Improving molecular dynamics by explictly including Nuclear Quantum Effects

Nastasia Mauger

Sorbonne Université

École Doctorale de Chimie Physique et de Chimie Analytique de Paris Centre
Laboratoire de Chimie Théorique - UMR 7616 CNRS

Improving Molecular Dynamics Simulations: Explicit Inclusion of Nuclear Quantum Effects

Mauger Nastasia

Thèse de doctorat de Chimie Théorique

Présentée et soutenue publiquement le 04/10/2022

devant un jury composé de:

Reviewers	Dammak Hichem Habershon Scott
Examiners	Bonella Sara Laage Damien Marry Virginie
Supervisors	Piquemal Jean-Philip Huppert Simon

Resumé

La dynamique moléculaire (DM) est un outil puissant pour étudier les propriétés de systèmes complexes. Cependant, cette méthode simule le mouvement des particules selon une approche classique via les équations de Newton. Les effets quantiques nucléaires (EQNs), tels que l'énergie du point zéro ou l'effet tunnel, ne sont donc pas pris en compte bien qu'ils aient une influence importante sur les mécanismes de réactions physico-chimiques impliquant des éléments légers, et ce, même à température ambiante. La méthode de référence pour inclure les EQNs est la dynamique moléculaire par intégrales de chemin (PIMD), issue du formalisme de Feynman appliqué à la fonction de partition quantique. Elle repose sur un isomorphisme entre une particule quantique et une chaîne d'oscillateurs harmoniques, appelé beads, reliées par des ressorts. Cependant, le nombre de beads pour converger reste un facteur limitant pour l'application de cette méthode, augmentant drastiquement le temps de calcul (typiquement plusieurs dizaines de fois celui d'une simulation classique) et réduisant la possibilité d'étudier les EQNs sur des systèmes d'intérêt biologique.

Le bain thermique quantique (ou QTB) est une alternative intéressante. Cette méthode repose sur l'équation de Langevin où les différents degrés de liberté classiques du système sont couplés à un ensemble d'oscillateurs harmoniques quantiques. Dans le cas classique, l'équation de Langevin obéit au Théorème de Fluctuation-Dissipation (TFD) classique, qui correspond à l'équipartition de l'énergie. Au contraire, le QTB vise à imposer le TFD quantique qui implique d'injecter davantage d'énergie dans les hautes fréquences que dans les basses fréquences pour reproduire les effets d'énergie de point zéro. Cependant, une fuite d'énergie apparaît lorsque la méthode est appliquée à des systèmes réalistes à cause du couplage des modes de hautes fréquences avec ceux de basses fréquences. Pour corriger cela, une méthode adaptative (adQTB) a été proposée où le théorème de fluctuation-dissipation quantique est utilisé comme un critère afin de corriger de façon systématique la fuite d'énergie au cours d'une simulation.

Les travaux menés dans cette thése ont porté sur l'eau, qui est un composé chimique principalement constitué d'hydrogène. La faible masse de cet atome rend l'impact des EQNs non négligeable. Ces derniers jouent donc un rôle crucial dans la dynamique du solvant. De plus, il a été montré que la prise en compte des EQNS au cours d'une dynamique est déterminante pour l'obtention de propriétés telles que la densité ou l'enthalpie de vaporisation. Afin de vérifier la validité de la méthode adQTB sur un système très anharmonique comme l'eau, les différentes méthodes (PIMD et adQTB) ont été implémentées dans le code de dynamique moléculaire TINKER-HP (CPU et GPU). Les différents résultats obtenus avec le modèle d'eau q-TIP4P/F étant très encourageants pour l'adQTB, un nouveau modèle d'eau polarisable a été développé : Q-AMOEBA. Ce nouveau modèle a permis d'étudier l'effet des EQNs sur une forme fonctionnelle de champ de forces plus complexe qui inclut la polarisation. Cette étude a démontré que l'impact des EQNs dépend du modèle choisi et qu'une généralisation de leurs effets n'est pas possible. Ainsi, avec ce nouveau champ de forces polarisable et grâce au coût de calcul quasi-classique de l'adQTB, les EQNs peuvent maintenant être inclus dans les simulations de DM. Ceci ouvre la voie à l'étude des EQNs dans des systèmes complexes tels que les protéines offrant de potentielles nouvelles applications en Biologie et Pharmacologie.

Abstract

Molecular dynamics (MD) is a powerful tool to study properties of complex systems. However, it treats particles as classical ones by using Newton's equations of motion. Therefore, Nuclear Quantum Effects (NQEs), such as zero point energy or tunneling, are not taken into account, although they can have an influence on the different mechanisms of chemical reactivity that imply light atoms even at ambient temperatures. The reference method to include NQEs is the Path Integral Molecular Dynamics (PIMD) based on Feynman's formalism of quantum mechanics applied to the quantum partition function. It relies on an isomorphism between a quantum particle and a chain of harmonic oscillators, called beads, coupled through harmonic springs. However, the number of beads needed to reach convergence (typically several tens) remains the limiting factor and then reduces the possibility to study NQEs on more complex systems.

The Quantum Thermal Bath (QTB) is an interesting alternative. This method relies on the Langevin equation where the different classical degrees of freedom are coupled to a chain of quantum oscillators. In the classical case, the Langevin equation obeys the classical Fluctuation-Dissipation Theorem (FDT), which corresponds to the equipartition of energy. The QTB aims to impose the quantum FDT which implies to inject more energy in the high frequency modes to reproduce the effects of the zero point energy. However, an energy leakage appears when the method is applied to realistic systems due to the coupling between high and low frequency modes. To correct this leakage, an adaptive method has been proposed where the FDT is used as a criterion to systematically correct this unphysical flow of energy.

The work presented in this thesis focuses on water which is an important chemical with a large proportion of hydrogen atoms. Because of their low mass, NQEs cannot be neglected. These NQEs can have a major impact on the solvent dynamics. Moreover, it has been shown that NQEs should be included in the dynamics to correctly recover some thermodynamical properties such as the density or the enthalpy of vaporization. In order to validate the method on highly anharmonic systems such as water, the PIMD and adQTB methods were implemented inside the molecular dynamics software TINKER-HP (CPU and GPU). The different results obtained with the q-TIP4P/F water model were very promising for the adQTB and a new polarizable water model has been developed: Q-AMOEBA. This new model allowed to study NQEs on a more complex functional form which includes polarization. This study shows that the impact of NQEs can no longer be generalized to all water models. Therefore, with this new model and the classical cost of the adQTB method, NQEs can now be included in MD simulations to study more complex systems such as proteins, opening the paths for their extended study in Biology and Pharmacology.

Contents

Introduction

The first molecular dynamics (MD) simulation was performed in the 50's by Fermi, Pasta and Ulam. [1]. Indeed, the equations of motion for a one-dimensional chain of nonlinear oscillators were for the first time integrated numerically to quantify the degree of ergodicity and energy equipartitioning in it. Thanks to developments in computing hardware, high-performance computing [2–5] and improved simulation techniques [6, 7], it is now possible to run molecular dynamics simulations even on personal computers. These advances enable the study of larger and more complex systems to compute the structural and thermodynamical properties of molecules and solids, but also to reach time scales of milliseconds [4]. Then, it helps to predict unobserved phenomena, such as protein folding or DNA-protein interactions, which makes molecular dynamics a powerful valuable tool to design new drugs and materials [8–10]. Although MD is a powerful tool to investigate and predict the properties of complex systems, it relies on Newton's equation of motion and treats particles from a classical point of view. It uses a potential energy function, called also Force Field (FF), to compute the dynamical evolution of the nuclei of the system. However, the studied system should also obey the laws of quantum mechanics. Therefore, Nuclear Quantum Effects (NQEs), such as tunneling, delocalization and zero-point energy motion, are usually neglected, although they could play an important role on the properties of the studied system [11, 12]. NQEs are relevant to describe processes occurring at low temperature [13], high pressure [14, 15] but also at ambient conditions such as reactions involving proton transfer [16] or isotopic effects [17–20], especially if it involves light atoms such as hydrogen [20–22].

Including NQEs in MD simulations is still a theoretical and computational challenge. To include NQES, the standard method is based on Feynman's path integral formulation of quantum mechanics (PIMD) [23, 24] It provides a classical Hamiltonian which is isomorphic to a ring-polymer composed of P copies of the system, called beads and which partition function approaches the quantum mechanical one. Although it requires only the computation of energies and forces for P beads, the computational convergence is rather slow because the number of replicas should be high enough to take into account for the Zero Point Energy (ZPE) given by $\hbar\omega_0\beta < P$ with ω_0 the highest frequency of the system. Therefore, lot of efforts have been made throughout the years to develop different methods to efficiently take into account NQEs inside MD. One of them is the Quantum Thermal Bath (QTB) which uses a Langevin equation with a stochastic noise to simulate NQEs [25, 26]. This stochastic noise includes both thermal and quantum fluctuations. However, it tries to impose the second-kind Fluctuation-Dissipation Theorem (FDT). Therefore, the classical description of the forces couples the different modes and energy of high frequency modes are transferred to the low frequency ones. An alternative method, called adaptive QTB (adQTB), uses the first-kind FDT as a criterion to correct this unphysical flow of energy and recover the correct energy distribution [27]. The method was already studied on model-like systems and its good results motivate to extend the study of the adQTB method on more realistic systems.

Water is one of the most important liquids due to its role in many fields. Its various physical phenomena, such as cooperative hydrogen bond network, large polarizability and strong permanent dipole, motivate researchers to understand and model this compound [20, 21]. Moreover, given the low mass of hydrogen atoms, classical mechanics is not accurate enough to describe the liquid, even at ambient temperature. It has been shown that NQEs have an impact on the structural properties of water due to the interplay between the stretching ZPE and bending ZPE. The former strengthens the hydrogen bonds whereas the latter weakens it [26]. Then, NQEs have a critical impact on the solvent structure. Thus, they should be included in the simulations to correctly recover some thermodynamical properties, such as the density or the enthalphy of vaporization.

Through this thesis, the adQTB method was compared to PIMD on water to study its theoretical limit on a complex and highly anharmonic system. Therefore, it allows to study both the adQTB method and also the ubiquitous NQEs in water. To do so, these methods were implemented inside the TINKER-HP software (CPU and GPU) [5, 28]. Several water models are available and can be used. However, most of them are fitted without taking into account NQEs explicitly and recover different condensed phase properties mainly through fitting of the experimental quantities. Then, using one of these FFs with the adQTB would give a double counting of the NQEs: one time through their implicit inclusion in the FF parameters and one more time with the dynamics. Moreover, many recent models use high quality *ab initio* data as a basis for their FF functional parametrization and it has been shown that with such a high accuracy, NQEs should be explicitly included to accurately reproduce thermodynamic observables [29]. Therefore, we have decided to use the q-TIP4P/F water model [30]. It was specifically fitted on various thermodynamical properties, such as radial distribution function and infrared spectra with RPMD simulations. Moreover, this model is in good agreement for a broad range of thermodynamical observables outside the reference data. Thereby, the adQTB accuracy and efficiency were studied with this model on different properties in Chapter 3. It gives practical guidelines to use this method when studying NQEs on complex and anharmonic systems where the deconvolution procedure and corrections added to the potential and kinetic energy terms become essential to obtain reliable results. A comparison with PIMD results shows that, contrary to the QTB which suffers from ZPE leakage, adQTB is able to capture accurately NQEs with a computational cost comparable to classical MD simulations. Thanks to the promising results obtained with the adQTB, we have decided to develop a new polarizable model by using the AMOEBA functional form [31, 32]. The AMOEBA water model introduces more flexibility than the q-TIP4P/F water model, but it is fitted without taking into account explicit NQEs. Therefore, our new FF, called Q-AMOEBA, was developed using the adQTB method and AMOEBA functional form. The methodology used to derive this new polarizable model and the different results obtained are presented in Chapter 4. As previously done with q-TIP4P/F study, we compared PIMD (and TRPMD) results with the adQTB method with Q-AMOEBA on a broad range of properties. To study the impact of NQEs in water, we have also decided to compare our different results with Q-AMOEBA but without using adQTB (or PIMD) methods. The results obtained with classical MD simulations while using Q-AMOEBA model highlight that the effect of NQEs can no longer be generalized to all FF. Thereby, with the development and implementation of the adQTB method inside the TINKER-HP software and the different results obtained during this thesis, an open path is now available to study NQEs on more complex systems with a reasonable computational cost.

