* **147,** 164110 (2017) (cit. on p. 42).

77. Buchholz, M., Grossmann, F. & Ceotto, M. Simplified approach to the mixed time-averaging semiclassical initial value representation for the calculation of dense vibrational spectra. *The Journal of chemical physics* **148,** 114107 (2018) (cit. on p. 42).

78. De Leon, N. & Heller, E. J. Semiclassical quantization and extraction of eigenfunctions using arbitrary trajectories. *The Journal of Chemical Physics* **78,** 4005–4017 (1983) (cit. on p. 42).

79. Ceotto, M., Atahan, S., Shim, S., Tantardini, G. F. & Aspuru-Guzik, A. First-principles semiclassical initial value representation molecular dynamics. *Physical Chemistry Chemical Physics* **11,** 3861–3867 (2009) (cit. on p. 42).

80. Ceotto, M., Tantardini, G. F. & Aspuru-Guzik, A. Fighting the curse of dimensionality in first-principles semiclassical calculations: Non-local reference states for large number of dimensions. *The Journal of chemical physics* **135,** 214108 (2011) (cit. on p. 42).

81. Ceotto, M., Di Liberto, G. & Conte, R. Semiclassical "Divide-and-Conquer" method for spectroscopic calculations of high dimensional molecular systems. *Physical Review Letters* **119,** 010401 (2017) (cit. on p. 43).

82. Gabas, F., Di Liberto, G., Conte, R. & Ceotto, M. Protonated glycine supramolecular systems: the need for quantum dynamics. *Chemical science* **9,** 7894–7901 (2018) (cit. on p. 43).

83. Gabas, F., Di Liberto, G. & Ceotto, M. Vibrational investigation of nucleobases by means of divide and conquer semiclassical dynamics. *The Journal of Chemical Physics* **150,** 224107 (2019) (cit. on p. 43).

84. Gabas, F., Conte, R. & Ceotto, M. Semiclassical vibrational spectroscopy of biological molecules using force fields. *Journal of chemical theory and computation* **16,** 3476–3485 (2020) (cit. on pp. 43, 116).

85. Cazzaniga, M., Micciarelli, M., Moriggi, F., Mahmoud, A., Gabas, F. & Ceotto, M. Anharmonic calculations of vibrational spectra for molecular adsorbates: A divide-and-conquer semiclassical molecular dynamics approach. *The Journal of Chemical Physics* **152,** 104104 (2020) (cit. on pp. 43, 96).

86. Cazzaniga, M., Micciarelli, M., Gabas, F., Finocchi, F. & Ceotto, M. Quantum Anharmonic Calculations of Vibrational Spectra for Water Adsorbed on Titania Anatase (101) Surface: Dissociative versus Molecular Adsorption. *The Journal of Physical Chemistry C* **126,** 12060–12073 (2022) (cit. on pp. 43, 78, 96).

87. Di Liberto, G., Conte, R. & Ceotto, M. "Divide and conquer" semiclassical molecular dynamics: A practical method for spectroscopic calculations of high dimensional molecular systems. *The Journal of Chemical Physics* **148,** 014307 (2018) (cit. on p. 43).

88. Rognoni, A., Conte, R. & Ceotto, M. Caldeira–Leggett model vs ab initio potential: A vibrational spectroscopy test of water solvation. *The Journal of Chemical Physics* **154,** 094106 (2021) (cit. on p. 44).

89. Langevin, P. *On the theory of Brownian motion* (1908) (cit. on p. 46).

90. Brown, R. *The Miscellaneous Botanical Works of Robert Brown, Esq....* (Ray Society, 1866) (cit. on p. 46).

91. Huppert, S., Plé, T., Bonella, S., Depondt, P. & Finocchi, F. Simulation of Nuclear Quantum Effects in Condensed Matter Systems via Quantum Baths. *Applied Sciences* **12,** 4756 (2022) (cit. on pp. 47, 48).

92. Brieuc, F., Bronstein, Y., Dammak, H., Depondt, P., Finocchi, F. & Hayoun, M. Zero-point energy leakage in quantum thermal bath molecular dynamics simulations. *Journal of chemical theory and computation* **12,** 5688–5697 (2016) (cit. on pp. 47, 48).

93. Dammak, H., Antoshchenkova, E., Hayoun, M. & Finocchi, F. Isotope effects in lithium hydride and lithium deuteride crystals by molecular dynamics simulations. *Journal of Physics: Condensed Matter* **24,** 435402 (2012) (cit. on p. 48).

94. Bronstein, Y., Depondt, P., Bove, L. E., Gaal, R., Saitta, A. M. & Finocchi, F. Quantum versus classical protons in pure and salty ice under pressure. *Physical Review B* **93,** 024104 (2016) (cit. on p. 48).

95. Leimkuhler, B. & Matthews, C. Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research eXpress* **2013,** 34–56 (2013) (cit. on p. 49).

96. Leimkuhler, B. & Matthews, C. Robust and efficient configurational molecular sampling via Langevin dynamics. *The Journal of chemical physics* **138,** 05B601_1 (2013) (cit. on p. 49).

97. Valasek, J. Piezo-electric and allied phenomena in Rochelle salt. *Physical review* **17,** 475 (1921) (cit. on p. 55).

98. Zhong, W., Vanderbilt, D. & Rabe, K. Phase transitions in BaTi O 3 from first principles. *Physical Review Letters* **73,** 1861 (1994) (cit. on p. 55).

99. Bussmann-Holder, A. & Dalal, N. Order/disorder versus or with displacive dynamics in ferroelectric systems. *Ferro-and Antiferroelectricity,* 1–21 (2006) (cit. on p. 56).

100. Esswein, T. & Spaldin, N. A. Ferroelectric, quantum paraelectric, or paraelectric? Calculating the evolution from $BaTiO_3$ to $SrTiO_3$ to $KTaO_3$ using a single-particle quantum mechanical description of the ions. *Physical Review Research* **4,** 033020 (2022) (cit. on p. 56).

101. Koval, S., Kohanoff, J., Migoni, R. & Tosatti, E. Ferroelectricity and isotope effects in hydrogen-bonded KDP crystals. *Physical Review Letters* **89,** 187602 (2002) (cit. on p. 56).

102. Bastow, T. J., Elcombe, M. M. & Howard, C. J. Ferro- and antiferroelectric phases of the alkali hydroxides. *Ferroelectrics* **79,** 269–271 (1988) (cit. on p. 56).

103. Jacobs, H., Mach, B., Lutz, H.-D. & Henning, J. Bindungsverhältnisse in den kristallinen Phasen von Rubidiumhydroxid und -deuterohydroxid, RbOH und RbOD. *Z. Anorg. Allg. Chem.* **544,** 28–54 (1987) (cit. on p. 56).

104. Bastow, T. J., Elcombe, M. M. & Howard, C. J. Low temperature phase transition in CsOH and CsOD. *Solid state communications* **62,** 149–151 (1987) (cit. on p. 56).

105. Jacobs, H., Mach, B., Harbrecht, B., Lutz, H.-D. & Henning, J. Bindungsverhältnisse in den kristallinen Phasen von Caesiumhydroxid und -deuterohydroxid, CsOH und CsOD. *Z. Anorg. Allg. Chem.* **544,** 55–73 (1987) (cit. on p. 56).

106. Amm, D. T., Bastow, T. J., Jeffrey, K., Heyding, R. D. & Segel, S. L. A low temperature phase transition in NaOD near 150 K. *Thermochim. Acta* **95,** 447–451 (1985) (cit. on p. 56).

107. Bleif, H.-J. & Dachs, H. Cystalline modifications and structural phase transitions of NaOH and NaOD. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **38,** 470–476 (1982) (cit. on p. 56).

108. White, M. A., Perrott, A., Britten, D. & Van Oort, M. J. M. Thermodynamic characterization of the low-temperature phase transformations in KOH and KOD. *J. Chem. Phys.* **89,** 4346–4348 (1988) (cit. on pp. 56, 57).

109. Jacobs, H., Kockelkorn, J. & Tacke, T. Hydroxide des Natriums, Kaliums und Rubidiums: Einkristallzüchtung und röntgenographische Strukturbestimmung an der bei Raumtemperatur stabilen Modifikation. *Z. Anorg. Allg. Chem.* **531,** 119–124 (1985) (cit. on p. 56).