Chapter 1

State of the Art in Molecular Dynamics Simulations

Contents

1.1 An Overview of Quantum Chemistry Methods

The beginning of the twentieth century behold a revolution in physics with the birth of quantum mechanics (QM) thanks to the formulated energy quantization as a linear partial differential eigenvalue problem given by Schrödinger in 1926 [33]. It allows to quantitatively describe the many spectroscopy experiments for which classical mechanics failed. The main equation in QM is the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = H\Psi(r,t) \tag{1.1}$$

The resolution of the quantum mechanical problem requires to find the eigenvalues and eigenstates of the Hamiltonian H, which is usually done by solving the time-independent Schrödinger equation:

$$\hat{H}\Psi(r) = E\Psi(r) \tag{1.2}$$

Equation (1.1) is used to determine the wave function $\Psi(r,t)$ of a system at a given time where r is the collection of three-dimensional coordinates of the involved atomic nuclei and electrons. Most of the quantum chemistry methods relies on solving equation (1.2) to find the energy E of a given system. The quantum Hamiltonian operator $\hat{H}$ can be expressed as:

$$\hat{H} = -\sum_{\alpha=1}^{N}\frac{\hbar^2}{2\,M_\alpha}\nabla_\alpha^2 - \frac{\hbar^2}{2\,m_e}\sum_{i=1}^{n}\nabla_i^2 + \sum_{i<j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,\alpha}\frac{Z_\alpha\,e^2}{|\mathbf{R}_\alpha-\mathbf{r}_i|} + \sum_{\alpha<\beta}\frac{Z_\alpha\,Z_\beta\,e^2}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|} \tag{1.3}$$

with $\hbar$ the reduced Plank constant. The first two terms represent the kinetic energy of the nuclei and electrons respectively with M_α the mass of the nuclei and m_e the electron's mass. $e^2/|\mathbf{r}_i-\mathbf{r}_j|$ is the electron-electron interaction and $-Z_\alpha\,e^2/|\mathbf{R}_\alpha-\mathbf{r}_i|$ is the electron-nuclei interaction with Z_α the nuclear charge of the α-nucleus. The last term represents the nuclei-nuclei interaction. Due to their high cost, these equations can be exactly solved only for systems with one or two electrons. However, it is possible to separate the motion of the nuclei from the motion of the electrons thanks to the Born-Oppenheimer (BO) approximation [34]. The BO approximation relies on the fact that the nuclei have a significantly greater mass than the electrons (approximately three orders of magnitude) and moves much slower. Therefore, we assume that the electrons react instantaneously to any change in the nuclear configurations. Hence, the electrons are always in their ground state. On the other hand, the nuclei do not sense the different electrons on a particular position in space but rather via a potential energy surface (PES) created by the wavefunction of the different electrons. Therefore, the nucleus-nucleus repulsion potential can be treated as constant $\hat{V}_{ext}$. Thus, it is possible to write equation (1.3) as:

$$\begin{aligned}
&\left(-\frac{\hbar^2}{2\,m_e}\sum_{i=1}^{n}\nabla_i^2 + \sum_{i<j}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|} - \sum_{i,\alpha}\frac{Z_\alpha\,e^2}{|\mathbf{R}_\alpha-\mathbf{r}_i|}\right)\Psi_0(\mathbf{R},\mathbf{r}) = E_0(\mathbf{R})\Psi_0(\mathbf{R},\mathbf{r}) \\
&\left(\frac{-\hbar^2}{2}\sum_{\alpha=1}^{N}\frac{\nabla_\alpha^2}{M_\alpha} + \sum_{\alpha<\beta}\frac{Z_\alpha Z_\beta e^2}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}\right)\chi(\mathbf{R}) = E(\mathbf{R})\chi(\mathbf{R})
\end{aligned} \tag{1.4}$$

where $\chi(\mathbf{R})$ is the nuclear wave function and $\Psi_0(\mathbf{r},\mathbf{R})$ is the ground state electronic wavefunction, which

depends parametrically on the position $\mathbf{R}$ of the nuclei. The eigenvalues give a PES which corresponds to the ground state on which the nuclei evolves. It is a $3N$-dimensional complex partial differential equations where N is the number of electrons and atomic nuclei combined, which can be computationally expensive if N becomes too large. For larger systems, finding exact solutions is not possible and approximations have to be found. To do so, it is important to incorporate into these approximated solutions the same properties carried by the exact ones. One of those is that the wave function should be antisymmetric with respect to the interchange of the coordinate of two electrons:

$$\Psi(r_1, r_2, \ldots, r_{N_e}) = -\Psi(r_1, r_2, \ldots, r_{N_e}) \tag{1.5}$$

To impose equation (1.5), it is possible to express the electronic wave function Ψ in terms of Slater determinant (SD):

$$\Psi(\xi_1, \xi_2, \ldots, \xi_{N_e}) = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \phi_a(\xi_1) & \ldots & \phi_m(\xi_1) \\ \phi_a(\xi_2) & \ldots & \phi_m(\xi_2) \\ \vdots & & \vdots \\ \phi_a(\xi_{N_e}) & \ldots & \phi_m(\xi_{N_e}) \end{vmatrix} \tag{1.6}$$

where $\phi_i(\xi)$, called as spin-orbitals, is composed by both the spatial part of the wave function and its spin.

To circumvent the high cost of solving equations (1.1) and (1.2), other methods can be used to calculate approximate solutions. The simplest one is the Hartree-Fock (HF) method which transforms the electronic Schrödinger equation for a system of N_e interacting electrons into N_e one particle equations. It computes the anti-symmetrized product Ψ of one electron function to find the 'best' set of spin orbitals $\{\phi_i\}$ to express the electronic wave function as a single SD. To do so, it minimizes $\int \Psi\, H\, \Psi \mathrm{d}\tau$ by using a finite (and thus incomplete) basis set to approximate some eigenvalues E. Unfortunately, the electronic correlation, which is a property of the wave function, is not taken into account.

To correct the missing electronic correlation contribution, other methods were developed such as Post Hartree-Fock methods and the Density Functional Theory (DFT). The DFT method computes the molecular electron density ρ instead of the wave function on the basis of the Hohenberg and Kohn theorem which states that the non degenerated electronic ground state energy of a given system is uniquely defined by its electron density [35]. Dealing with ρ instead of the wave function Ψ reduces the number of coordinates from $3N_e$ to 3 and allows to treat systems of a reasonable size (several hundreds of atoms). The energy E of the system (as a function of the density ρ) is formulated as follows [36]:

$$E[\rho] = T_s[\rho] + E_{exch}[\rho] + J[\rho] + V_{ne}[\rho] \tag{1.7}$$

with $T_s[\rho]$ being the Kohn and Sham kinetic energy evaluated as if the electrons were non-interacting [36]. Then, $T_s[\rho]$ is defined as a fictitious one electron wavefunction called as Kohn and Sham orbitals. It is also possible to compute $T_s[\rho]$ from a SD. $E_{exch}[\rho]$ is the exchange correlation energy, $J[\rho]$ the electron-electron

repulsion energy and $V_{ne}[\rho]$ the nuclei-electron interaction energy. Each terms can be expressed as:

$$\begin{aligned} T_s[\rho] &= \sum_{i=1}^{N} \left\langle \phi_i \left| -\frac{1}{2}\nabla_i^2 \right| \phi_i \right\rangle \\ J[\rho] &= \frac{1}{2} \iint \frac{\rho(r)\rho(r')}{|r-r'|} \, \mathrm{d}V \mathrm{d}V' \\ V_{ne}[\rho] &= -\sum_{j}^{N} Z_j \int \frac{\rho(r)}{|r-R_j|} \, \mathrm{d}V \end{aligned} \tag{1.8}$$

In practice, for a many electrons system, the exchange correlation functional is unknown. Therefore, several approximations of $E_{exc}[\rho]$ were made and can be divided into different classes:

- Local-Density Approximation (LDA) relies on the assumption that ρ is slowly varying and can be locally treated as a uniform gas.

- Generalized Gradient Approximations (GGA) employ not only ρ but also its gradient in the exchange correlation energy expression. Functionals containing also the Laplacian of the density in their expressions are often called meta-GGA methods [37, 38].

- Hybrid functional where the HF exchange energy is included in the exchange correlation energy function [39, 40]. They are successful for predicting bond lengths, atomization energies or, in general, short-range properties. However, they have difficulties to describe long-range properties such as charge-transfer.

Even if DFT is a powerful method in the quantum field, it encounters difficulties to describe dispersion forces [41] .

Concerning Post Hartree-Fock approaches, the Coupled Cluster (CC) theory is often used as a reference in *ab initio* calculation due to its high accuracy. It is based on an exponential ansatz that can be written as [42]:

$$\Psi_{CC} = e^{\hat{T}} \Psi_0 \tag{1.9}$$

where Ψ_{CC} is the CC wave function. $\hat{T}$ is the cluster operator when applied to the reference wave function Ψ_0 produces a new wave function Ψ_{CC} containing cluster functions where each of them correlated the motion of electrons within specific orbitals. Usually Ψ_0 is a SD constructed using HF orbitals. If $\hat{T}$ includes all possible contributions from all possible orbitals for the N_e systems, then the exact wave-functions within the given one-electron basis may be obtained. The cluster operator is often called excitation operator because it is described as a sum of cluster operators of different excitation level $\hat{T} = \hat{T_1} + \hat{T_2} + ... + \hat{T_N}$. It can be truncated at the single and double excitation (CCSD) or higher with adding triple excitation (CCSD(T)) which is the 'golden' standard. Hence, CCSD(T) will be our *ab initio* energy reference when discussing results obtained through this thesis.

These different methods allow to compute electronic structures of different systems, but one may be interested in modeling analytical forces, spectrum or being able to study larger systems made of millions of atoms. Hence, a different method which allows to sample the phase space is needed: Molecular Dynamics (MD) where different approaches are possible:

- *ab initio* MD where energies and gradients are derived from electronic calculations and used to compute the classical dynamics of the nuclei. It became really popular thanks to the 'Car-Parrinello' approaches where the classical dynamics of nuclei is computed simultaneously with the orbitals used in the wavefunction allowing to study electrons as quantum particles. However, it is really computationally expensive and the size of the studied system is still the major limiting factor [43].

- Classical MD allows to study larger systems (up to millions of atoms). It does not take into account the dynamics of the electrons and hence are not suitable to study reactivity through forming/breaking chemical bonds. However, thanks to the low computational cost of these approaches to compute the potential energy surfaces, more complex systems can be explored. This is the method used during this thesis.

1.2 Classical Molecular Dynamics

Molecular Dynamics (MD) was firstly introduced by Alder and Wainwright [44] and the work of Fermi, Pasta and Ulam in 1955 to simulate systems consisting of hard spheres using a discontinuous potential [1]. However, it is not based on QM and it uses a classical potential. Hence, to change from a fully quantum description to a classical one, approximations have to be done. The first one is the BO approximation. Using Newton's second law of motion, it is possible to derive the classical equations of motion [45, 46]:

$$\begin{cases} m\dot{\mathbf{r}}_i = \mathbf{p}_i \\ m\ddot{\mathbf{r}}_i = \mathbf{F}_i(\mathbf{r}_1, \cdots, \mathbf{r}_N) \end{cases} \tag{1.10}$$

with m the mass of the i-th particle going from 1 to N. It is used to express the velocity $\mathbf{v}_i$ as functions of momenta $\mathbf{p}_i$. $\mathbf{F}$ is the vector of the forces that act on the particles and depend on their positions $\mathbf{r}_i$. In the BO approximation, the effect of the electrons is approximated by a Potential Energy Surface (PES) function V. In this case, the forces are related to V by:

$$\mathbf{F}_i(\mathbf{r}_1, \ldots, \mathbf{r}_N) = -\nabla_{\mathbf{r}_i} V(\mathbf{r}_1, \ldots, \mathbf{r}_N) \tag{1.11}$$

In the context of atomistic simulation equation (1.10) is often also expressed in the equivalent form of Hamilton's equation of motion :

$$\begin{cases} \dot{q_\alpha}(t) = & \frac{\partial H}{\partial p_\alpha} \\ \dot{p_\alpha}(t) = & -\frac{\partial H}{\partial q_\alpha} \end{cases} \tag{1.12}$$

where $p_\alpha(t)$ and $q_\alpha(t)$ are the conjugate momenta and positions respectively with α going from 1 to $3N$ and H is the classical Hamiltonian defined as:

$$H(\mathbf{p}, \mathbf{q}) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}) \tag{1.13}$$

Hence, going from Newton's second law of motion to Hamilton's equation gives $6N$ first-order differential equations instead of $3N$ second-order differential equations. Moreover, an another advantage of the Hamil-

ton formulation is the connection between the Hamiltonian and the PES of a high-dimensional system. H represents the total energy of the system. Furthermore, a trajectory which fulfills equation (1.12) implies that:

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{\alpha=1}^{3N} \left[\frac{\partial H}{\partial q_\alpha} \dot{q}_\alpha + \frac{\partial H}{\partial p_\alpha} \dot{p}_\alpha \right] = 0 \tag{1.14}$$

It is the mathematical transcription of energy conservation, which defines a $6N$-1 dimensional surface in phase space.

1.2.1 Phase Space Sampling and Observables

Having access to the time evolution of the atomic coordinates and velocities of a system corresponds to having access to its phase space which allows to predict the properties of a classical system. The aim of MD simulations is to observe motion of atoms or molecules to understand chemical or biological processes such as molecular arrangement or protein folding. Hence, one may ask how atomic (or microscopic) structures are related to macroscopic properties like temperatures or pressure and how to deal with an immeasurable number of degrees of freedom (1 mol = 6.022×10^{23} particles) and with the complex and non-linear interactions which occur in realistic systems. The relation between macroscopic thermodynamics and microscopic laws of motion is expressed by statistical mechanics [46, 47]. It gives a mathematical framework to explain how macroscopic observables of a system do not depend on the explicit dynamical motion of each particles but rather on a collection of systems sharing macroscopic properties. It is called an ensemble and was firstly introduced by Gibbs. Hence, averages performed over an ensemble yield the thermodynamic quantities of a system and it can be defined by a wide range of thermodynamic variables. So, it is important to know the probability distribution $\mathscr{P}(\mathbf{p}, \mathbf{q})$ in the canonical ensemble as:

$$\mathscr{P}(\mathbf{p}, \mathbf{q}) = e^{-\beta\, H(\mathbf{p},\mathbf{q})}/Z \tag{1.15}$$

where the partition function is defined as $Z = \int \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{q}\, e^{-\beta\, H(\mathbf{p},\mathbf{q})}$ and $\beta = (k_B T)^{-1}$. One of the benefit of the classical ensemble is that there is no correlation between $\mathbf{p}$ and $\mathbf{q}$. Hence, the probability distribution can be expressed and treated separately as:

$$\mathscr{P}(\mathbf{p}, \mathbf{q}) = \mathscr{P}(\mathbf{p}) \cdot \mathscr{P}(\mathbf{q}) = \frac{e^{-\beta \sum_i \frac{\mathbf{p}_i^2}{2M_i}}}{\int \mathrm{d}\mathbf{p}\, e^{-\beta \frac{\mathbf{p}_i^2}{2M_i}}} \cdot \frac{e^{-\beta\, V(\mathbf{q})}}{\int \mathrm{d}\mathbf{q}\, e^{-\beta\, V(\mathbf{q})}} \tag{1.16}$$

The main difficulty is to determine the configurational part $\mathscr{P}(\mathbf{q}) = e^{-\beta\, V(\mathbf{q})} / \int \mathrm{d}\mathbf{q}\, e^{-\beta\, V(\mathbf{q})}$ which depends on the PES. Moreover, the expectation value of any configuration dependent property only depends on the particle coordinates $A(\boldsymbol{q})$ and can be expressed as an integral over the probability distribution:

$$\langle A \rangle = \int \mathrm{d}\mathbf{q}\, A(\mathbf{q})\, \mathscr{P}(\mathbf{q}) = \frac{\int \mathrm{d}\mathbf{q}\, A(\mathbf{q})\, e^{-\beta\, V(\mathbf{q})}}{\int \mathrm{d}\mathbf{q}\, e^{-\beta\, V(\mathbf{q})}} \tag{1.17}$$

To calculate equilibrium observables which are ensemble averages, different approaches can be used such as the Monte Carlo (MC) method. In this thesis we focus on the Molecular Dynamics (MD) family of methods. By making the assumption that our system is ergodic (given an infinite amount of time, any MD will be able to visit all configurations N on the PES), microcanonical phase space averages can be replaced by time averages over the trajectory:

$$\langle A \rangle = \lim_{t\to\infty} \frac{1}{T} \int_0^T A(\mathbf{q})\, \mathrm{d}t = \int \mathrm{d}\mathbf{q}\, \mathscr{P}(\mathbf{q})\, A(\mathbf{q}) \equiv A \tag{1.18}$$