110. Bastow, T. J., Elcombe, M. M. & Howard, C. J. Low temperature phase transition in NaOD. *Solid state communications* **57,** 339–341 (1986) (cit. on p. 56).

111. White, M. A. & Moore, S. A. A calorimetric investigation of the low-temperature phase transition in NaOD. *J. Chem. Phys.* **85,** 4629–4632 (1986) (cit. on p. 56).

112. Elschner, S. & Bastow, T. J. Critical relaxation of 23Na in NaOD near the antiferroelectric transition. *Solid state communications* **60,** 75–77 (1986) (cit. on p. 56).

113. Bastow, T., Elcombe, M. & Howard, C. Low temperature phase transition in KOH and KOD. *Solid state communications* **59,** 257–259 (1986) (cit. on pp. 56, 57, 58, 59, 64, 65, 66, 72, ii).

114. Mach, B., Jacobs, H. & Schäfer, W. Bindungsverhältnisse in kristallinen Phasen von Kaliumdeuterohydroxid, KOD. *Zeitschrift für anorganische und allgemeine Chemie* **553,** 187–195 (1987) (cit. on pp. 56, 57, 58, 64, 65, 66).

115. Bastow, T. J., Segel, S. L. & Jeffrey, K. R. Antiferroelectric transition in KOH(D): A 2H and 39K NMR study. *Solid state communications* **78,** 565–568 (1991) (cit. on pp. 56, 57).

116. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta crystallographica section A: crystal physics, diffraction, theoretical and general crystallography* **32,** 751–767 (1976) (cit. on p. 56).

117. Schaack, S., Mangaud, E., Depondt, P., Huppert, S., Fallacara, E. & Finocchi, F. When quantum fluctuations meet structural instabilities: the isotope- and pressure-induced phase transition in the quantum paraelectric NaOH. *Submitted* (2022) (cit. on p. 57).

118. Adams, D. M., Christy, A. G. & Haines, J. High-pressure phase transition in lithium deuteroxide: a neutron powder diffraction and vibrational spectroscopic study. *J. Phys. Chem.* **96,** 8173–8176 (1992) (cit. on p. 57).

119. Krobok, M. & Holzapfel, W. High-pressure Raman and FTIR studies of solid potassium, rubidium and caesium hydroxides. *Journal of Physics: Condensed Matter* **6,** 9789 (1994) (cit. on p. 57).

120. Otto, J. W. & Holzapfel, W. B. High-pressure structural studies of MOH layered compounds (M = Na, K, Rb, Cs). *J. Phys.: Condens. Matter* **7,** 5461–5476 (1995) (cit. on p. 57).

121. Loveday, J. S., Marshall, W. G., Nelmes, R. J., Klotz, S., Hamel, G. & Besson, J. M. The structure and structural pressure dependence of sodium deuteroxide-V by neutron powder diffraction. *J. Phys.: Condens. Matter* **8,** L597–L604 (1996) (cit. on p. 57).

122. Krobok, M. P., Johannsen, P. G. & Holzapfel, W. B. Raman and FTIR study of NaOH and NaOD under pressure. *J. Phys.: Condens. Matter* **4,** 8141–8150 (1992) (cit. on p. 57).

123. Jacobs, H., Kockelkorn, J. & Tacke, T. Hydroxide des Natriums, Kaliums und Rubidiums: Einkristallzüchtung und röntgenographische Strukturbestimmung an der bei Raumtemperatur stabilen Modifikation. *Zeitschrift Fur Anorganische Und Allgemeine Chemie* **531,** 119–124 (Dec. 1985) (cit. on p. 57).

124. Ibers, J., Kumamoto, J. & Snyder, R. Structure of potassium hydroxide: An x-ray and infrared study. English (US). *Journal of Chemical Physics* **33,** 1164–1170 (Jan. 1960) (cit. on p. 57).

125. Snyder, R. G., Kumamoto, J. & Ibers, J. A. Vibrational spectrum of crystalline potassium hydroxide. *The Journal of Chemical Physics* **33,** 1171–1177 (1960) (cit. on p. 57).

126. Kamisuki Toshio, M. S. Infrared Intensities of Sodium, Calcium, and Potassium Hydroxide Crystals. *Bulletin of the Chemical Society of Japan* **45,** 1345–1348 (1972) (cit. on p. 57).

127. Kanesaka, I., Kawahara, H. & Kawai, K. The vibrational spectrum and the normal coordinate analysis of KOX (X = H or D). *Journal of Raman spectroscopy* **15,** 165–168 (1984) (cit. on p. 57).

128. Fallacara, E., Depondt, P., Huppert, S., Ceotto, M. & Finocchi, F. Thermal and Nuclear Quantum Effects at the Antiferroelectric to Paraelectric Phase Transition in KOH and KOD Crystals. *The Journal of Physical Chemistry C* **125,** 22328–22334 (2021) (cit. on pp. 57, 64, 68).

129. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry* **27,** 1787–1799 (2006) (cit. on p. 59).

130. Barone, V., Casarin, M., Forrer, D., Pavone, M., Sambi, M. & Vittadini, A. Role and effective treatment of dispersive forces in materials: Polyethylene and graphite crystals as test cases. *Journal of computational chemistry* **30,** 934–939 (2009) (cit. on p. 59).

131. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of chemical physics* **132,** 154104 (2010) (cit. on p. 59).

132. Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics* **113,** 9901–9904 (2000) (cit. on pp. 60, 73, 80).

133. Baroni, S., De Gironcoli, S., Dal Corso, A. & Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **73,** 515–562 (2001) (cit. on pp. 61, 73).

134. Ubbelohde, A. R. & Gallagher, K. J. Acid-base effects in hydrogen bonds in crystals. *Acta Crystallogr.* **8,** 71–83 (1955) (cit. on p. 67).

135. Benoit, M., Marx, D. & Parrinello, M. Tunnelling and zero-point motion in high-pressure ice. *Nature* **392,** 258–261 (1998) (cit. on p. 67).

136. Li, X.-Z., Walker, B. & Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. USA* **108,** 6369–6373 (2011) (cit. on p. 67).

137. Bessonette, P. W. R. & White, M. A. Why is there no low-temperature phase transition in NaOH? *J. Chem. Phys.* **110,** 3919–3925 (1999) (cit. on p. 67).

138. Angiolari, F., Huppert, S. & Spezia, R. Quantum versus Classical Unimolecular Fragmentation Rate Constants and Activation Energies at Finite Temperature from Direct Dynamics Simulations (2022) (cit. on p. 70).

139. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter* **21,** 395502 (2009) (cit. on p. 73).

140. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77,** 3865–3868 (1996) (cit. on p. 73).

141. Ceriotti, M., More, J. & Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Comput. Phys. Commun.* **185,** 1019–1026 (2014) (cit. on p. 73).

142. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostatting of path integral molecular dynamics. *J. Chem. Phys.* **133,** 124104 (2010) (cit. on p. 73).

143. Diebold, U. The surface science of titanium dioxide. *Surface science reports* **48,** 53–229 (2003) (cit. on pp. 75, 108).

144. Balajka, J., Hines, M. A., DeBenedetti, W. J., Komora, M., Pavelec, J., Schmid, M. & Diebold, U. High-affinity adsorption leads to molecularly ordered interfaces on TiO2 in air and solution. *Science* **361,** 786–789 (2018) (cit. on p. 75).

145. Galoppini, E. Linkers for anchoring sensitizers to semiconductor nanoparticles. *Coordination Chemistry Reviews* **248,** 1283–1297 (2004) (cit. on p. 75).

146. Pujari, S. P., Scheres, L., Marcelis, A. T. & Zuilhof, H. Covalent surface modification of oxide surfaces. *Angewandte Chemie International Edition* **53,** 6322–6356 (2014) (cit. on p. 75).

147. Hagfeldt, A. & Graetzel, M. Light-induced redox reactions in nanocrystalline systems. *Chemical reviews* **95,** 49–68 (1995) (cit. on p. 75).

148. Persson, P., Bergström, R., Ojamäe, L. & Lunell, S. Quantum-chemical studies of metal oxides for photoelectrochemical applications. *Advances in Quantum Chemistry* **41,** 203–263 (2002) (cit. on p. 75).

149. Persson, P., Lunell, S. & Ojamäe, L. Electronic interactions between aromatic adsorbates and metal oxide substrates calculated from first principles. *Chemical physics letters* **364,** 469–474 (2002) (cit. on p. 75).

150. Zhang, Q.-L., Du, L.-C., Weng, Y.-X., Wang, L., Chen, H.-Y. & Li, J.-Q. Particle-size-dependent distribution of carboxylate adsorption sites on TiO2 nanoparticle surfaces: insights into the surface modification of nanostructured TiO2 electrodes. *The Journal of Physical Chemistry B* **108,** 15077–15083 (2004) (cit. on p. 75).

151. Schiffmann, F., VandeVondele, J., Hutter, J., Wirz, R., Urakawa, A. & Baiker, A. Protonation-dependent binding of ruthenium bipyridyl complexes to the anatase (101) surface. *The Journal of Physical Chemistry C* **114,** 8398–8404 (2010) (cit. on p. 75).

152. Deiana, C., Sakhno, Y., Fabbiani, M., Pazzi, M., Vincenti, M. & Martra, G. Direct synthesis of amides from carboxylic acids and amines by using heterogeneous catalysts: Evidence of surface carboxylates as activated electrophilic species. *ChemCatChem* **5,** 2832–2834 (2013) (cit. on p. 75).

153. Fabbiani, M., Pazzi, M., Vincenti, M., Tabacch, G., Fois, E. & Martra, G. Does the Abiotic Formation of Oligopeptides on TiO2 Nanoparticles Require Special Catalytic Sites? Apparently Not. *Journal of nanoscience and nanotechnology* **18,** 5854–5857 (2018) (cit. on p. 75).

154. Martra, G., Deiana, C., Sakhno, Y., Barberis, I., Fabbiani, M., Pazzi, M. & Vincenti, M. The formation and self-assembly of long prebiotic oligomers produced by the condensation of unactivated amino acids on oxide surfaces. *Angewandte Chemie* **126,** 4759–4762 (2014) (cit. on p. 75).

155. Lambert, J.-F. Adsorption and polymerization of amino acids on mineral surfaces: a review. *Origins of Life and Evolution of Biospheres* **38,** 211–242 (2008) (cit. on p. 75).

156. Henderson, M. A. Complexity in the decomposition of formic acid on the TiO2 (110) surface. *The Journal of Physical Chemistry B* **101,** 221–229 (1997) (cit. on pp. 75, 76).

157. Yun, H. J., Lee, H., Joo, J. B., Kim, W. & Yi, J. Influence of aspect ratio of TiO2 nanorods on the photocatalytic decomposition of formic acid. *The Journal of Physical Chemistry C* **113,** 3050–3055 (2009) (cit. on p. 75).

158. Petrik, N. G. *et al.* Conversion of Formic Acid on Single-and Nano-Crystalline Anatase TiO2 (101). *The Journal of Physical Chemistry C* **125,** 7686–7700 (2021) (cit. on pp. 75, 76, 82, 98).

159. Rezaei, M. & Chermahini, A. N. A DFT study on production of hydrogen from biomass-derived formic acid catalyzed by Pt–TiO2. *International Journal of Hydrogen Energy* **45,** 20993–21003 (2020) (cit. on p. 75).

160. Pellegrino, F., Sordello, F., Mino, L., Minero, C., Hodoroaba, V.-D., Martra, G. & Maurino, V. Formic acid photoreforming for hydrogen production on shape-controlled anatase TiO2 nanoparticles: assessment of the role of fluorides, 101/001 surfaces ratio, and platinization. *ACS Catalysis* **9,** 6692–6697 (2019) (cit. on p. 75).

161. Kwon, S., Lin, T. C. & Iglesia, E. Elementary steps and site requirements in formic acid dehydration reactions on anatase and rutile TiO2 surfaces. *Journal of Catalysis* **383,** 60–76 (2020) (cit. on p. 75).

162. Xu, M., Gao, Y., Moreno, E. M., Kunst, M., Muhler, M., Wang, Y., Idriss, H. & Wöll, C. Photocatalytic activity of bulk TiO 2 anatase and rutile single crystals using infrared absorption spectroscopy. *Physical Review Letters* **106,** 138302 (2011) (cit. on p. 76).

163. Busca, G. Infrared studies of the reactive adsorption of organic molecules over metal oxides and of the mechanisms of their heterogeneously-catalyzed oxidation. *Catalysis today* **27,** 457–496 (1996) (cit. on p. 76).

164. Sayago, D. I., Polcik, M., Lindsay, R., Toomes, R. L., Hoeft, J. T., Kittel, M. & Woodruff, D. P. Structure determination of formic acid reaction products on TiO2 (110). *The Journal of Physical Chemistry B* **108,** 14316–14323 (2004) (cit. on p. 76).

165. Pang, C. L., Lindsay, R. & Thornton, G. Structure of clean and adsorbate-covered single-crystal rutile TiO2 surfaces. *Chemical reviews* **113,** 3887–3948 (2013) (cit. on p. 76).

166. Vittadini, A., Selloni, A., Rotzinger, F. & Grätzel, M. Formic acid adsorption on dry and hydrated TiO2 anatase (101) surfaces by DFT calculations. *The Journal of Physical Chemistry B* **104,** 1300–1306 (2000) (cit. on pp. 76, 77, 80).

167. Gong, X.-Q. & Selloni, A. Role of steps in the reactivity of the anatase TiO2 (101) surface. *Journal of catalysis* **249,** 134–139 (2007) (cit. on p. 76).

168. Nunzi, F. & De Angelis, F. DFT investigations of formic acid adsorption on single-wall TiO2 nanotubes: effect of the surface curvature. *The Journal of Physical Chemistry C* **115,** 2179–2186 (2011) (cit. on p. 76).

169. Kou, L., Frauenheim, T., Rosa, A. & Lima, E. Hybrid density functional calculations of formic acid on anatase TiO2 (101) surfaces. *The Journal of Physical Chemistry C* **121,** 17417–17420 (2017) (cit. on p. 76).

170. Miller, K. L., Falconer, J. L. & Medlin, J. W. Effect of water on the adsorbed structure of formic acid on TiO2 anatase (1 0 1). *Journal of catalysis* **278,** 321–328 (2011) (cit. on p. 76).

171. Li, Y. & Gao, Y. Carboxylic Acid Group-Induced Oxygen Vacancy Migration on an Anatase (101) Surface. *Langmuir* **34,** 546–552 (2018) (cit. on p. 76).

172. Tabacchi, G., Fabbiani, M., Mino, L., Martra, G. & Fois, E. The case of formic acid on anatase TiO2 (101): where is the acid proton? *Angewandte Chemie* **131,** 12561–12564 (2019) (cit. on pp. 76, 78, 79, 80).

173. Wang, Y., Wen, B., Dahal, A., Kimmel, G. A., Rousseau, R., Selloni, A., Petrik, N. G. & Dohnálek, Z. Binding of Formic Acid on Anatase TiO2 (101). *The Journal of Physical Chemistry C* **124,** 20228–20239 (2020) (cit. on pp. 76, 77, 80, 82, 85, 100).

174. Wen, B. & Selloni, A. Hydrogen Bonds and H3O+ Formation at the Water Interface with Formic Acid Covered Anatase TiO2. *The Journal of Physical Chemistry Letters* **12,** 6840–6846 (2021) (cit. on p. 76).

175. Xu, M., Noei, H., Buchholz, M., Muhler, M., Wöll, C. & Wang, Y. Dissociation of formic acid on anatase TiO2 (1 0 1) probed by vibrational spectroscopy. *Catalysis Today* **182,** 12–15 (2012) (cit. on pp. 76, 82, 83).

176. Popova, G. Y., Andrushkevich, T., Chesalov, Y. A. & Stoyanov, E. In situ FTIR study of the adsorption of formaldehyde, formic acid, and methyl formiate at the surface of TiO2 (anatase). *Kinetics and catalysis* **41,** 805–811 (2000) (cit. on p. 76).