1.2.2 Integration Algorithms using Finite Differences

MD simulations are performed by solving the classical equations of motion and allow to know the trajectory of the particles in the system. Unfortunately, equation (1.12) cannot be integrated analytically except for simple problems. So, one may find a solution by using approximated schemes to evolve along a MD trajectory. To do so, it requires the use of an integrator which will generate phase space vectors at discrete times that are multiples of the time step of discretization Δt. Then, using the initial conditions, MD will generate phase space vectors $\mathbf{x}_{n\Delta t}$ where n=0,...,N_{step} are generated by applying the integration algorithm iteratively and $\mathbf{x}$ is the 6N-1 dimensional phase space vector of the system. Thus, the ensemble average of a property is related to the time step as:

$$\langle A \rangle = \lim_{N_{step}\to\infty} \frac{1}{N_{step}} \sum_{n=1}^{N_{step}} A(\mathbf{x}_{n\Delta t}) \equiv A \tag{1.19}$$

Due to the extreme sensitivity of the trajectory to the initial conditions, it is important to have an integrator with good stability and accuracy. One of the simplest integrator uses Taylor expansion where the position of a particle at time $t + \Delta t$ and $t - \Delta t$ are expressed in terms of position, velocity and acceleration at time t according to:

$$\begin{aligned} \mathbf{r}_i(t+\Delta t) &\approx \mathbf{r}_i(t) + \Delta t\, \dot{\mathbf{r}}_i(t) + \frac{1}{2}\Delta t^2\, \ddot{\mathbf{r}}_i(t) + \frac{1}{3!}\Delta t^3\, \dddot{\mathbf{r}}_i(t) + \mathcal{O}(\Delta t^4) \\ \mathbf{r}_i(t-\Delta t) &\approx \mathbf{r}_i(t) - \Delta t\, \dot{\mathbf{r}}_i(t) + \frac{1}{2}\Delta t^2\, \ddot{\mathbf{r}}_i(t) - \frac{1}{3!}\Delta t^3\, \dddot{\mathbf{r}}_i(t) + \mathcal{O}(\Delta t^4) \end{aligned} \tag{1.20}$$

By adding both two previous equations and using Newton's second law one may find:

$$\mathbf{r}_i(t+\Delta t) + \mathbf{r}_i(t-\Delta t) = 2\mathbf{r}_i(t) + \frac{\Delta t^2}{2m_i}\mathbf{F}_i(t) \tag{1.21}$$

where equation (1.21) is known as the Verlet algorithm which does not involve velocities [48]. If needed, it is possible to compute velocities at any point in the trajectory via the centered difference formula [49]:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{2\Delta t} \tag{1.22}$$

The Verlet algorithm is subjected to error in the order of $\mathcal{O}(\Delta t^4)$ for positions whereas the error is in the order of $\mathcal{O}(\Delta t^2)$ for velocities. There is also the velocity Verlet algorithm which as an error in the order of $\mathcal{O}(\Delta t^3)$ for both velocities and positions. Other algorithms are available such as the leap-frog one which allows to propagate positions coupled to velocities [50].

1.2.3 The Need of Other Ensembles

Since the equations of motion conserve the Hamiltonian *H(q,p)*, a trajectory computed using equation (1.12) will generate configuration belonging to a microcanonical ensemble with fixed energy E (NVE). A disadvantage of this ensemble is that conditions of constant total energy E are not those under which experiments are performed. Even though it does conserve the canonical ensemble, and then a collection of trajectories with the Boltzmann distribution would yield correct averages, a single trajectory does not allow for fluctuations of the Hamiltonian H. However, it is possible to modify equation (1.12) to allow for energy fluctuations and to obtain an ergodic sampling of the ensemble. This is referred as thermostats. Consequently, the particle number N, the volume V and the temperature T are control variables in simulations in the canonical ensemble, also referred to as NVT. It characterizes a system in thermal equilibrium with a heat bath where the energy of endothermic and exothermic processes are exchanged with a thermostat. The easiest way to drive the system toward a desired temperature T_0 is to scale all the velocities by a factor λ. The factor is chosen such that equipartition of energy holds yielding an average kinetic energy of $3N/2k_BT$ with N the total number of particles. The difference between the reference temperature T_0 and the instantaneous temperature $T(t)$ is given by:

$$\begin{aligned} T_0 - T(t) &= \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i\lambda^2 \left|\dot{\mathbf{r}}_i\right|^2 - \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i \left|\dot{\mathbf{r}}_i\right|^2 \\ &= (\lambda^2 - 1)T(t) \end{aligned} \tag{1.23}$$

which gives $\lambda = \sqrt{\frac{T_0}{T(t)}}$. Scaling the velocities by λ is referred as the isokinetic thermostat.

It is also possible to introduce a parameter τ to modulate the strength of the coupling between the system and the thermal bath of temperature T_0 by defining the temperature change occurring in a time step Δt as:

$$\begin{aligned} \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{T_0 - T(t)}{\tau} \\ \Delta T &= \frac{\Delta t}{\tau}(T_0 - T(t)) \end{aligned} \tag{1.24}$$

Thus, it is possible to express λ as:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_0}{T(t)} - 1\right)} \tag{1.25}$$

which is referred as the Berendsen thermostat [51], but other thermostats following the idea of velocity rescaling exist such as Nosé-Hoover [52]. Although these thermostats reach the desired temperature quickly, they are not capable to reproduce the correct canonical fluctuations of the instantaneous T [53]. Hence, they are mainly used for equilibration.

With the temperature, the pressure is also an important parameter to be controlled in MD simulation. It means that a barostat has to be added where the pressure is controlled by scaling the system volume. Following the ideas of velocity rescaling and assuming the simulation box being cubic, the rescaling factor

λ_P can be expressed as:

$$\lambda_P = \sqrt[3]{1 + \frac{\beta \Delta t}{\tau_p} \left(P_o - P(t)\right)} \tag{1.26}$$

where β is the compressibility factor and τ_p is analogous to τ. Equation (1.26) is referred as the Berendsen barostat [51]. By controlling the temperature and the pressure within the dynamics, the results obtained belongs to the NPT ensemble and are closer to the laboratory conditions. Because the global thermostat variables are coupled and control directly only global system quantities, they rely on the efficient energy transfer within the system to achieve equipartition within the canonical distribution. Therefore, in a system where the energy transfer between its different parts is different, the thermostat may have difficulties to maintain the same temperature for the different parts of the system.

Stochastic processes, such as Langevin-type equations, can also be used as thermostats. They provide a direct control of equipartitioning with their independent thermalization of each degree of freedom and hence sample correctly the canonical ensemble, which is not the case of velocity rescaling. The Langevin equation was initially developed in 1908 [54] to model the Brownian motion. It introduces two additional forces to describe the effect of the fluid on the particle: a friction force γ and a random force $\mathbf{R}(t)$ such as it is possible to re-write equation (1.12) as:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{q}}{\mathrm{d}t} &= \frac{\mathbf{p}}{m} \\ \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} &= -\nabla V(\mathbf{q}) - \gamma\, \mathbf{p} + \mathbf{R}(t) \end{aligned} \tag{1.27}$$

with $\mathbf{R}(t)$ a stochastic and stationary force described by a Gaussian distribution (technically the time derivative of a Wiener process). It has the following properties:

- It is a stationary process.
- Its mean value is zero

$$\langle \mathbf{R}(t) \rangle = 0 \tag{1.28}$$

- Its time-correlation τ is infinitely short-ranged:

$$\langle \mathbf{R}_i(t) \mathbf{R}_j(t+\tau) \rangle = 2 m k_B T \gamma \delta(\tau) \delta_{ij} \tag{1.29}$$

  where for two different times t and $t+\tau$, $\mathbf{R}_i(t)$ and $\mathbf{R}_j(t+\tau)$ are uncorrelated and T is the target temperature.

The magnitude of the random force and the friction are related to guarantee the Fluctuation-Dissipation Theorem (FDT), hence the NVT statistics. So, the Hamiltonian follows a stochastic thermalization, which acts locally (each degree of freedom is thermalized through its own noise). In this thesis, only the Langevin thermostat was used to sample the NVT ensemble.

1.2.4 Advanced Integrators

The different integrators presented in Section 1.2.2 are not exact. The energy of the exact Hamiltonian H is not conserved throughout the numerical trajectory. Moreover, certain formal properties of the Hamiltonian H should be preserved by numerical integration methods. Thus, it is important to develop a formal structure that allows numerical solvers to be generated more rigorously. The framework is based on the phase space integrator defined as:

$$i\mathcal{L}a = \{a, H\} = \frac{\mathrm{d}a}{\mathrm{d}t} \tag{1.30}$$

where the curly bracket is the Poisson bracket. It introduces an operator $\mathcal{L}$, called the Liouville operator on the phase space. Equation (1.30) can also be written as a differential operator:

$$i\mathcal{L} = \sum_{\alpha=1}^{3N} \left[\frac{\partial H}{\partial p_\alpha} \frac{\partial}{\partial q_\alpha} - \frac{\partial H}{\partial q_\alpha} \frac{\partial}{\partial p_\alpha} \right] \tag{1.31}$$

where α runs over all the $3N$ degrees of freedom of the system (with N the number of atoms). Furthermore, the equation $\frac{\mathrm{d}a}{\mathrm{d}t} = i\mathcal{L}a$ can be solved formally for any function $a(x_t)$ to have a formal solution to Hamilton's equation:

$$a(x_t) = e^{i\mathcal{L}t} a(x_0) \tag{1.32}$$

where $e^{i\mathcal{L}t}$ is the classical propagator. As equation (1.31) suggests, the Liouville operator can be written as a sum of two terms $\mathcal{L}_1$ and $\mathcal{L}_2$:

$$\begin{aligned} i\mathcal{L}_1 &= \sum_{\alpha=1}^{3N} \frac{\partial H}{\partial p_\alpha} \frac{\partial}{\partial q_\alpha} \\ i\mathcal{L}_2 &= -\sum_{\alpha=1}^{3N} \frac{\partial H}{\partial q_\alpha} \frac{\partial}{\partial p_\alpha} \end{aligned} \tag{1.33}$$

where $i\mathcal{L}_1$ corresponds to the kinetic energy part and $i\mathcal{L}_2$ to the forces. However, these two operators do not commute and the classical propagator $e^{i\mathcal{L}t} = e^{(i\mathcal{L}_1 + i\mathcal{L}_2)t}$ cannot be separated into a simple product $e^{i\mathcal{L}_1 t} e^{i\mathcal{L}_2 t}$. One thus need an approximation of the classical propagator known as the Trotter theorem which states that given two operators A and B for which $[A, B] \neq 0$, then:

$$e^{A+B} = \lim_{P \to \infty} \left[e^{B/2P} e^{A/P} e^{B/2P} \right]^P \tag{1.34}$$

where P is an integer. By defining $\Delta t = t/P$ and taking $1/P$ power of both sides:

$$e^{i\mathcal{L}\Delta t} = e^{i\mathcal{L}_2 \Delta t/2} e^{i\mathcal{L}_1 \Delta t} e^{i\mathcal{L}_2 \Delta t/2} + \mathcal{O}(\Delta t^3) \tag{1.35}$$

which is the velocity Verlet integrator.

One of the most ubiquitous aspects of complex systems in classical mechanics is the presence of forces that generate motion with different time scales. The forces derived from the potential will have large and rapidly varying components due to the intramolecular terms and smaller, slowly varying components due to

long-range contribution to the non-bonded interactions. Hence, the derivative of the non-bonded terms are computed as often we compute the bond-term forces, as it is done in the velocity Verlet scheme. However, an ideal integrator will juggle between at least two time step lengths, one for high frequency and the other for the low-frequency energy terms. Thus, $\mathcal{L}_2$ can be defined as:

$$\begin{aligned} i\mathcal{L}_2 &= i\mathcal{L}_{slow} + i\mathcal{L}_{fast} \\ &= \sum_{j=1}^{N} \left[\mathbf{F}_{j,slow}(\mathbf{r}) + \mathbf{F}_{j,fast}(\mathbf{r})\right] \frac{\partial}{\partial \mathbf{p}_j} \end{aligned} \tag{1.36}$$

Then, it is possible to redefine two operators such as:

$$\begin{aligned} i\mathcal{L} &= i\mathcal{L}_{slow} + i\mathcal{L}^{'}_{fast} \\ i\mathcal{L}_{slow} &= \sum_{j=1}^{N} \left[\mathbf{F}_{j,slow}(\mathbf{r}) \frac{\partial}{\partial \mathbf{p}_j}\right] \\ i\mathcal{L}^{'}_{fast} &= \sum_{j=1}^{N} \left[\mathbf{F}_{j,fast}(\mathbf{r}) \frac{\partial}{\partial \mathbf{p}_j} + \frac{\mathbf{p}_j}{m} \frac{\partial}{\partial \mathbf{r}_j}\right] \end{aligned} \tag{1.37}$$

Applying the Trotter schemes gives:

$$e^{i\mathcal{L}\Delta t} \approx e^{i\mathcal{L}_{slow}\Delta t/2} e^{i\mathcal{L}^{'}_{fast}\Delta t} e^{i\mathcal{L}_{slow}\Delta t/2} \tag{1.38}$$

It is possible to apply the Trotter theorem to the $e^{i\mathcal{L}^{'}_{fast}\Delta t}$ term. If one defines a smaller time step $\delta t = \Delta t/n$, then:

$$e^{i\mathcal{L}^{'}_{fast}\Delta t} = \left(e^{i\mathcal{L}_{fast}\delta t/2} e^{i\mathcal{L}_1 \delta t} e^{i\mathcal{L}_{fast}\delta t/2}\right)^n \tag{1.39}$$

By using equation (1.38) and (1.39) one may find:

$$e^{i\mathcal{L}\Delta t} \simeq e^{i\mathcal{L}_{slow}\Delta t/2} \left(e^{i\mathcal{L}_{fast}\delta t/2}\, e^{i\mathcal{L}_1 \delta t}\, e^{i\mathcal{L}_{fast}\delta t/2}\right)^n e^{i\mathcal{L}_{slow}\Delta t/2} \tag{1.40}$$

This leads to the so-called REference System Propagator Algorithm (RESPA) introduced by Tuckerman et al [55] where the time step Δt is chosen according to the time scale of the slow forces and δt is chosen according to the natural time scale of $\mathbf{F}_{fast}$.