177. Grinter, D. C., Nicotra, M. & Thornton, G. Acetic acid adsorption on anatase TiO2 (101). *The Journal of Physical Chemistry C* **116,** 11643–11651 (2012) (cit. on p. 76).

178. Sun, J., Ruzsinszky, A. & Perdew, J. P. Strongly constrained and appropriately normed semilocal density functional. *Physical review letters* **115,** 036402 (2015) (cit. on p. 77).

179. Manzhos, S. & Ihara, M. Computational vibrational spectroscopy of molecule–surface interactions: what is still difficult and what can be done about it. *Physical Chemistry Chemical Physics* (2022) (cit. on p. 78).

180. Giannozzi, P. *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter* **21,** 395502 (2009) (cit. on pp. 79, 108).

181. Bader, R. F. Atoms in molecules. *Accounts of Chemical Research* **18,** 9–15 (1985) (cit. on p. 81).

182. Bronstein, Y. *A study of nuclear quantum effects in hydrogen bond symmetrization via the quantum thermal bath* PhD thesis (Université Pierre et Marie Curie-Paris VI, 2016) (cit. on p. 81).

183. Finocchi, F., Haque, F., Chenot, S., Jupille, J. & Stankic, S. Water dissociation on the low-coordinated sites of MgO nanopowders. *Journal of Materials Research* **34,** 408–415 (2019) (cit. on p. 82).

184. Haque, F., Finocchi, F., Chenot, S., Jupille, J. & Stankic, S. Interplay between single and cooperative H2 adsorption in the saturation of defect sites at MgO nanocubes. *The Journal of Physical Chemistry C* **122,** 17738–17747 (2018) (cit. on p. 82).

185. Miller, K. L., Lee, C. W., Falconer, J. L. & Medlin, J. W. Effect of water on formic acid photocatalytic decomposition on TiO2 and Pt/TiO2. *Journal of Catalysis* **275,** 294–299 (2010) (cit. on p. 82).

186. Liao, L.-F., Wu, W.-C., Chen, C.-Y. & Lin, J.-L. Photooxidation of formic acid vs formate and ethanol vs ethoxy on TiO2 and effect of adsorbed water on the rates of formate and formic acid photooxidation. *The Journal of Physical Chemistry B* **105,** 7678–7685 (2001) (cit. on pp. 82, 84).

187. Nanayakkara, C. E., Dillon, J. K. & Grassian, V. H. Surface adsorption and photochemistry of gas-phase formic acid on TiO2 nanoparticles: the role of adsorbed water in surface coordination, adsorption kinetics, and rate of photoproduct formation. *The Journal of Physical Chemistry C* **118,** 25487–25495 (2014) (cit. on pp. 82, 84, 93, 94, 97).

188. Valiev, M. *et al.* NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* **181,** 1477–1489 (2010) (cit. on p. 85).

189. Nejad, A., Suhm, M. A. & Meyer, K. A. Increasing the weights in the molecular work-out of cis-and trans-formic acid: Extension of the vibrational database via deuteration. *Physical Chemistry Chemical Physics* **22,** 25492–25501 (2020) (cit. on pp. 88, 89, 91, 92, 94, 96, 98).

190. Qu, C. & Bowman, J. M. High-dimensional fitting of sparse datasets of CCSD (T) electronic energies and MP2 dipole moments, illustrated for the formic acid dimer and its complex IR spectrum. *The Journal of Chemical Physics* **148,** 241713 (2018) (cit. on pp. 104, 105, 110, vi).

191. Schran, C., Brieuc, F. & Marx, D. Converged colored noise path integral molecular dynamics study of the zundel cation down to ultralow temperatures at coupled cluster accuracy. *Journal of Chemical Theory and Computation* **14,** 5068–5078 (2018) (cit. on pp. 104, 105, vi).

192. Van Setten, M. J., Giantomassi, M., Bousquet, E., Verstraete, M. J., Hamann, D. R., Gonze, X. & Rignanese, G.-M. The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications* **226,** 39–54 (2018) (cit. on p. 108).

193. Treacy, J. P. *et al.* Geometric structure of anatase Ti O 2 (101). *Physical Review B* **95,** 075416 (2017) (cit. on p. 109).

194. Ceriotti, M., More, J. & Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Computer Physics Communications* **185,** 1019–1026 (2014) (cit. on p. 110).

195. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109,** 1528–1532 (2015) (cit. on p. 110).

196. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020) (cit. on p. 110).

197. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953) (cit. on p. 114).

198. Watson, J. D. & Crick, F. H. *The structure of DNA* in *Cold Spring Harbor symposia on quantitative biology* **18** (1953), 123–131 (cit. on p. 114).

199. Shukla, M. K. & Leszczynski, J. Tautomerism in nucleic acid bases and base pairs: a brief overview. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **3,** 637–649 (2013) (cit. on p. 114).

200. Commons, W. *keto-enol-tautomerism of Guanine* `https://commons.wikimedia.org/wiki/File:Guanine_tautomerism.svg` (cit. on p. 115).

201. Rejnek, J. & Hobza, P. Hydrogen-bonded nucleic acid base pairs containing unusual base tautomers: complete basis set calculations at the MP2 and CCSD (T) levels. *The Journal of Physical Chemistry B* **111,** 641–645 (2007) (cit. on p. 115).

202. Löwdin, P.-O. Proton tunneling in DNA and its biological implications. *Reviews of Modern Physics* **35,** 724 (1963) (cit. on p. 115).

203. Löwdin, P.-O. in *Advances in quantum chemistry* 213–360 (Elsevier, 1966) (cit. on p. 115).

204. Srivastava, R. The role of proton transfer on mutations. *Frontiers in Chemistry* **7,** 536 (2019) (cit. on p. 115).

205. Slocombe, L., Al-Khalili, J. & Sacchi, M. Quantum and classical effects in DNA point mutations: Watson–Crick tautomerism in AT and GC base pairs. *Physical Chemistry Chemical Physics* **23,** 4141–4150 (2021) (cit. on p. 115).

206. Slocombe, L., Sacchi, M. & Al-Khalili, J. An open quantum systems approach to proton tunnelling in DNA. *Communications Physics* **5,** 1–9 (2022) (cit. on p. 115).

207. Soler-Polo, D., Mendieta-Moreno, J. I., Trabada, D. G., Mendieta, J. & Ortega, J. Proton transfer in guanine-cytosine base pairs in B-DNA. *Journal of chemical theory and computation* **15,** 6984–6991 (2019) (cit. on p. 115).

208. Gheorghiu, A., Coveney, P. & Arabi, A. The influence of base pair tautomerism on single point mutations in aqueous DNA. *Interface focus* **10,** 20190120 (2020) (cit. on p. 115).

209. Gabas, F., Conte, R. & Ceotto, M. Quantum Vibrational Spectroscopy of Explicitly Solvated Thymidine in Semiclassical Approximation. *The journal of physical chemistry letters* **13,** 1350–1355 (2022) (cit. on p. 116).

210. Ponder, J. W. *et al.* TINKER: Software tools for molecular design. *Washington University School of Medicine, Saint Louis, MO* **3** (2004) (cit. on p. 116).

211. Yoosefian, M. & Mola, A. Solvent effects on binding energy, stability order and hydrogen bonding of guanine–cytosine base pair. *Journal of Molecular Liquids* **209,** 526–530 (2015) (cit. on p. 116).

212. Yurenko, Y. P., Zhurakivsky, R. O., Samijlenko, S. P. & Hovorun, D. M. Intramolecular CH… O hydrogen bonds in the AI and BI DNA-like conformers of canonical nucleosides and their Watson-Crick pairs. Quantum chemical and AIM analysis. *Journal of Biomolecular Structure and Dynamics* **29,** 51–65 (2011) (cit. on pp. 116, 117).

213. Brauer, B., Gerber, R. B., Kabelac, M., Hobza, P., Bakker, J. M., Abo Riziq, A. G. & de Vries, M. S. Vibrational Spectroscopy of the GC Base Pair: Experiment, Harmonic and Anharmonic Calculations, and the Nature of the Anharmonic Couplings. *The Journal of Physical Chemistry A* **109,** 6974–6984 (2005) (cit. on pp. 118, 129).