When using the Langevin equations of motion (1.27), a slightly different splitting scheme is introduced to integrate the friction and random forces. By applying the phase propagator and by splitting equation (1.27) into three exactly solvable parts, one may finds [56, 57]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} q \\ p \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ -\nabla V \end{pmatrix}}_{\mathrm{B}} + \underbrace{\begin{pmatrix} p/m \\ 0 \end{pmatrix}}_{\mathrm{A}} + \underbrace{\begin{pmatrix} 0 \\ -\gamma p + \mathrm{R}(t) \end{pmatrix}}_{\mathrm{O}} \tag{1.41}$$

The three parts may be solved numerically within a time step Δt. The A part corresponds to the evolution

of the position due to the momentum. B is the evolution of the momentum due to the conservative forces. The O step has its velocity propagated under the action of the friction and the random forces. The most simple integration scheme would be to solve A first with a time step Δt, then B to finally end with the O part. This lead to the so-called ABO method. However, the splitting is asymmetric causing unstability in the averaged phase space quantities [58, 59]. To circumvent that, symmetric splitting have been proposed, such as BAOAB, with five time steps of three different types (A, B and O) which can be applied to the phase space vector. The O part is a standard Ornstein-Uhlenbeck process which can be integrated analytically [59]:

$$p(t+\Delta t) = e^{-\gamma \Delta t} p(t) + \sqrt{\frac{m}{\beta}\left(1 - e^{-2\gamma \Delta t}\right)}\,\mathrm{r} \tag{1.42}$$

The random number r usually consists in a Gaussian white noise, which simulates the thermal fluctuations of the system. Using this equation, it is possible to decompose the BAOAB algorithm such as:

$$\begin{aligned}
\mathrm{B} &: v\left(t + \frac{\Delta t}{2}\right) = v(t) - \frac{\Delta t}{2}\frac{\nabla V}{m}(q(t)) \\
\mathrm{A} &: q\left(t + \frac{\Delta t}{2}\right) = q(t) + \frac{\Delta t}{2} v\left(t + \frac{\Delta t}{2}\right) \\
\mathrm{O} &: \tilde{v}\left(t + \frac{\Delta t}{2}\right) = e^{-\gamma \Delta t} v\left(t + \frac{\Delta t}{2}\right) + \sqrt{\frac{m}{\beta}\left(1 - e^{-2\gamma \Delta t}\right)}\,\mathrm{r} \\
\mathrm{A} &: q\left(t + \Delta t\right) = x\left(t + \frac{\Delta t}{2}\right) + \Delta t \tilde{v}\left(t + \frac{\Delta t}{2}\right) \\
\mathrm{B} &: v\left(t + \Delta t\right) = \tilde{v}\left(t + \frac{\Delta t}{2}\right) - \frac{\mathrm{d}t}{2}\frac{\nabla V}{m}(q(t + \Delta t))
\end{aligned} \tag{1.43}$$

The O part uses the previous velocity v calculated in the first step B. Hence, the velocity obtained with step O is referred as $\tilde{v}$. During this thesis, this was the integrator used.

1.2.5 Force Fields

The production of reliable trajectories and thus observables computed as time-averaged quantities strongly depends on the accuracy of the PES from which the forces are computed. In chemistry and biology, the functional form enabling to compute thes PES are referred as Force Field (FF) or as sometime interatomic potential in physics. The general equation associated with a FF can be partitioned as [60]:

$$E_{total} = E_{intramolecular} + E_{intermolecular} \tag{1.44}$$

The intramolecular and intermolecular terms, also called respectively bonded and non bonded terms, can be decomposed as:

$$\begin{aligned}
E_{intramolecular} &= E_{bond} + E_{angle} + E_{torsion} + E_{improper} \\
E_{intermolecular} &= E_{elec} + E_{vdW}
\end{aligned} \tag{1.45}$$

The Intramolecular Energy Terms

The intramolecular energy terms consist in bond stretching, angle bending and torsion energy terms and are usually expressed as:

$$\begin{aligned} E_{stretching}(r) &= \sum_{bonds} k_s(l_{ij} - l_0)^2 \\ E_{bending}(\theta) &= \sum_{angles} k_b(\theta_{ijk} - \theta_0)^2 \\ E_{torsions}(\phi) &= \sum_{torsions} V_n[\cos(n\phi - \phi_0)] \end{aligned} \tag{1.46}$$

The bond stretching and angle bending are usually represented by a harmonic potential where k_s and k_b are stiffness constant. l_{ij} is the bond length deformation (stretching or compression). l_0 and θ_0 are respectively the equilibrium bond length and valence angle. The harmonic approximation is usually reasonable for the bond stretching and angle bending since they usually involve small displacements from the equilibrium. Anharmonicity can be taken into account by adding higher terms which is usually preferred to Morse-like potentials due to computational reasons [61]. Sometimes, other terms are added to optimize the fitting to the vibrational spectra such as the Urey-Bradley potential:

$$E_{UB} = \sum_{angles} k_{UB}(s_{ijk} - s_0)^2 \tag{1.47}$$

with s the distance between two external atoms forming the angle. Although bond stretching and angle bending terms involve small displacements from the equilibrium, this is not the case for torsion because the PES is rather flat and larger displacements from the equilibrium usually take place. This lead to the use of truncated Fourier series where ϕ is the torsional angle and V_n determines the height of the potential barrier with n the number of minima or maxima between 0 and 2π [62]. To ensure the planarity of some particular groups, an improper torsion term (out of plane motions) is added and defined as

$$E_{improper}(\omega) = \sum_{impropers} k_{im}(\omega - \omega_0)^2 \tag{1.48}$$

where k_{im} is the improper torsion constant and ω the improper dihedral angle which deviated from planarity. ω_0 is the equilibrium improper dihedral angle between four atoms *i*, *j*, *k* and *l*. The different components of the intramolcular energy terms are represented with the non-bonded ones in Figure 1.1.

Intermolecular Energies Terms

The intermolecular energy terms consist of electrostatics, dispersion and short-range repulsion interaction where the latter two are often referred as the so-called van der Waals term (vdW). The vdW interaction arises from the balance between repulsive and attractive forces where the repulsive one is due to the overlap of the electrons clouds (Pauli repulsion) and the attractive one is linked to London's dispersion. In most classical FF, like CHARMM [63], AMBER [64], OPLS [65], the Lennard-Jones potential is generally used to

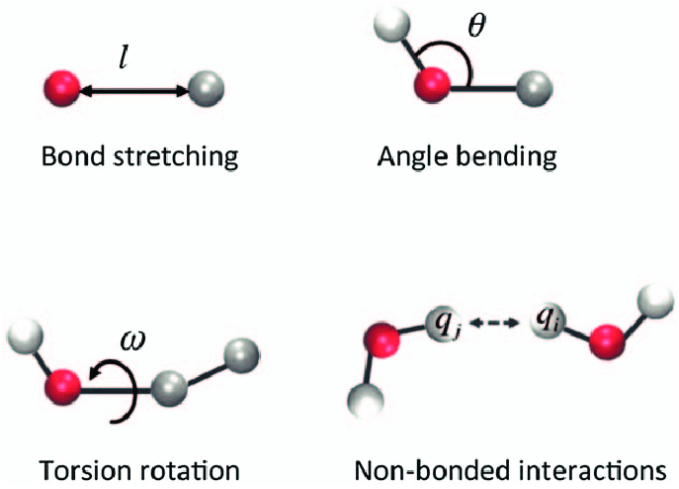


Figure 1.1: Schematic representation of the different intramolecular energy terms (bond stretching, angle bending and torsional rotation) with the non-bonded interactions (vdW and electrostatics).

describe these interactions where the attractive parts varies as $1/r^6$ and the repulsive part as $1/r^{12}$:

$$E_{vdW} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{1.49}$$

with ε_{ij} the depth of the potential (or more physically, the highest attraction energy between atom i and j) and σ_{ij} the distance at which particle-particle potential energy is zero (Figure 1.2). Improvement in the vdW form can be made by including higher terms or by modeling the short-range repulsion by more physically grounded expressions [66, 67].

To this potential is added the electrostatic component defined as:

$$E_{elec} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{1.50}$$

with ε_0 is the dielectric constant. To model this electrostatic attraction between different polarized parts of molecules (due to the asymmetric distribution of electrons), partial charges are assigned to each atom and the interaction is calculated using Coulomb's law by summing over atom pairs.

The choice of the different parameters is very important due to their impact on the accuracy of the FF. Furthermore, it can be really challenging due to the very large number of parameters involved. Torsional, k_0 and r_0 parameters are usually obtained either from *ab initio* or by fitting to experimental data such as neutron, X-ray and electron diffraction, Raman and neutron spectroscopy. Moreover, neutron scattering deals with the same properties obtained in MD simulations and the spatial/time scale which are measured through experiments. The partial charges can be derivated from experimental data or by *ab initio* and deriving from quantum mechanical potential. Coulomb's law is then used to compute their contribution to the total energy.

Simulations of relevant systems require the introduction of empirically parametrized PES referred as FF. One of the first FF was developed in the 1960's to predict molecular structures, vibrational spectra and enthalpies of isolated molecules. In the beginning, FF were developed to treat small organic molecules [68, 69]. More widely applicable FF were developed to deal with more complex systems such as the previous

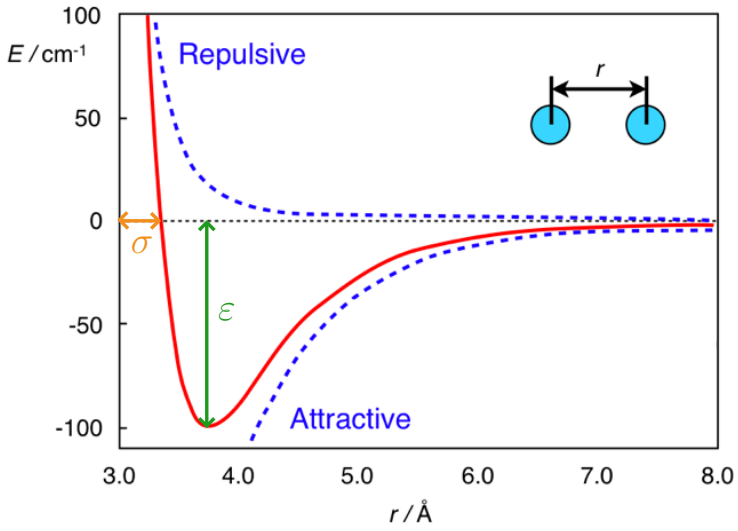


Figure 1.2: Schematic representation of the Lennard-Jones potential: the blue dotted lines are the attractive and repulsive energy parts of two particles. The plain red line is the total LJ intermolecular energy where the depth of the potential ε and the distance at which particle-particle potential energy is zero σ are represented in green and orange respectively.

cited which provided over the years very extensive results on many fields of applications ranging from biology to ionic liquid [70]. However, the majority of FFs describe molecular electrostatics by fixed point atomic point charges place at atomic center interacting via Coulomb's law giving an inexpensive description of the PES. Hence, these descriptions do not take into account the electron mobility which corresponds to a response of the system to a change of its environment. Thus, a many-body polarization term must be added.

1.2.6 Polarizable Force Fields

In early developments, polarization was not taken explicitly into account. Hence, much effort has been made to explicitly include many-body polarization in new FFs. This has been made by two main approaches. The first one introduces polarization by allowing atomic charge to fluctuate as response of a change in the electrostatic potential according to the electronegativity equalization principle. This is the so called fluctuating charge method (FQ). The second approach introduces polarization by atomic dipole as a response of an electric field. Two variations of this approach can be identified depending of the source of the induced dipole. In the induced point dipole model (PD), the induced dipole are introduced at the dipole level by adding atom-centered dipole polarizabilities. Polarization can be either included through Drude oscillators (DO) where induced dipole are added through an additional charge virtual particle connected to the nuclei by a spring.

The Fluctuating Charge Model

The FQ model is based on a redistribution of the atomic partial charges to recreate the fluctuation of the electronic density. Hence, the electrostatic energy E_{elec} is expressed as a second order Taylor expansion in terms of atomic point charges $\{Q_i\}$ such as:

$$E_{elec} = E_0 + \mathbf{Q}^\dagger \boldsymbol{\chi} + \frac{1}{2}\mathbf{Q}^\dagger \boldsymbol{\eta} \mathbf{Q} \tag{1.51}$$

where χ_i and η_{ij} are respectively the atomic electronegativities and hardness matrix elements. E_0 is the reference energy term which is usually not considered [71]. The equilibrium charges are obtained by minimizing equation (1.51) where global Lagrange multiplier was added to ensure that the total charge of the system is conserved. Atomic electronegativities and diagonal hardness elements are free parameters while out off diagonal hardness elements are the interaction between charges which are usually given by Coulomb interaction. However, at short distances, the Coulomb interaction must be screened, thus many other functional forms have been proposed [72]. Because charge conservation is introduced through a Lagrange multiplier, unphysical effect arise where coupling between atoms appears regardless of their distances which induces long-range charge transfer that should not be observed [73, 74]. To solve this problem, some groups have reformulated the FQ model in terms of split charge variables where the charge is transferred between atoms pairs, the so-called Split Charge Equilibration (SCQE) [75, 76]. It is used in FF such as CHARMM-FQ [77, 78].

The Drude Oscillators

In the DO model, the atomic partial charge is partitioned into a nuclear and a fictitious mass-less components, called Drude particles, connected to the nuclei via harmonic springs [79]. Fluctuation of the electrostatic environment will then have direct repercussions on the point charge's dynamic, as a charge moving in an external electric field. It can expressed as:

$$E_{drude} = \sum_{i=1}^{N} \frac{1}{2} k_D (r_d(i) - r_i)^2 + \sum_{\substack{i=0,N \\ i,N_D}}^{i \neq j} \frac{q_D(j) q_i}{|r_D(j) - r(i)|} + \sum_{\substack{i,j=1 \\ i<j}}^{i \neq j} \frac{q_D(j) - q_D(i)}{|r_D(j) - r_D(i)|} \tag{1.52}$$

where $r_D(i)$, q_D and k_D are the position of the Drude particle attached to atom i, its partial charge and the stiffness constant of the spring respectively. Equation (1.52) has to be minimized at each time step to find the correct position of Drude's particles which can be computationally expensive. To reduce this high cost, extended Lagrangian techniques is usually used where the mass of each atom is partitioned between the fictitious particle and the parent atom to which it is attached. However, the mass attached to the fictitious particle should be chosen wisely. If it is too small, high-frequency motions will appear, then a very small time step Δt should be used to be correctly simulated. On the other hand, if it is too heavy, the response of the Drude particles will not be fast enough compared to the evolution of the nuclei, which is contradictory with the BO approximation. Moreover within the extended Lagrangian scheme, it is not possible for more advanced multi-time steps integration techniques, due to the alteration of the dynamical properties. Hence, it has made polarizable MD simulations based on PD model more effective [80].