214. Bakker, J. M., Compagnon, I., Meijer, G., von Helden, G., Kabeláč, M., Hobza, P. & de Vries, M. S. The mid-IR absorption spectrum of gas-phase clusters of the nucleobases guanine and cytosine. *Physical Chemistry Chemical Physics* **6,** 2810–2815 (2004) (cit. on pp. 118, 129).

215. Nir, E., Janzen, C., Imhof, P., Kleinermanns, K. & De Vries, M. Pairing of the nucleobases guanine and cytosine in the gas phase studied by IR–UV double-resonance spectroscopy and ab initio calculations. *Physical Chemistry Chemical Physics* **4,** 732–739 (2002) (cit. on pp. 118, 129, 130, 132, 133).

216. Nir, E., Plützer, C., Kleinermanns, K. & De Vries, M. Properties of isolated DNA bases, base pairs and nucleosides examined by laser spectroscopy. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics* **20,** 317–329 (2002) (cit. on pp. 118, 129).

217. Abo-Riziq, A., Grace, L., Nir, E., Kabelac, M., Hobza, P. & De Vries, M. S. Photochemical selectivity in guanine–cytosine base-pair structures. *Proceedings of the National Academy of Sciences* **102,** 20–23 (2005) (cit. on p. 118).

218. Bende, A. & Muntean, C. M. The influence of anharmonic and solvent effects on the theoretical vibrational spectra of the guanine–cytosine base pairs in Watson–Crick and Hoogsteen configurations. *Journal of molecular modeling* **20,** 1–12 (2014) (cit. on p. 118).

219. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of chemical physics* **98,** 10089–10092 (1993) (cit. on p. 120).

220. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of chemical physics* **103,** 8577–8593 (1995) (cit. on p. 120).

221. Rackers, J. A., Wang, Z., Lu, C., Laury, M. L., Lagardère, L., Schnieders, M. J., Piquemal, J.-P., Ren, P. & Ponder, J. W. Tinker 8: software tools for molecular design. *Journal of chemical theory and computation* **14,** 5273–5289 (2018) (cit. on p. 126).

222. Adjoua, O. *et al.* Tinker-hp: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using gpus and multi-gpu systems. *Journal of chemical theory and computation* **17,** 2034–2053 (2021) (cit. on p. 127).

223. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S. & Roitberg, A. E. TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *Journal of chemical information and modeling* **60,** 3408–3415 (2020) (cit. on p. 130).

224. Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter* **29,** 273002 (2017) (cit. on p. 130).

225. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical science* **8,** 3192–3203 (2017) (cit. on p. 130).

226. Smith, J. S., Nebgen, B., Lubbers, N., Isayev, O. & Roitberg, A. E. Less is more: Sampling chemical space with active learning. *The Journal of chemical physics* **148,** 241733 (2018) (cit. on p. 130).

227. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1, A data set of 20 million calculated off-equilibrium conformations for organic molecules. *Scientific data* **4,** 1–8 (2017) (cit. on pp. 130, 136).

228. Smith, J. S., Zubatyuk, R., Nebgen, B., Lubbers, N., Barros, K., Roitberg, A. E., Isayev, O. & Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Scientific data* **7,** 1–10 (2020) (cit. on p. 130).

229. Smith, J. S., Nebgen, B. T., Zubatyuk, R., Lubbers, N., Devereux, C., Barros, K., Tretiak, S., Isayev, O. & Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature communications* **10,** 1–8 (2019) (cit. on p. 130).

230. Devereux, C., Smith, J. S., Huddleston, K. K., Barros, K., Zubatyuk, R., Isayev, O. & Roitberg, A. E. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *Journal of Chemical Theory and Computation* **16,** 4192–4202 (2020) (cit. on pp. 130, 136).

231. Inizan, T. J., Plé, T., Adjoua, O., Ren, P., Gökcan, H., Isayev, O., Lagardère, L. & Piquemal, J.-P. Scalable Hybrid Deep Neural Networks/Polarizable Potentials Biomolecular Simulations including long-range effects. *arXiv preprint arXiv:2207.14276* (2022) (cit. on p. 136).

# RÉSUMÉ LONG

La liaison hydrogène est une interaction fondamentale dans la matière inorganique et organique. Ses caractéristiques dans différents environnements chimiques en font un objet très fascinant à étudier. Les systèmes avec des liaisons hydrogène sont intrinsèquement affectés par la forme de la surface d'énergie potentielle (PES), qui induit une contribution anharmonique, qui peut être plus ou moins importante. Il est en outre très difficile de prédire le comportement du proton dans un système à liaisons hydrogène, où les effets quantiques des noyaux (NQEs) entrent en jeu. Les NQEs peuvent avoir un fort impact sur l'échantillonnage de l'espace de configuration car les distributions quantique (Bose-Einstein) et classique (Boltzmann) sont assez différentes à basse température. En règle générale, les NQEs renforcent les effets anharmoniques, rendant ainsi l'interprétation des spectres vibrationnels assez complexe. La dynamique des atomes légers (tels que hydrogène et deutérium) peut également être impactée par des phénomènes d'effet tunnel, contrairement à la dynamique classique. L'interaction complexe entre les NQEs et l'anharmonicité du PES complique grandement le problème. Par conséquent, les phénomènes liés aux liaisons hydrogène ne peuvent pas être facilement prédits dans un scénario aussi varié.

Dans cette thèse, nous nous sommes concentrés sur l'étude du comportement du proton dans trois différentes systèmes à liaisons hydrogène, au-delà de la vision classique et harmonique. L'interaction entre la délocalisation quantique du proton et le confinement atomique est très différente pour chaque système étudié. Pour une description adéquate du mouvement des noyaux, nous avons adopté des méthodes quantiques approximatives, en utilisant l'approximation de Born-Oppenheimer (séparation entre les degrés de liberté des électrons et des noyaux) et en traitant la structure électronique soit par des approches *ab initio* ou des champs de force. Pour le calcul des propriétés indépendantes du temps, nous avons principalement utilisé des méthodes de dynamique moléculaire par intégrales de chemins (PIMD) [12], qui sont basées sur la formulation de la mécanique quantique de Feynman. Alternativement, des approches de bains quantiques avec un thermostat Langevin comme la méthode du bain thermique quantique (QTB) peuvent être utilisées pour reproduire les propriétés quantiques statistiques [13, 14]. Pour avoir une comparaison directe avec les expériences, nous avons calculé les spectres vibrationnels à l'aide de méthodes semi-classiques basées sur la théorie de la représentation de la valeur initiale (SCIVR) [15], qui traitent à la fois des effets anharmoniques et quantiques dans les spectres vibrationnels.

## 1) La transition de phase antiferroélectrique à paraélectrique dans l'hydroxyde de potassium

Les NQEs peuvent avoir un rôle essentiel dans les transitions de phase impliquant des atomes d'hydrogène. L'hydroxyde de potassium cristallin est caractérisé par la présence de liaisons hydrogène faibles et subit une transition de phase ordre-désordre (IVa $\rightarrow$ II). La même transition se produit dans le cristal deutéré mais la température de Curie augmente d'environ 24 K [113]. L'identification des trois structures monocliniques (fig. I) a été une première étape pour la reconstruction de l'énergie potentielle décrivant la transition de phase. Il y a deux phases ordonnées, qui sont caractérisées par des chaînes dipolaires de groupes OH parallèles à l'axe *b*. Ces chaînes forment des liaisons hydrogène faibles entre les deux couches K-O, en stabilisant la structure cristalline de KOH. Les chaînes de dipôles OH voisines peuvent avoir la même orientation (ferroélectrique) ou des orientations opposées (antiferroélectrique). La troisième structure est la phase paraélectrique, où il n'y a pas de formation de liaisons d'hydrogène. Cette structure n'est pas stable et représente un maximum de la PES, tandis que les phases ferroélectrique et antiferroélectrique sont des minimums de la PES.