The Polarizable Point Dipole Model

In the Polarizable point Dipole (PD) model, polarization is introduced at the dipole level where the electric field perceived at an atomic position will drive the polarization of its electronic densities as [81]:

$$E_{elec} = E_0 + \mathbf{F}^{\dagger} \boldsymbol{\mu} + \frac{1}{2} \boldsymbol{\mu}^{\dagger} \mathbf{T} \boldsymbol{\mu} \tag{1.53}$$

with E_0 the interaction energy of the permanent dipole. $\boldsymbol{\mu}$ is a $3N$ vector with all atomic induced dipole moment components. $\mathbf{F}$ is the electric field components arising from the permanent dipole and $\mathbf{T}$ is the

dipole-dipole interaction matrix defined as:

$$\mathbf{T}_{ii} = \boldsymbol{\alpha}_i^{-1} \tag{1.54}$$

$\boldsymbol{\alpha}_i^{-1}$ is the inverse atomic polarizabilities tensor of atom i. Therefore, derivating equation (1.53) with respect to the induced dipole one finds:

$$\begin{aligned} \frac{\partial E_{elec}}{\partial \boldsymbol{\mu}} =& \mathbf{T}\boldsymbol{\mu} + \mathbf{F} = 0 \\ \boldsymbol{\mu} =& -\mathbf{T}^{-1}\mathbf{F} \end{aligned} \tag{1.55}$$

The PD model yields many advantages compared to the DO and FQ ones. It has a higher flexibility in terms of time integration whereas the DO model was limited. It also allows to directly use experimental or *ab initio* data thanks to the explicit presence of polarizabilities $\boldsymbol{\alpha}_i$ while the DO requires a non-trivial balancing between k_D and q_d to reproduce correct polarizabilites. Moreover, because the atomic charges are purely fixed parameters, there is no risk of non-physical charge transfer compared to the FQ model. Furthermore, it has been shown that charge polarization are generally large than dipole ones and that polarization should be included from scratch while developing a FF instead to be added to already existing FFs [82]. The induced dipole model have proven to be best suited for polarizable systems such as ionic liquid than Drude oscillators yielding a better accuracy [83–85]. The AMOEBA FF is a well-known polarizable FF based on PD model where the set of permanent dipole are introduced up to quadrupoles [32]. Other models also exist which combined fluctuating charges with polarizable dipoles [86, 87].

One may define the molecular polarizability as the inverse of the polarization matrix restricted to the atoms within a single molecules. Then, it is related to the molecular induced dipole moment such as $\boldsymbol{\mu} = \boldsymbol{\mu}_A + \boldsymbol{\mu}_B$ for a diatomic system. However, Applequist et al. showed that the molecular polarizabilities could diverge when atoms are close [88]. This is known as the polarization catastrophe. To circumvent that, they first choose lower polarizability parameters to reduce the molecular polarizabilities amplitudes. Thole et al. proposed a more general solution by using a damping function to compensate the divergence at short distances and hence effectively avoiding the catastrophe in simulations [89]. Therefore, the Thole damping scheme is widely use in MD and implemented in many FF.

1.2.7 Periodic Boundary Conditions and Particle Mesh Ewald

In most cases we are interested in the bulk properties of a liquid or solid system. Hence, some periodic boundaries conditions (PBC) need to be imposed. Thus, the simulated box is surrounded by an infinite number of replicas of itself. Only the number of atoms inside the main cell are considered explicitly but as soon as the atoms leaves the cell an image particle enters from the opposite side of the box to replace it as illustrated in Figure 1.3. However, electrostatic interaction are slowly decaying (r^{-1}), so interaction arising from many of the periodic image should be included. Hence, our system box has an infinite numbers of neighbours which makes the non-bonded interaction very cumbersome to compute. A distance cutoff is introduced to limit the interaction ranges, thus, the complexity of computing pairwise interaction with PBC and a cutoff scale to $\mathcal{O}(N^2)$.

However, electrostatic interactions have different scales. Electrostatics encompasses a part which is short-ranged and that should be treated in the direct space and a long-range contribution to the Coulomb interaction

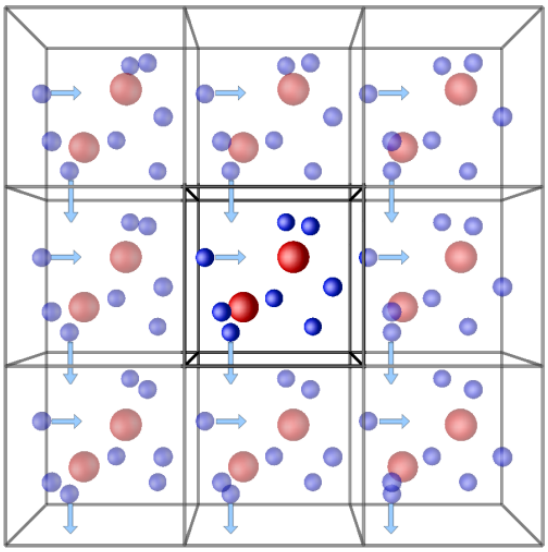


Figure 1.3: Schematic representation of the periodic Boundary Condition: when a particle goes out of the simulation box, an another one is reintroduced on the opposite side [90].

which needs to be handled separately in the Fourier space [85, 91]:

$$\begin{aligned} E_{elec} =& E_{direct} + E_{recip} + E_{self} \\ E_{direct} =& \sum_{i=1}^{N} \sum_{j=i+1}^{N} q_i q_j \frac{erfc(\beta\, r_{ij})}{r_{ij}} \\ E_{recip} =& \sum_{i=1}^{N} \sum_{j=i+1}^{N} q_i q_j\, \Phi_{rec}(\mathbf{r}_{ij}, \beta) \\ E_{self} =& \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \end{aligned} \tag{1.56}$$

where *erfc* is a complementary error factor defined as $2/\sqrt{\pi}\ \int_x^\infty e^{-t^2}\ \mathrm{d}t$ which allows to smooth the switching off of the direct contribution as a function of the distance r_{ij}. The $\Phi_{rec}(\mathbf{r}_{ij}, \beta)$ potential is defined as:

$$\Phi_{rec}(\mathbf{r}_{ij}, \beta) = \frac{1}{\pi V} \sum_{m \neq 0} \frac{e^{\frac{-\pi^2\, \mathbf{m}^2}{\beta^2}}}{\mathbf{m}^2} e^{2\pi i \mathbf{m} \cdot \mathbf{r}} \tag{1.57}$$

with $\mathbf{m}$ the linear combination of the reciprocal lattice vectors. β is a parameter which balances between the ratio between E_{direct} and E_{recip} terms [91]. Hence, Ewald summation scales as $\mathcal{O}(N^2)$ although a careful choice of the parameter β allow to reduce the scaling to $\mathcal{O}(N^{3/2})$ which is called as smooth particle mesh Ewald (SPME) [92].

Darden proposed a method to speed up the computation times of equation (1.57) called PME where the complex exponential are interpolated on a grid which allow to write Φ_{rec} as a convolution product which becomes a simple product in the Fourier space, allowing the algorithm to scale as $\mathcal{O}(N\ \log\ N)$ [93].

1.3 Conclusion

Molecular Dynamics is a powerful tool which allows a numerical 'thought experiment' to be carried out on computers using a model. Hence, it gives access to simulations at extreme conditions, where temperature and pressure that could be difficult to create in a physical laboratory. One of the earliest example was done in 1955 by Fermi, Pasta and Ulam to quantify the degree of ergodicity and energy equipartitioning in a system made of nonlinear oscillators. It was followed by the work of Alder and Wainwright to show that a solid-liquid phase transition exists [44] and by Stillinger and Rahman who carried out the first MD simulations of liquid water [94, 95]. The evolution of the field of MD has benefited by the many advances in high performance computing and make it possible to carry them out on desktop computers nowadays. One of the challenge of MD field is the long-time dynamics which is involved in many processes such as protein folding and hence new algorithms and theoretical methods are developed throughout the years such as metadynamics [96] or hyper-MD [97]. There is also the question of different length scales, where some phenomenon, such as catalysis, require a two level description of the system that enclose atomistic level (where chemical bonds are broken and formed) to the macroscopic one with the same level of detail. This problem was addressed by Warshel and Levitt to study enzymatic reactions which lead to their sharing Nobel prize in 2013 [98]. After all, MD is a key scientific tool used in many fields to help in the interpretation of experimental results or suggesting new experiments.

Although really powerful to investigate and predict the properties of complex systems, MD has some limitations. One of them, it that MD relies on Newton's equation of motion. Hence, it treats nuclei from a classical point of view. However, the studied system should also obey the law of QM due to the wave-particle duality which make it computationally difficult reducing the size of the desired studied system. Thus, Nuclear Quantum Effects (NQEs), such as zero-point energy (ZPE) and tunneling, are neglected, although they could play an important role on the properties of the studied system. For example, at room temperature, the ZPE of a typical chemical bond of frequency ω ($\sim \hbar\omega/2$) exceeds the thermal energy scale ($\sim k_b T$) of that coordinate at ambient temperature T by an order of magnitude [99]. These effects can make large changes to the structure and dynamics from proton delocalization and tunneling in enzymes [100], to the phase diagram of high pressure melts [101]. For this reason, these effects are important and cannot be neglected. Moreover, the PES on which the nuclei evolves have become more accurate thanks to the many development in the QM fields. It allows researchers to go beyond the previous empirical surfaces fit to experimental data and to uses PES fit from electronic structure. Although these quantum mechanical PES are used, the different computed properties are not accurate enough for many systems since it has been shown that NQEs should be included in the simulation [102]. Therefore, lot of effort have been made throughout the years to include NQEs inside MD by developing various approaches and theoretical methods to efficiently simulate them. Thereby, the next chapter will be devoted on different methods to include NQEs inside a MD simulation.

Chapter 2

Theoretical framework for Nuclear Quantum Effects

Contents

2.1 Introduction

Since direct resolution of Schrödinger's equation is only possible for a few degrees of freedom, a lot of effort is being made to develop various approximate approaches to treat Nuclear Quantum Effects (NQEs), in complex molecular systems. In particular, methods based on Feynman's path integral formulation of quantum mechanics have been developed, such as path integral molecular dynamics (PIMD) and Path Integral Monte Carlo (PIMC) [23, 103]. These methods are based on an isomorphism between a quantum particle and a chain of classical beads, coupled by harmonic springs. It provides numerically exact estimates for thermal equilibrium observables when then number of beads is sufficiently large (32 to 64 beads for water at room temperature). Unfortunately, the number of beads to converge increases as the temperature decreases, making low temperature simulations computationally expensive. Moreover, dynamical properties are not directly accessible through the path integral formalism although it is exact for static equilibrium properties, even for highly anharmonic systems [46]. Several approximate methods have been proposed to deal with this issue such as Centroid Molecular Dynamics (CMD) [104] and Ring-Polymer Molecular Dynamic (RPMD) [105]. These methods are described in Section 2.2 and have proved accurate for a wide range of systems and properties [106, 107], although they still present some drawbacks, in particular their computational cost, that can reduce their range of applicability. An overview of the different methods to include NQEs in MD simulations can be found in the literature [22, 107–109].

To overcome this computational limitation, alternative methods based on a generalized Langevin equation have been proposed. [25, 26]. The one used during this Thesis and described in Section 2.3 is the Quantum Thermal Bath (QTB) that accounts for quantum statistics with a colored random noise. This noise simulates both thermal and quantum fluctuations. It is chosen such that its power spectrum matches the quantum Fluctuation-Dissipation Theorem (FDT) [110, 111], whereas in the classical Langevin dynamics, the random force is a white noise which fulfills the classical FDT. The QTB allows to treat particles as quantum particles within MD simulations at the same computational cost as standard MD. However, QTB has some drawbacks, in particular the ZPE leakage (ZPEL) where part of the energy of high frequency modes is transferred to low frequencies and leads to a wrong energy distribution [15, 112–116] Thus, an adaptive QTB method, called adaptive QTB (adQTB), has been proposed to compensate the leakage and to recover the correct energy distribution. Although not explicitly derivable from the quantum statistical formalism, the adQTB has shown good results in different systems [27] and it will be further investigated in this thesis via a systematic comparison with reference methods.

2.2 Path Integral Molecular Dynamics

In order to simplify notations, a one-dimensional system is considered in this section. All results can be extended to a complex system in three dimensions where q is replaced by the $3N$-dimensional vector $\mathbf{q}$ of all atomic positions.

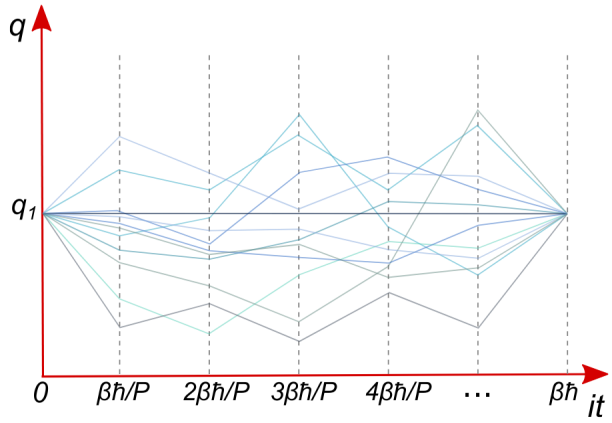


Figure 2.1: Schematic representation of all discretized path in imaginary time which contribute to the partition function (equation 2.6). Picture extracted from [47].

2.2.1 The partition function

Within Feynman's path integral formalism [103], it is possible to express the quantum mechanical partition function in the canonical ensemble as:

$$Z = \oint \mathcal{D}\left[q(\tau)\right] e^{-S[q(\tau)]} \tag{2.1}$$

where $\oint \mathcal{D}\left[q(\tau)\right]$ is a functional integral over all the possible closed paths in imaginary time τ. S is an action-like integral over the path which can be defined as:

$$S[q(\tau)] = -\frac{1}{\hbar}\int_0^{\beta\hbar} H\left[q(\tau)\right] \mathrm{d}\tau = -\frac{1}{\hbar}\int_0^{\beta\hbar} \left[V(q(\tau)) + \frac{m}{2}\dot{q}(\tau)^2\right] \mathrm{d}\tau \tag{2.2}$$

where the Hamiltonian H is written as the sum of kinetic and potential energy terms. For the paths which are statistically unlikely, the action S becomes large and their contribution to the partition function $\mathcal{Z}$ vanishes. The imaginary time τ can be discretized as $\tau = n\beta\hbar/P$ with $n = 0, \ldots, P-1$ and P an integer (Figure 2.1). To derive equation (2.1), the quantum partition function in the position representation is used:

$$\mathcal{Z} = \int \mathrm{d}q \left\langle q \left| e^{-\beta\hat{H}} \right| q \right\rangle \tag{2.3}$$

Because, the kinetic and potential operators, denoted $\hat{T}$ and $\hat{V}$ respectively, do not commute, the Trotter theorem defined in equation (1.34) has to be used:

$$e^{-\beta\hat{H}} = \left(e^{-\frac{\beta\hat{H}}{P}}\right)^P \approx \left(e^{-\beta_p\hat{V}/2} e^{-\beta_P\hat{T}} e^{-\beta_P\hat{V}/2}\right)^P + \mathcal{O}(\beta_P^2) \tag{2.4}$$

which becomes exact in the limit $P \longrightarrow \infty$. β_P is defined as $\beta_P = \beta/P$. Using P times the closure relation $\int \mathrm{d}q|q\rangle\langle q| = \mathrm{Id}$ and equation (2.4), it is possible to re-write equation (2.3) as:

$$Z = \lim_{P\to\infty} \int \mathrm{d}q_1, \ldots, \mathrm{d}q_P \prod_{j=1}^{P} \left[\left\langle q_{j+1} \left| e^{-\beta_P V(q_{j+1})/2} e^{-\beta_P\hat{T}} e^{-\beta_P V(q_j)/2} \right| q_j \right\rangle\right] \tag{2.5}$$

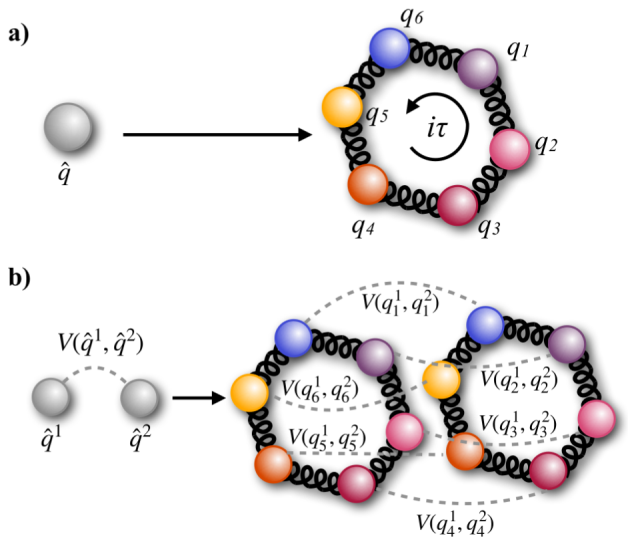


Figure 2.2: Schematic representation of classical and quantum particles within the path integral formalism. In (a), a single classical particle corresponds to one ring polymer with P=6. In (b), two classical particles are represented. They are interacting through a potential $V(q^{(1)}, q^{(2)})$ which corresponds to two ring polymers in the path integral formalism. Beads with the same index (represented with the same color in the sketch) interact through the potential V with their corresponding bead in the second ring polymer. Picture extracted from [47].

where $q_1 = q_{P+1}$. It can be understood as a product over several high temperature density matrices. Since the terms involving the potential energy are simple scalar values, it is possible to evaluate the necessary matrix elements, which are a series of terms corresponding to off-diagonal elements of the exponential of the kinetic energy operator. Using the internal product of position and momenta and introducing the ring polymer frequency as $\omega_P = 1/\beta_P\hbar$, one finally obtains the discrete path integral partition function as:

$$\mathcal{Z} = \lim_{P\to\infty} \left(\frac{1}{2\pi\hbar}\right)^P \left(\frac{2\pi m}{\beta_P}\right)^{P/2} \int \mathrm{d}q_1, \ldots, \mathrm{d}q_P \, e^{-\beta_P \sum_{j=1}^{P} \left\{ \frac{m\omega_P^2}{2}(q_{j+1}-q_j)^2 + V(q_j) \right\}} \tag{2.6}$$

It can be noticed that equation (2.6) does not depend on any operators and then corresponds to the classical partition function of a cyclic polymer composed of P atoms connected by harmonic springs with spring constant $k_P = m\omega_P^2$. Therefore, there is an isomorphism between the quantum system and a cyclic chain of P classical particles. In this chain, each particle is called a bead and corresponds to a different imaginary time slice. The isomorphism is in principle exact only when the number of beads P is infinite. The path integral formalism samples the potential in an extended ring polymer phase space at a temperature $P \times T$. Hence, P should be high enough to provide enough energy to account for the ZPE (an empirical criterion for convergence is that Pk_BT should be sufficiently larger than $\hbar\omega_0$, with ω_0 the highest frequency of the system). Therefore, when the temperature decreases, $\beta \longrightarrow \infty$ and P has to be very large compared to $\hbar\omega_0\beta$. Moreover, in a system with many degrees of freedom, the highest vibrational frequency should be used to estimate how many beads are needed. In practise, the number of beads P acts as a convergence parameter and the ring-polymer density is sampled via molecular dynamics using different types of thermostats, such as Nosé-Hoover chains or Langevin dynamics (the latter will be used in the rest of this Thesis).

2.2.2 Sampling with Path Integral Molecular Dynamics

The path integral formalism was first exploited in MD simulations by Parrinello and Rahman in 1984 [117]. They introduced in equation (2.6) an effective kinetic energy as a function of effective momenta of the polymer beads p_j:

$$\mathcal{Z} = \lim_{P\to\infty}\left(\frac{1}{2\pi\hbar}\sqrt{\frac{m}{m'}}\right)^P \int \mathrm{d}q_1,\ldots,\mathrm{d}q_p\mathrm{d}p_1,\ldots,\mathrm{d}p_P\, e^{-\beta_P \sum_{j=1}^{P}\left\{\frac{p_j^2}{2m'}+\frac{m\omega_P^2}{2}(q_{j+1}-q_j)^2+V(q_j)\right\}} \tag{2.7}$$

where it can be identified:

$$H_P = \sum_{j=1}^{P}\left\{\frac{p_j^2}{2m'}+\frac{m\omega_P^2}{2}(q_{j+1}-q_j)^2+V(q_j)\right\} \tag{2.8}$$

H_P referred to as the ring polymer Hamiltonian, also called effective Hamiltonian. It is a classical Hamiltonian that can be used to generate a classical dynamics in the extended phase space of the beads. The masses m' have no physical meaning and one can choose them to arbitrary values with no effect on the configurational ensemble generated by H_P. Therefore, the molecular dynamics generated in this case is just a mean to evaluate quantum statistical averages, and the concept of 'time' has no physical meaning. The equations of motion generated by equation (2.8) are given by:

$$\begin{aligned}\dot{p}_j &= -m\omega_P^2\left(2q_j - q_{j+1} - q_{j-1}\right) - \frac{\partial V(q_j)}{\partial q_j}\\ \dot{q}_j &= \frac{p_j}{m'}\end{aligned} \tag{2.9}$$

The harmonic springs between the beads generally vibrate at frequencies much higher than the physical frequencies of motion of the system. Then, a small time step will be required to integrate equation (2.9) [46, 118]. To circumvent that, one may work in the normal mode representation of the ring polymer by introducing new variables:

$$\begin{aligned}Q_n &= \sum_{i=1}^{P} T_{in}q_i\\ P_n &= \sum_{i=1}^{P} T_{in}p_i\end{aligned} \tag{2.10}$$

with $n = 0,\ldots,P-1$. T is the unitary transfer matrix whose columns are eigenvectors of the symmetric $P\times P$ matrix Ω_P^2 defined as:

$$\Omega_P^2 = \omega_P^2\begin{pmatrix} 2 & -1 & & & -1\\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1\\ -1 & & & -1 & 2\end{pmatrix} \tag{2.11}$$

where all undefined terms are zeros. Therefore, it is possible to express equation (2.8) as:

$$H_P = \sum_{i=1}^{P} V\left(\sum_{n=0}^{P-1} T_{in} Q_n\right) + \sum_{n=0}^{P-1} \frac{1}{2m} P_n^2 + \sum_{n=0}^{P-1} \left(\frac{1}{2} m\omega_n^2 Q_n^2\right) \tag{2.12}$$

with P_n the normal mode momenta obtained with equation (2.10) and ω_n^2 the eigenvalues of equation (2.11). Q_n and P_n represent the phase-space coordinates of the normal modes of the free ring polymer, that are coupled to one another via the physical potential V. Going to the normal mode representation allows not to be obliged to use very small time steps, thanks to an appropriate splitting of the time integrator. Then, the normal modes are harmonic oscillators which can be exactly integrated at each time step. Although the harmonic frequencies involved in the system can be high, it is possible to use a moderately small time step. Another possibility is to transform the representation to the so-called 'staging variables' [46] which have a similar effect. The normal mode decomposition is also at the basis of the ring-polymer contraction method that allows to further reduce the computational cost by evaluating the long-range part of the interaction potential on a smaller number of beads (potentially even only on the centroid) [119, 120].

2.2.3 Path Integral Estimators

Since the fundamental equations and background of path integral formalism have been derived previously, one may be interested to obtain averages of observables of the form:

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

In the case where the operator $\hat{A}$ is a function of position only $\hat{A}(q)$ and using the invariance under cyclic permutation of the partition function, it can be shown that:

$$\langle \hat{A} \rangle = \lim_{P\to\infty} \frac{1}{Z} \int \mathrm{d}q_1, \ldots, \mathrm{d}q_P \mathrm{d}p_1, \ldots, \mathrm{d}p_P \left(\frac{1}{P} \sum_{j=1}^{P} A(q_j)\right) e^{-\beta_P H_P} \tag{2.14}$$

where H_P is given by equation (2.8). This expression can be sampled straightforwardly. by using molecular dynamics. One can define the instantaneous estimator A_P for the observable A as:

$$A_P[q(t)] = \frac{1}{P} \sum_{j=1}^{P} A[q_j(t)] \tag{2.15}$$

which is valid only for position-dependent observables. However, one may be interested in observables that depend on the momentum such as the kinetic energy.

The Primitive Kinetic Estimator

To derive an estimator for the kinetic energy, the relation between the total energy and the partition function is used:

$$\langle E \rangle = -\frac{1}{\mathcal{Z}} \left(\frac{\partial \mathcal{Z}}{\partial \beta}\right) \tag{2.16}$$

Inserting the path integral partition function for a one-dimensional system given by equation (2.6) leads to:

$$\langle E \rangle = \left\langle \frac{Pk_BT}{2} - \sum_{j=1}^{P} \frac{1}{2} m\omega_P^2 (q_{j+1} - q_j)^2 + \frac{1}{P} \sum_{j=1}^{P} V(q_j) \right\rangle \tag{2.17}$$

where the last term is the potential energy estimator. The average on the right-hand side is taken over the ring-polymer statistical probability distribution. The last term of equation (2.17) corresponds to the average potential energy, estimated according to equation (2.15). Therefore, the sum of the first two terms provides an estimator for the kinetic energy $\mathcal{K}_{prim}$:

$$\mathcal{K}_{prim} = \frac{Pk_BT}{2} - \sum_{j=1}^{P} \frac{1}{2} m\omega_P^2 (q_{j+1} - q_j)^2 \tag{2.18}$$

which is a difference between two terms. However, both grow larger with P, increasing the fluctuations in the instantaneous values of $\mathcal{K}_{prim}$ along a path integral trajectory. Hence, the statistical error also grows with P. Therefore, longer trajectories are needed to average out these fluctuations and obtain an accurate value of the kinetic energy. To circumvent that, one may derive an alternative estimator $\mathcal{K}_{virial}$ whose fluctuations do not grow with the number of replicas P [121].

The Centroid Virial Kinetic Estimator

To obtain a better convergence behavior for the kinetic estimator, one can employ a coordinate scaling approach [122, 123]. The ring polymer positions q_j are scaled relative to their centroid position $\{q_j^s\}$ via:

$$q_j^s = q_c + \lambda(q_j - q_c) \tag{2.19}$$

where the centroid q_c is defined as:

$$q_c = \frac{1}{P} \sum_{j=1}^{P} q_j \tag{2.20}$$

The Jacobian associated with this transformation is:

$$\mathrm{d}q_1^s, \ldots, \mathrm{d}q_P^s = \lambda^{P-1} \mathrm{d}q_1, \ldots, \mathrm{d}q_P \tag{2.21}$$

Therefore, a new expression for the partition function may be derived as:

$$\mathcal{Z} = \left[\lambda \left(\frac{1}{2\pi\hbar^2} \right)^{P/2} \int \mathrm{d}q_1^s, \ldots, \mathrm{d}q_P^s \, e^{-\frac{1}{2\hbar^2} \sum_{j=1}^{P} (q_{j+1}^s - q_j^s)^2 - \frac{\beta}{P} \sum_{j=1}^{P} V(q_c + \lambda^{-1}(q_j^s - q_c))} \right] \tag{2.22}$$

where $\lambda = \sqrt{\beta'/\beta}$ and β' a fictitious temperature. Equation (2.22) can be inserted in equation (2.16) to evaluate the total energy in the temperature scaled coordinates. Transforming back to the unscaled ring polymer gives the total energy estimator as:

$$\langle E \rangle = \left\langle \frac{1}{2\beta} + \frac{1}{2P} \sum_{j=1}^{P} \frac{\partial V(q_j)}{\partial q_j} (q_j - q_c) + \frac{1}{P} \sum_{j=1}^{P} V(q_j) \right\rangle \tag{2.23}$$

One may identify the centroid virial estimator for the kinetic energy as:

$$\mathcal{K}_{virial} = \frac{k_B T}{2} + \frac{1}{2P}\sum_{j=1}^{P}(q_j - q_c)\frac{\partial V}{\partial q_j} \tag{2.24}$$

Thanks to the use of the forces and the distances of each replica to the centroid, the virial estimator provides an efficient way to evaluate the kinetic energy from a path integral molecular dynamics. It can also be derived using the classical virial theorem [46] on the primitive kinetic estimator [121]. However, if the simulation is correctly performed over a sufficiently long trajectory, the averaged kinetic and virial estimator are equal. Hence, during the thesis, both estimators were computed to ensure they were equal. If they are not equal, it highlights a fundamental error in the performed simulation. Therefore, with its smaller statistical errors, the virial estimator was mainly used to evaluate the kinetic energy.