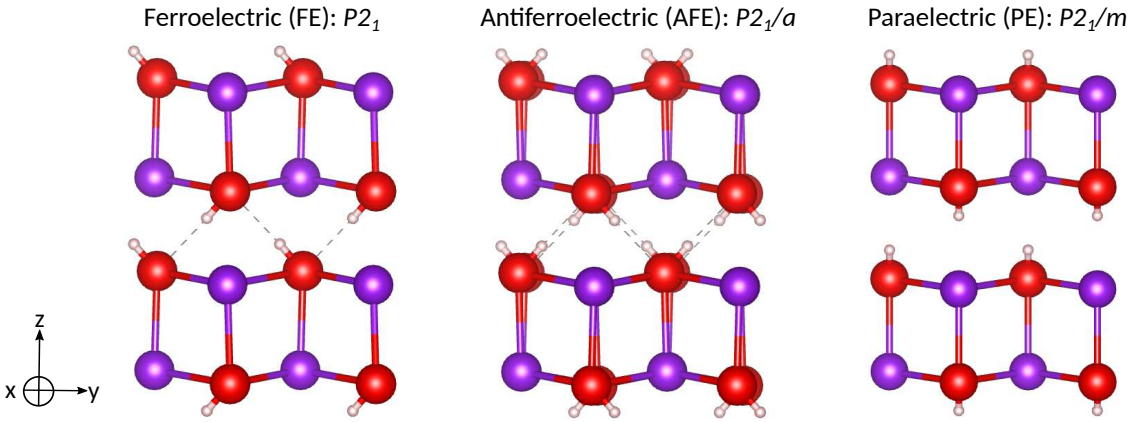



Figure I: Les structures *statiques* du KOH monoclinique: ferroélectrique (FE), antiferroélectrique (AFE) et paraélectrique (PE).

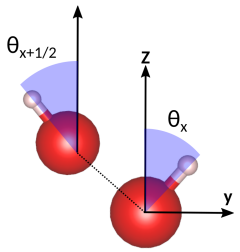



Figure II: Des angles $\theta$ consécutifs dans la direction de l'axe $x$. Leur corrélation caractérise l'ordre de la transition de phase.

Au moyen de *ab initio* PIMD, nous avons montré que la transition de phase peut être rationalisée en termes de basculement des protons, qui est associée à un mouvement de libration des hydroxyles défini par le paramètre d'ordre $\theta$, c'est-à-dire l'angle polaire du groupe OH (fig. II).

Dans la phase à basse température (IVa), les positions des atomes d'hydrogène sont corrélées dans un arrangement antiferroélectrique au sein d'un cristal monoclinique de groupe d'espace $P2_1/a$ (fig. III, T = 77 K). En revanche, la phase à haute température (II) est caractérisée par un mouvement non corrélé des atomes d'hydrogène et, par conséquent, par un désordre dynamique des états ferroélectrique/antiferroélectrique, c'est-à-dire une phase paraélectrique dynamique avec une demi-occupation des sites d'hydrogene, qui restitue un plan miroir (symétrie $P2_1/m$) (fig. III, T = 350 K).

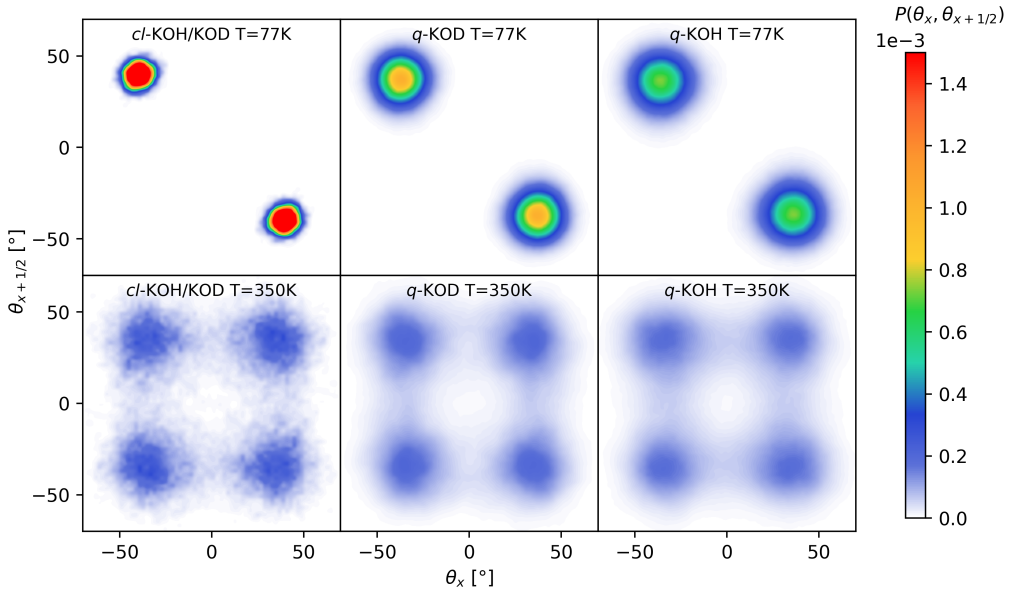



Figure III: Distribution de probabilité conjointe $\theta_x$ et $\theta_{x+1/2}$ obtenue par des simulations *ab initio* classiques (*cl*-KOH/D) et PIMD de l'hydroxide de potassium hydrogéné (*q*-KOH) et deutéré (*q*-KOD).

Le mécanisme de la transition de phase est le même pour l'hydroxide de potassium hydrogéné et deutéré et il est régi par le profil d'énergie du double puits en fonction de l'angle polaire $\theta$. Le va-et-vient des atomes d'hydrogène et deutérium est entraîné par les fluctuations thermiques et du point zéro et génère deux orientations possibles des dipôles OH et OD. Les deux types de fluctuations contribuent à la transition de phase, qu'une image purement classique ne peut pas capturer correctement. Le déplacement isotopique de 24 K de la température de Curie lors de la deutération peut être principalement expliqué en termes de mouvement d'énergie du point zéro plus important du mode de libration OH par rapport à OD.

La présence d'un réseau de liaisons hydrogène faibles est une caractéristique spécifique, qui distingue les cristaux KOH et KOD et la plupart des hydroxydes alcalins des autres ferroélectriques à liaison hydrogène tels que le phosphate de monopotassium (KDP), où les liaisons hydrogène sont fortes. La transition de phase IVa↔II dans KOH et KOD est régie par les liaisons hydrogène faibles, qui sont fortement impactés par les fluctuations quantiques comme mis en évidence, sous deutération, par un effet isotopique géométrique H/D dans la phase antiferroélectrique à basse température. Sous deutération la liaison hydrogène est plus courte et plus forte (fig. IV). En conséquence, les paramètres de réseau $c$ et $\beta$ pour KOH et KOD présentent une différence significative: dans KOD, les bi-couches sont plus proches, donc $c \sin \beta$ est plus petit, provoquant une contraction structurelle

des couches KO.

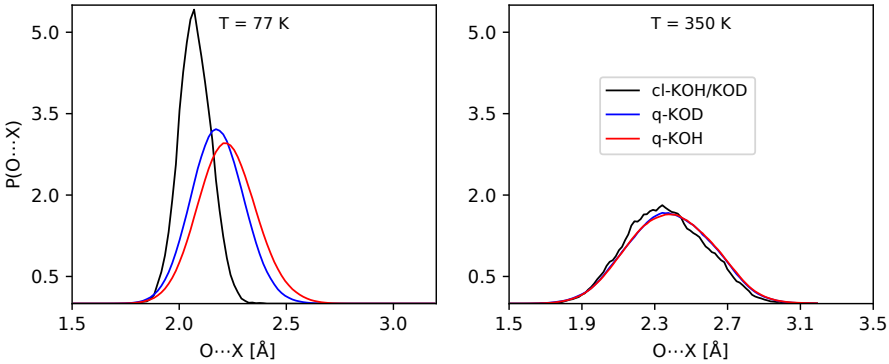



Figure IV: Distribution de probabilité de la longueur O· · · X obtenue par des simulations *ab initio* classiques (*cl*-KOH/D) et PIMD de l'hydroxide de potassium hydrogéné (*q*-KOH) et deutéré (*q*-KOD).

## 2) L'adsorption de l'acide formique sur la surface de l'anatase de titane (101)

Lorsqu'un complexe moléculaire forme des liaisons hydrogène avec un substrat, la délocalisation du proton quantique peut également jouer un rôle crucial dans l'adsorption. C'est le cas, par exemple, de l'adsorption de l'acide formique (HCOOH) sur l'anatase $TiO_2$ (101), qui présente des configurations d'adsorption concurrentes qui font encore débat. D'une part, la molécule d'acide formique peut s'adsorber sous une forme moléculaire; d'autre part, il peut se dissocier en formant une espèce formiate et des groupes des hydroxyles libres en surface. D'après nos calculs à température zéro par la théorie de la densité fonctionnelle, DFT/PBE, la configuration la plus stable est la monodentate moléculaire, qui présente une forte liaison hydrogène avec un site $O_{2c}$ de la surface (fig. V, MH). La dissociation via un mode monodenté similaire est fortement défavorisée et ne représente pas une configuration métastable. Au lieu de cela, la dissociation présente un minimum local pour une configuration bidentée, où le plan OCO du formiate est parallèle à la direction [010] (fig. V, BB(H)).

Cette étude a été menée en collaboration avec des expérimentateurs du laboratoire de l'INSP, qui ont réalisé de nouveaux spectres de spectroscopie infrarouge à transformée de Fourier (FTIR) sur des nanopoudres d'anatase. Les signaux spectroscopiques dans la région OCO sont la clé pour distinguer les configurations bidentées et monodentées alors que la présence ou non de signal du stretching OH peut révéler si la forme est moleculaire ou dissociée. En plus, la deutération sélective dans la position du proton acide, HCOOD, permet de mieux préciser où se situe le proton.

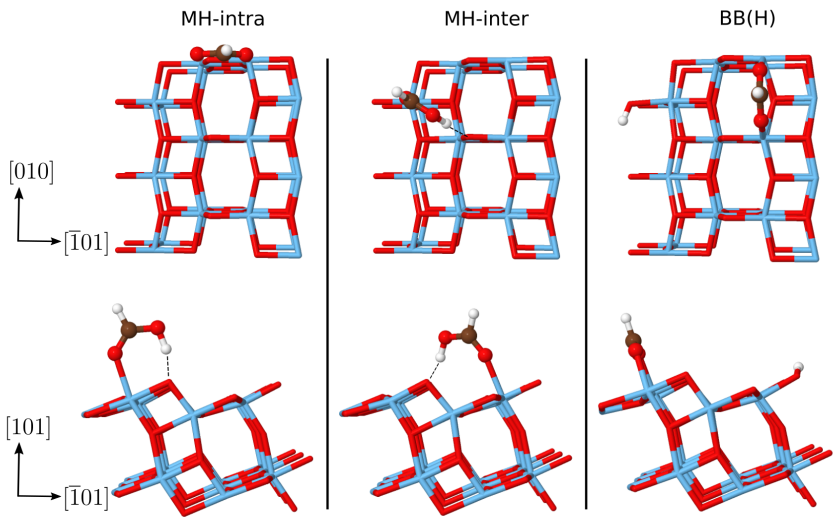



Figure V: Représentation des configurations d'absorption de l'acide formique sur la surface de l'anatase de titane (101): moléculaire monodenté MH intra ou inter et bidenté BB(H).

L'image d'adsorption semble être très complexe, avec des spectres expérimentaux à température ambiante et à basse température riches en différentes caractéristiques spectroscopiques. Les spectres à température ambiante sont dominés par la présence d'espèces bidentées en raison de la présence des caractéristiques $\nu_a$(OCO) et $\nu_s$(OCO) du formiate. Cependant, à basse température, l'adsorption de la molécule FA est différente. En effet, les expériences à basse température ont révélé la formation d'une espèce plus stable, qui, sous deutération spécifique de la position du proton acide (HCOOD), présente un décalage isotopique H/D de certaines bandes (fig. VI).

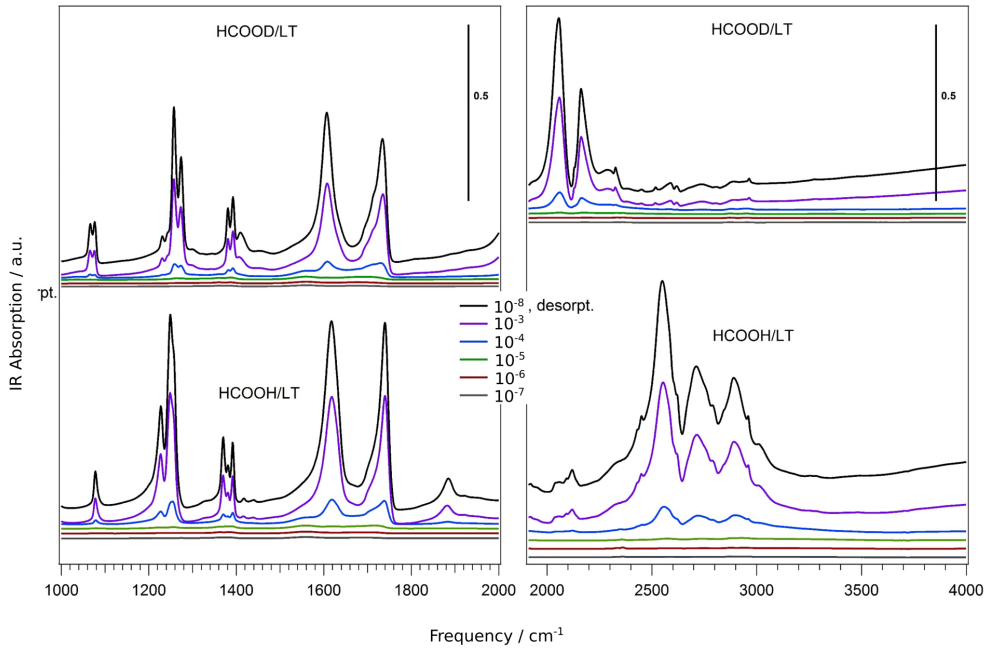



Figure VI: Spectres FTIR at à basse température (13 K) de l'acide formique protoné et deutéré adsorbé sur nanopoudres d'anatase exposé à $P_{FA}$ = $10^{-7}$–$10^{-3}$ mbar, par rapport à les spectres enregistrés après la désorption d'acide formique à $P_{FA} = 10^{-8}$ mbar (courbe noire).

Nous avons alors comparé les spectres FTIR avec les spectres calculés par les approches quasi-classique et semi-classique. Nos calculs sont compatibles avec l'attribution d'une espèce moléculaire monodentée à basse température, bien qu'un bidenté soit également présent. En particulier, nous nous sommes concentrés sur les vibrations du stretching de OH et OD des espèces moléculaires monodentées à liaison hydrogène. Les fréquences du stretching de OH pour le monodentate moléculaire adsorbé sont proches de la région du stretching de CH à $\sim 2550$ cm$^{-1}$. L'estimation harmonique est en bon accord avec ce résultat, mais échoue complètement à décrire le comportement de l'acide formique deutéré monodenté moléculaire, qui montre dans les expériences deux bandes principales à $\sim 2060$ et 2170 cm$^{-1}$. En effet, avec l'inclusion des effets anharmoniques, nous avons révélé que le stretching OH est significativement plus décalé vers le rouge par rapport à la phase gazeuse que le stretching OD, suggérant que les liaisons hydrogène sont plus fortes dans le HCOOH adsorbé que dans le HCOOD adsorbé, principalement en raison de la effets énergétiques du point zéro.

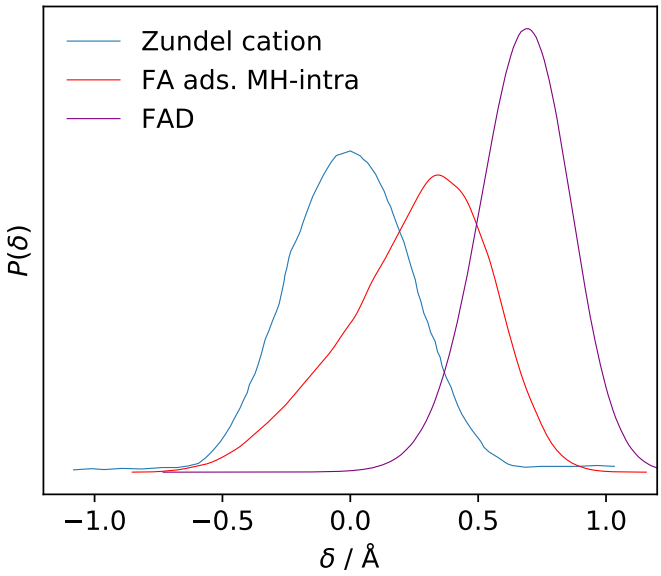



Figure VII: Distribution de probabilité à température 100 K de la coordonnée du "proton sharing" $\delta$ dans le cation zundel (Shran et al. [191], Figure 4, courbe à 2.4 Å) , l'acide formique adsorbé sur la surface d'anatase (101) et le dimère d'acide formique en utilisant la PES de Qu et al. [190].