2.2.4 Quantum Mechanical Time Correlation Functions

Although the imaginary-time path integrals formalism is a powerful technique to compute equilibrium properties, it cannot directly yield information about the dynamics of the system. However, many interesting quantities regarding the dynamic of a system can be expressed in terms of time-correlation functions. The standard real time correlation function between two generic operators A and B can be expressed as [124]:

$$\mathcal{C}_{AB}(t) = \frac{1}{Z}\mathrm{Tr}\left[e^{-\beta\hat{H}}\hat{A}(0)\hat{B}(t)\right] \tag{2.25}$$

where $\hat{B}(t)$ is propagated according to the quantum dynamics using:

$$B(t) = e^{i\hat{H}t/\hbar}\hat{B}e^{-i\hat{H}t/\hbar} \tag{2.26}$$

However, $\mathcal{C}_{AB}(t)$ can be a complex quantity contrary to its classical equivalent. Then, one may want to use the real part, $\tilde{\mathcal{C}}_{AB}(\omega)$, of equation (2.25) which displays a more 'classical' behavior. $\tilde{\mathcal{C}}_{AB}(\omega)$ can be related to the standard correlation function by using Fourier transforms:

$$\begin{aligned}\tilde{\mathcal{C}}_{AB}(\omega) &= f(\omega)\mathcal{C}_{AB}(\omega) = f(\omega)\int_{-\infty}^{+\infty}\mathrm{d}t\, e^{-i\omega t}\mathcal{C}_{AB}(t)\\ f(\omega) &= \frac{\beta\hbar\omega}{(1-e^{-\beta\hbar\omega})}\end{aligned} \tag{2.27}$$

It is also possible to reconstruct the standard correlation function from the Kubo-transformed correlation function (and vice versa) which is defined as [125, 126]:

$$\mathcal{K}_{AB}(t) = \frac{1}{\beta Z}\int_0^\beta \mathrm{d}\lambda\,\mathrm{Tr}\left[e^{-(\beta-\lambda)\hat{H}}\hat{A}(0)e^{-\lambda\hat{H}}\hat{B}(t)\right] \tag{2.28}$$

where the Boltzmann operator $e^{-\beta\hat{H}}$ as been 'smeared' around $\hat{A}(0)$. Working in the eigenstates basis of $\hat{H}$ allows to express the standard correlation function as a function of the Kubo-transformed correlation function as:

$$\tilde{\mathcal{C}}_{AB}(\omega) = \frac{\beta\hbar\omega}{1-e^{-\beta\hbar\omega}}\mathcal{F}\left[\mathcal{K}_{AB}(t)\right] \tag{2.29}$$

Then, the real part of the standard correlation function is:

$$\tilde{\mathcal{C}}_{AB}(\omega) = \frac{\beta\hbar\omega/2}{\tanh(\beta\hbar\omega/2)}\mathcal{K}_{AB}(\omega) \tag{2.30}$$

where $\mathcal{F}[\cdot]$ is the Fourier transform. The Kubo-transformed correlation function has the same symmetry properties as the corresponding classical correlation function [127]. Therefore, it plays the same role in quantum mechanical linear response theory as the classical correlation function in classical linear response theory [125, 126]. Centroid Molecular Dynamics (CMD) and Ring-Polymer Molecular Dynamic (RPMD) are used to generate approximations to $\mathcal{K}_{AB}$ and then compute dynamical properties, such as Infrared/ Raman spectra or diffusion coefficients.

2.2.5 Centroid Molecular Dynamics

The centroid molecular dynamics (CMD) was firstly introduced by J. Cao and G. A. Voth in 1993 [128]. Using the definition of the centroid of equation (2.20), CMD relies on the following equations of motion:

$$\begin{aligned} \dot{q}_c &= \frac{p_c}{m} \\ \dot{p}_c &= -\frac{\partial V_c(q_c)}{\partial q_c} \end{aligned} \tag{2.31}$$

where m is the physical mass, p_c is the conjugated momentum of q_c and $\partial V_c(q_c)/\partial q_c \equiv F_c(q_c)$ is the derivative of the centroid potential of mean force defined as:

$$V_c(q_c) = -\frac{1}{\beta}\ln\left\{\left(\frac{2\pi\beta\hbar^2}{m}\right)^{1/2}\oint \mathcal{D}q(\tau)\delta\left(q_0[q(\tau)] - q_c\right)e^{-\mathcal{S}[q(\tau)]/\hbar}\right\} \tag{2.32}$$

where

$$q_0[q(\tau)] = \frac{1}{\beta\hbar}\int_0^{\hbar\beta} q(\tau)\mathrm{d}\tau \tag{2.33}$$

is the centroid position of the path $q(\tau)$. The δ-function therefore restricts the integration to cyclic paths whose centroid position is q_c. Using the equation of motion of (2.31), it is possible to compute the forces and integrate the effective centroid dynamics to obtain the CMD approximation to the Kubo-transformed quantum time correlation function $\mathcal{K}_{AB}(t)$ as:

$$\langle\mathcal{K}_{AB}(t)\rangle \simeq \langle B_c(0)A_c(t)\rangle = \frac{1}{(2\pi\hbar)^P Z}\int \mathrm{d}p_c \int \mathrm{d}q_c\, A_c B_c\left[q_c(t), p_c(t)\right] e^{-\beta\left[\frac{p_c^2}{2m} + V_c(q_c)\right]} \tag{2.34}$$

where the function $B_c\left[q_c(t), p_c(t)\right]$ is evaluated using the time-evolved centroid variables generated by equation (2.31) using $\{q_c, p_c\}$ as initial conditions (equation (2.34) is valid for linear observables A and B, for non-linear ones, additional terms are needed in the estimator [104]). However, CMD requires to compute V_c at each time step of the trajectory which necessitates many PIMD steps. To reduce the computational cost, adiabatic decoupling can be employed [104, 129–131] where the non-centroid modes are decoupled from the centroid's motion and therefore speed up the calculation. Despite this, CMD is still computationally expensive and difficult for large and complex systems. Moreover, it is only exact in the limit of purely harmonic potential. Far from this limit, position auto-correlation functions are accurate up to $\mathcal{O}(t^3)$ [132]. In addition, CMD

suffers from a 'curvature problem' when applied to systems with both bond stretching and rotational modes [133, 134]. As the temperature is lowered, the ring polymer spreads out around the rotational modes and the centroid potential of mean force becomes softer in the stretching modes. It results in a spurious red-shift of the stretching frequencies which can become a serious issue in ice at low temperatures. To correct this curvature problem, several attempts have been made such as using curvilinear coordinates instead of cartesian coordinates [135]. However, CMD has been shown to provide satisfactory results in different cases [107, 136–138].

2.2.6 Ring-polymer Molecular Dynamics

The ring-polymer molecular dynamics (RPMD) was originally introduced by Craig and Manolopoulos [127] in 2004. The fictitious masses $m^{'}$ in equation (2.8) are chosen such that each imaginary time slice or bead have the physical mass m. Compared to CMD, RPMD uses the full chain to approximate time correlation functions. Hence, a quantum observables $\hat{A}(q)$ is represented by the following RPMD estimator:

$$A_P(t) = \frac{1}{P}\sum_{k=1}^{P} A\left[q_j(t)\right] \tag{2.35}$$

For observables which linearly depend on the position or momenta, equation (2.35) reduces to q_c and p_c respectively. The RPMD Kubo-transformed time correlation function $\mathcal{K}_{AB}$ is computed as:

$$\langle \mathcal{K}_{AB}(t)\rangle = \langle B_P(0)A_P(t)\rangle = \frac{1}{(2\pi\hbar)^P Z}\int \mathrm{d}p_c \int \mathrm{d}q_1,\ldots,\mathrm{d}q_p\, A_P(0)B_P(t)e^{-\beta H_P} \tag{2.36}$$

where H_P is defined in equation (2.8). Because $q_c(t)$ and $p_c(t)$ are propagated using the instantaneous centroid potential in RPMD, whereas in CMD they were propagated using the average centroid potential of mean force, the equations of motions are different and are given by:

$$\begin{aligned} \dot{q}_j &= \frac{p_j}{m} \\ \dot{p}_j &= -\frac{m}{\beta_P^2\hbar^2}\left(2q_j - q_{j-1} - q_{j+1}\right) - \frac{\partial V}{\partial q_j} \end{aligned} \tag{2.37}$$

making it less computationally expensive. In the original RPMD implementation, equations (2.37) where used without thermostat when computing correlation functions, in order to not perturb the dynamics.

Although RPMD is an approximation and cannot be rigorously derived from the quantum dynamics, it was shown to provide accurate results in a variety of cases [107, 136–139]. The performance of CMD compared to RPMD highlight that the chain frequencies from the ring polymer can interfere with the physical frequencies leading to non-physical peaks in the vibrational spectrum [133]. The lowest internal modes of a free ring polymer have frequencies of:

$$\omega_j = \frac{2P}{\beta\hbar}\sin\left(\frac{j\pi}{P}\right) \approx \frac{2\pi j}{\beta\hbar} \tag{2.38}$$

These modes are combined with the physical modes of frequency ω, hence the internal mode frequencies are shifted to $\sqrt{\omega^2+\omega_j^2}$. For example, in water at 300 K, an interaction between the second internal mode of the ring polymer and the water bending modes was observed, which gave a frequency of

$\sqrt{2600^2 + 1600^2} = 3052\,\mathrm{cm}^{-1}$. An analogous calculation at 350 K gives an oscillation with a wavenumber of $\sqrt{3033^2 + 1600^3} = 3429\,\mathrm{cm}^{-1}$. Both values are close to the stretching frequencies. Therefore, the physical peaks in the spectrum are not shifted as in CMD, but additional ones, called spurious peaks, can appear in the final spectrum [140]. The spurious frequency problem is only a problem if these features overlap with physical frequencies of the studied system which is the case for water. However, it has been shown that RPMD provides an effective way to include NQEs in the diffusion of liquids [107], but also for the orientational correlation times of molecular liquids [30]. It should be stressed that the zero point energy leakage does not happen in RPMD because its dynamic is consistent with the PIMD equilibrium distribution.

2.2.7 Thermostatted RPMD

To mitigate the spurious peaks problem one can apply a Langevin thermostat to the internal modes of the ring polymer during the dynamics. The resulting thermostatted RPMD (TRPMD) is characterized by the following equations of motion: [141]:

$$\begin{aligned} \dot{Q}_n &= \frac{P_n}{m} \\ \dot{P}_n &= -\frac{\partial H_P(P_n, Q_n)}{\partial Q_n} - \gamma P_n + \sqrt{\frac{2m\gamma}{\beta_P}}\, R(t) \end{aligned} \tag{2.39}$$

Applying a Langevin dynamics in the normal mode representation allows to tune the friction coefficient γ to match the critical damping value of each normal mode representation (while the centroid is under-damped). Although the thermostat in RPMD damps the oscillations of the internal modes of the ring polymer, it does not completely remove them which causes a broadening of the physical peaks that becomes more pronounced at low temperatures.

In conclusion, PIMD is the reference method to include NQEs and yields exact results for static observables when the number of beads is sufficiently large. Although structural properties of the system are exact, in the PIMD formalism, dynamical properties are not. Then, to compute real time correlation functions, approximate methods such as CMD and RPMD have to be used. However, their computational cost remains high, which has stimulated the research of alternative methods, such as those based on generalized Langevin equations presented in the next section [25, 26].

2.3 The Quantum Thermal bath for Molecular Dynamics

As an alternative to path integral methods, a Generalized Langevin Equation (GLE) can be used to approximate the quantum statistical distribution (in particular ZPE). In this section, the classical version of the GLE and the associated fluctuation-dissipation theorem (FDT) will be presented (first and second kind). In Section 2.3.2, the Quantum Thermal Bath (QTB) framework will be studied. It modifies the time-correlation of the stochastic forces to impose the quantum version of the second-kind FDT in the quantum case. However, this method suffers from Zero Point Energy Leakage (ZPEL). To overcome that, the recently developed adaptive QTB (adQTB) will be introduced which aims to correct ZPEL by using a criterion derived using the first-kind FDT.

2.3.1 The Classical Fluctuation-dissipation Theorem

The Langevin equation was firstly introduced by Langevin [142, 143] to describe Brownian motion. It is obtained by considering a small portion of the system that can exchange energy with its environment, called the bath. It is widely used in MD simulations to sample the canonical ensemble as it has been presented in Chapter 1 Section 1.2.3. Generalizations of the Langevin equation have later been introduced to account for cases in which the strength of the interaction with the bath is frequency-dependent. This formulation allows describing a wide range of processes in physics and chemistry [144, 145]. In order not to take into account explicitly the bath degrees of freedom, a frequency-dependent stochastic thermostat is added to Newton's equation of motion. Then, the equations of motion of the dynamics become non-Markovian:

$$\begin{aligned} \dot{q} &= \frac{p}{m} \\ \dot{p} &= -\frac{\partial V}{\partial q} - \int_0^\infty \mathrm{d}\tau\, \gamma(\tau) p(t-\tau) + R(t) \end{aligned} \tag{2.40}$$

where $\gamma(\tau)$ is the memory kernel. $R(t)$ is a stationary Gaussian stochastic force with a correlation function characterized by:

$$\langle R(t)R(t+\tau)\rangle = 2mk_BT\gamma(\tau) \tag{2.41}$$

This formula relates the random force amplitude (fluctuations) to the memory kernel of the friction force (dissipation) in the GLE. Both terms model the effect of the bath and ensure the canonical statistics. Therefore, equation (2.41) is designated as second-kind Fluctuation-Dissipation Theorem (FDT) [146, 147]. In the standard Langevin equation, the friction is Markovian (without memory), i.e. such that $\gamma(\tau) = \gamma\delta(\tau)$ with $\delta(\tau)$ the Dirac delta function. In this case $R(t)$ becomes a white noise with an amplitude given by equation (1.29).