L'adsorption moléculaire est la plus favorisée même en incluant les effets quantiques via PIMD. La forte liaison hydrogène entre la forme moléculaire monodentate et la surface font faire la navette au proton entre sa position stable sur la molécule d'acide formique et la surface. Les distributions PIMD prédisent un effet isotope géométrique pour la longueur des liaisons hydrogène, qui est plus courte dans HCOOH conformément à ce que nous avons trouvé pour les caractéristiques vibrationnelle. De plus, nous avons montré par l'approche PIMD que le système monodentate acide formique/$TiO_2$ se comporte assez différemment par rapport à d'autres systèmes bien connus, le cation zundel et le dimère d'acide formique (fig. VII). En résumé, dans le système acide formique/$TiO_2$ l'adsorption moléculaire est favorisée même en incluant les effets quantiques des noyaux, bien que le proton soit fréquemment partagé entre la surface et la molécule.

## 3) Une première investigation sur la pertinence des effets quantiques et anharmoniques dans le paire de bases guanine et cytosine en phase phase gazeuse et solvatate

La dynamique des paires de nucléobases, qui sont les éléments constitutifs de l'ADN, est fortement affectée par l'environnement dans lequel ils sont simulés. Pour bien décrire les conditions physiologiques, une étape fondamentale est la solvatation. Cela reste un problème très difficile puisque des calculs *ab initio* précis sont impossibles à obtenir. Une description des liaisons hydrogène dans le dimère de guanine et de cytosine (dG·dC) dans des conditions physiologiques peut être effectuée en utilisant le champ de force polarisable AMOEBA. Le dimère de guanine et de cytosine présente trois liaisons hydrogène (fig. VIII).

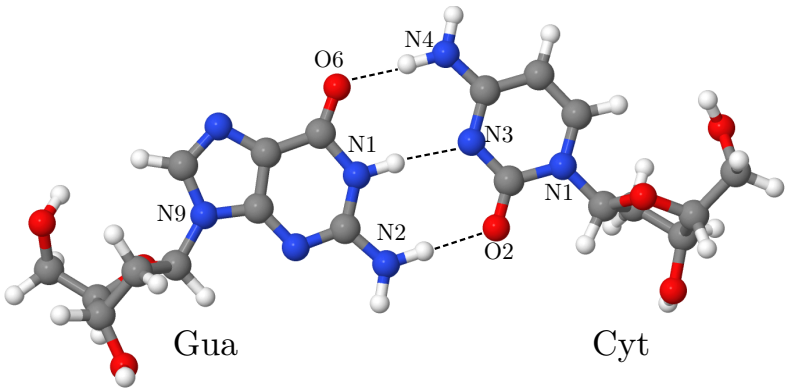



Figure VIII: Représentation de la paire canonique Watson-Crick désoxyguanosine et désoxycytosine nucléoside (dG·dC) utilisée dans les calculs AMOEBABIO18.

Dans le dimère en phase gazeuse, la délocalisation du proton due au mouvement de l'énergie du point zéro est significative également à température ambiante, comme le suggèrent les simulations adQTB et PIMD (fig. IX).

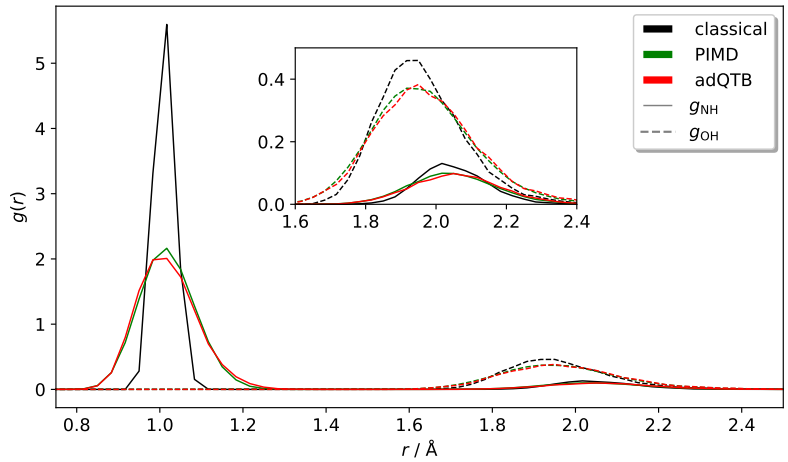



Figure IX: Fonction de distribution radiale $g(r)$ des paires interbases NH et OH de dG·dC en phase gazeuse à 300 K à partir de adQTB, PIMD et des simulations Langevin classiques utilisant AMOEBABIO18.

L'effet du solvant n'est pas facile à anticiper. Les spectres vibrationnels de dG·dC solvatés simulés

avec la méthode de trajectoire quasi-classique révèlent un décalage bleu des fréquences d'étirement NH liées à la région interbase (fig. X). Cependant, le dG·dC solvaté est instable à l'échelle de la nanoseconde en raison de l'absence de la chaîne principale de l'ADN, qui assure la stabilité du dimère. Dans un futur proche, nous prévoyons d'analyser plus en détail le rôle du solvant sur l'équilibre et les propriétés vibrationnelles du dimère dG·dC, en simulant un système de type ADN.

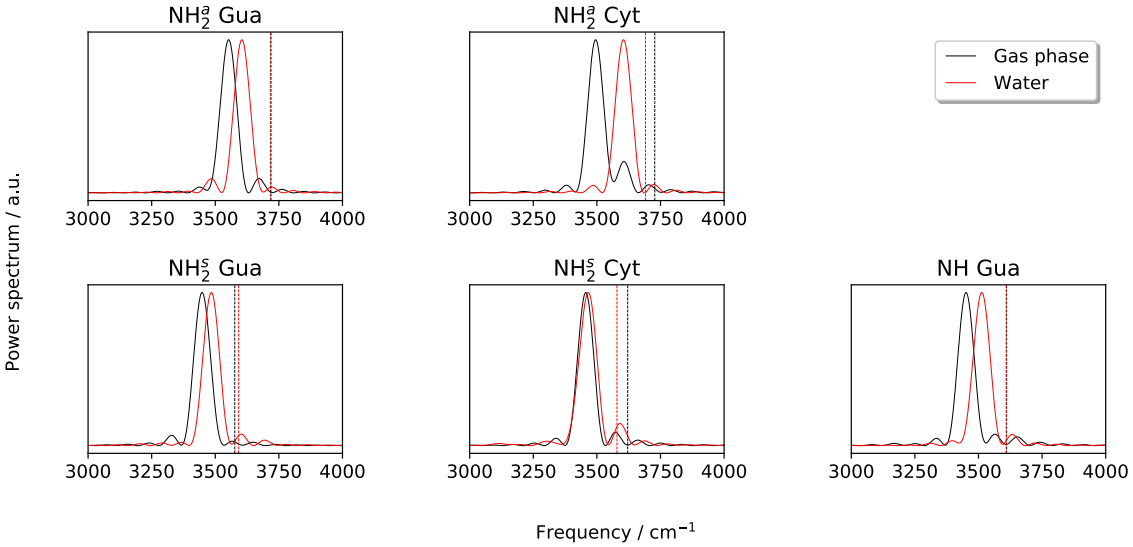



Figure X: Spectres QCT des vibrations du stretching de NH interbase en phase gazeuse et dG·dC solvaté à l'aide d'AMOEBABIO18. Les valeurs harmoniques correspondantes sont indiquées sous forme de lignes pointillées verticales.