The second-kind FDT characterizes the interaction of the system with the thermostat, as it relates the two Langevin forces (the friction and the random force), it is thus specific of the GLE. However, an other relation can be derived from the linear response theory, the first-kind FDT, which is valid even outside the context of the GLE and is therefore more general. It can be expressed as [110]:

$$\frac{\mathcal{C}_{vv}(\omega)}{2} = \mathrm{Re}\left[\chi_{vq}(\omega)\right] k_BT \tag{2.42}$$

where $\mathcal{C}_{vv}(\omega)$ is the Fourier transform of the velocity time autocorrelation function. Equation (2.42) is called first-kind FDT. The term $\chi_{vq}(\omega)$ is the susceptibility that characterizes the velocity linear response $\Delta v(t)$ for a system at thermal equilibrium undergoing a small perturbative force $\Delta R(t) = \mathrm{Re}\left[\Delta R(\omega)e^{i\omega t}\right]$, oscillating at the angular frequency ω:

$$\Delta v(\omega) = \chi_{vq}(\omega)\Delta R(\omega) \tag{2.43}$$

For a harmonic system, $\mathrm{Re}\left[\chi_{vq}(\omega)\right]$ is a sum of Dirac delta functions located at the different frequencies which are characteristic of the system. Each vibrational mode yields a Dirac delta function, then $\mathrm{Re}[\chi_{vq}(\omega)]$ represents the vibrational density of states. Therefore, equation (2.42) states that each vibrational mode at frequency ω is thermalized with an average kinetic energy $k_BT/2$. The first-kind FDT therefore expresses a fundamental property of classical statistical physics, the equipartition of energy, each degree of freedom (or each vibrational modes) have the same average energy k_BT.

2.3.2 The QTB and Zero-point Energy Leakage

The Quantum Thermal Bath (QTB) was introduced by Dammak and coworkers in 2009 [25] (similar ideas were simultaneously developed by Ceriotti et al. with a slightly different formalism under the name Quantum thermostat [26]). The QTB aims at accounting for what is usually the most relevant NQEs: the fact that the quantum distribution of energy differs from the classical equipartition of the energy due to zero point energy (ZPE). The classical fluctuation-dissipation theorem described above expresses the equipartition of energy. However, in the quantum case, one expects the frequency distribution of the energy distribution of a system to be:

$$\theta(\omega, T) = \hbar|\omega| \left(\frac{1}{2} + \frac{1}{e^{\hbar|\omega|/k_B T} - 1} \right) \tag{2.44}$$

which depends on the angular frequency $|\omega|$ (since its zero-point energy is $\hbar\omega/2$). The idea of the QTB is to modify the Langevin thermostat to enforce the quantum energy distribution given by equation (2.44). To do so, one may use a GLE as defined in equation (2.40) with a Markovian friction (as in standard Langevin), but modifying the frequency energy distribution of the stochastic force to impose:

$$\mathcal{C}_{RR}(\omega) = 2m\gamma\theta(\omega, T) \tag{2.45}$$

The QTB method modifies the time-correlation of the stochastic forces according to equation (2.45) which corresponds to the second-kind FDT in the quantum case. Since its spectrum depends on ω, the random force $R(t)$ of the QTB is not a white noise but a colored noise, even if the friction remains Markovian. At high temperature the classical limit is reached and $\theta(\omega, T) \approx k_B T$. On the other hand, at low temperature, the function $\theta(\omega, T)$ diverges from the classical behavior and NQEs can no longer be neglected. In particular, in the limit $T \to 0$, where the function $\theta(\omega, T) \to \hbar\omega/2$ is the ZPE of the harmonic oscillator. Therefore, the equipartition theorem does not hold and the average kinetic energy of any degree of freedom includes ZPE and thus is greater than $k_B T/2$. Hence, the effective temperature is greater than the physical temperature T and it only becomes equal to T in the classical high-temperature limit.

However, using a thermostat given by equation (2.45) is not sufficient to enforce the quantum FDT of the first kind for anharmonic systems. The classical force $-\partial V/\partial q$ couples the different modes and drives the system towards the classical equipartition of energy whereas the thermostat tries to impose the quantum distribution of energy. As the anharmonicity is increased, the energy in the high frequency modes decreases while the energy in the low frequencies increases, getting closer to the equipartition of the energy. Therefore, the average thermal energy per mode is lower than $\theta(\omega, T)$ for large ω and greater than it for small ω. This unphysical energy flow from high to low frequencies constitutes the ZPEL. To correct the ZPEL, several solutions have been suggested such as modifying the noise power spectrum to obtain the expected power distribution [148]. Although really promising, this solution has the drawback to be system dependent. Therefore, a study was done by Brieuc et al. to investigate the ZPEL by testing the QTB on different model systems [149]. They increased the friction coefficient γ to prevent ZPEL. However, the increase of gamma affects the dynamical properties of the system (in particular with a broadening of the vibration spectra), and although it reduces ZPEL, it does not entirely suppress it. Even with large gamma coefficients, ZPEL can remain problematic in some very anharmonic systems such as liquid water, where ZPEL is particularly massive from the high-frequency intramolecular vibration modes to the slower intermolecular motion. To

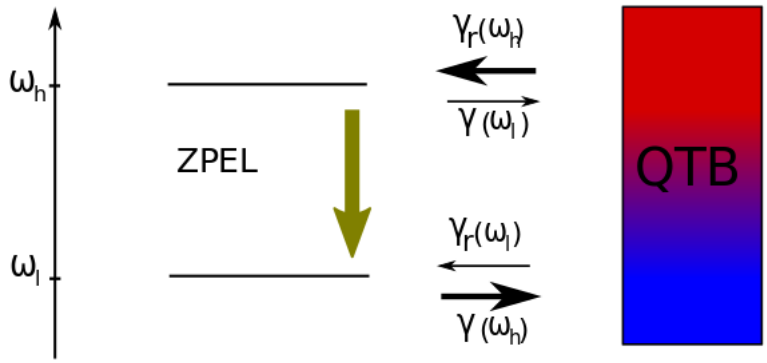


Figure 2.3: Schematic representation of the adQTB scheme to correct the ZPEL where $\gamma_r(\omega)$ is adapted accordingly to the deviation from the quantum FDT $\Delta_{FTD}(\omega)$ (equation 2.51). $\gamma_r(\omega)$ will be increased at high frequency and reduced at low frequency to minimize the deviation and reach the correct energy distribution for each frequency [27].

address this problem, an adaptive QTB (adQTB) was developed based on the measurement of deviations from the first kind of FDT.

2.3.3 The Adaptive QTB

In the QTB method, each degree of freedom follows a Langevin equation such as (2.40). Each vibrational mode is thermalized at the effective thermal energy $\theta(\omega, T)$ which increases with the mode frequency ω and accounts for ZPE. However, for anharmonic system, an unphyscial effect arises from the competition between the classical Newtonian evolution and the thermostat. One may solve the problem by measuring the deviations from the the first kind quantum FDT, which is equivalent to equation (2.42), only that the thermal energy k_BT is replaced by $\theta(\omega, T)$. The main difficulty is to estimate the linear susceptibility $\chi_{vq}(\omega)$. It has been shown that it can be expressed as a ratio of two correlation functions [27, 146]:

$$\mathrm{Re}[\chi_{vq}(\omega)] = \frac{\mathrm{Re}[\mathcal{C}_{vR}(\omega)]}{\mathcal{C}_{RR}(\omega)} \tag{2.46}$$

Equation (2.46) was obtained by using the property that the frequency components of the force R are independent from another ($\langle R(\omega)R(\omega')\rangle = 0$ if $\omega \neq \omega'$). Hence, the random force is treated at each particular frequency as a small perturbative force and it is used to compute the linear response $\chi_{vq}(\omega)$ using equation (2.43). Moreover, $\mathcal{C}_{RR}(\omega)$ is given by equation (2.45), then the quantum FDT in the QTB framework can be rewritten as:

$$\mathrm{Re}\left[\mathcal{C}_{vR}(\omega)\right] = m\gamma\mathcal{C}_{vv}(\omega) \tag{2.47}$$

Note that $\mathrm{Re}\left[\mathcal{C}_{vR}(\omega)\right]$ also represents the power injected by the thermostat at frequency ω and $\gamma\mathcal{C}_{vv}(\omega)$ corresponds to the dissipated power. Then, equation (2.47) relates the velocity response to the stochastic force $\mathrm{Re}\left[\mathcal{C}_{vR}(\omega)\right]$ and the internal dynamical properties. Equation (2.47) should be verified for each degrees of freedom to ensure that the equilibrium distribution of energy follows the quantum FDT. As explained

earlier, this is the case for harmonic potential. However, the equality is not generally fulfilled for anharmonic systems. Thus, equation (2.47) is used as a criterion to quantify the deviations from the quantum FDT of the first kind:

$$\Delta_{FDT}(\omega) = \mathrm{Re}\left[\mathcal{C}_{vR}(\omega)\right] - \gamma m \mathcal{C}_{vv}(\omega) \tag{2.48}$$

To enforce the quantum FDT in QTB simulations, the power spectrum of the stochastic force given by equation (2.45) is replaced by:

$$\mathcal{C}_{RR}(\omega) = 2m\gamma_r(\omega)\theta(\omega, T) \tag{2.49}$$

In the standard QTB $\gamma_r(\omega) = \gamma$ independently of ω. In the adaptive QTB, γ_r is an adjustable parameter of the amplitude of the Langevin random forces. It is adapted to enforce the first kind FDT. Using equation (2.49), it is possible to express the quantum FDT for this modified spectrum as:

$$\mathrm{Re}\left[\mathcal{C}_{vR}(\omega)\right] = m\gamma_r(\omega)\mathcal{C}_{vv}(\omega) \tag{2.50}$$

and then the deviation from the quantum FDT:

$$\Delta_{FDT}(\omega) = \mathrm{Re}[\mathcal{C}_{vR}(\omega)] - m\gamma_r(\omega)\mathcal{C}_{vv}(\omega) \tag{2.51}$$

When the quantum FDT of the first kind is enforced $\Delta_{FDT}(\omega) = 0$. To be sure this relation is respected during the dynamics, the right side is calculated regularly and the values of $\gamma_r(\omega)$ are then adapted for each frequency to reduce $\Delta_{FDT}(\omega)$: it will increase at high frequencies (the modes that loose energy due to the leakage), and decrease at low energy to compensate the effects of the ZPEL (Figure 2.3).

2.3.4 Colored Noise Generation

In the QTB and adQTB methods, one needs to generate a random force with any arbitrary correlation function $C_{RR}(\omega) = g(\omega)$. Because the stochastic force is no more a white noise, the standard way to generate the random force in Langevin thermostat cannot be used. A filter $\tilde{H}(\omega)$ is introduced such as [150]:

$$\tilde{H}(\omega) = \sqrt{g(\omega)} \tag{2.52}$$

The inverse Fourier transform of $\tilde{H}(\omega)$ gives $H(t)$. The noise $R(t)$ is obtained by convoluting $H(t)$ with a white noise $r(t)$:

$$R(t) = \int_{-\infty}^{+\infty} \mathrm{d}s\, H(s) r(t-s) \tag{2.53}$$

The noise can be implemented by making a discretization on the filter $\tilde{H}(\omega)$ such as:

$$\tilde{H}_k = \tilde{H}(k\Delta\omega) \tag{2.54}$$

with $\Delta\omega$ the angular frequency step defined as $\Delta\omega = 2\pi/(N_{seg}\Delta t)$. N_{seg} is the length of a segment of trajectory. In MD simulation, the time is discretized as $t_n = n\Delta t$ and a colored noise $R_n = R(t_n)$ is then

generated for a segment of trajectory of duration $N_{seg}\Delta t$.

One may notice that equation (2.44) diverges for high frequency. These high frequencies are not expected to be present in real systems. Thus, a cutoff frequency ω_{cut} is introduced with the function $f(\omega) = 1/(1 + e^{(\omega-\omega_{cut})/\omega_{smear}})$. A correction factor $C(\omega)$ is also used to account for finite time step errors defined as:

$$C(\omega) = \frac{1 - 2e^{-\gamma\Delta t}\cos(\omega\Delta t) + e^{-2\gamma\Delta t}}{\gamma^2 + \omega^2} \frac{1}{\Delta t^2} \tag{2.55}$$

Therefore, the filter $\tilde{H}(\omega)$ is defined as [150]:

$$\tilde{H}(\omega) = \sqrt{g(\omega)C(\omega)f(\omega)} \tag{2.56}$$

Because $\tilde{H}(\omega)$ is real and even, its Fourier transform $H(t)$ is given by:

$$H_n = \frac{1}{2N_{seg}} \sum_{k=-N_{seg}}^{N_{seg}-1} \tilde{H}_k \cos\left(\frac{\pi}{N_f} kn\right) \tag{2.57}$$

with $H_n = H(n\Delta t)$, for $k = -N_{seg}, \ldots, N_{seg} - 1$. A convolution procedure is then performed to obtain the discrete colored noise:

$$R_n = R(n\Delta t) = \sum_{m=-N_{seg}}^{N_{seg}-1} H_m r_{n-m} \tag{2.58}$$

with r_m a white noise with a Gaussian distribution. The colored noise can be either stored, generated 'on the fly' or generated for each segment. To avoid the storage and allow to study more complex systems, the random noise is generated regularly every N_{seg} time steps, for the next segment of trajectory during this thesis.

2.4 Summary

NQEs are an important feature which should be taken into account in MD simulations. They are relevant to describe processes occurring at low temperatures [151, 152] or high pressure [14, 15, 112], but also close to ambient conditions such as reactions involving proton transfer [16, 20] or isotopic effects [19, 153–155]. The path integral formalism gives an elegant framework to include NQEs. The quantum thermal density is evaluated as the Boltzmann probability distribution of an isomorphic classical system referred to as a ring polymer. Within this isomorphism, each quantum particle is represented using beads coupled by harmonic springs. However, the number of beads should be sufficiently large to converge to the exact solution. Consequently, the numerical cost can be prohibitive for complex and large system or when computer resources are limited. Therefore, it motivates the development of alternative methods such as the Quantum Thermal Bath. Compared to path integrals methods, the QTB is much less computationally expensive because no replica is needed. The quantum statistics is taken into account with a random colored noise which simulates both thermal and quantum fluctuations. It is chosen such that its power spectral density matches the quantum FDT. However, QTB has some drawbacks due to the ZPEL: part of the energy of high frequency modes is transferred to the low frequency ones and leads to an incorrect energy distribution. Therefore, an adaptive

method, called adQTB, was developed to correct the leakage and to recover the correct energy distribution. It enforces the first kind quantum FDT by adjusting 'on the fly' the parameters of the thermostat. The method was already studied on model-like systems [15, 112] and shows good results which motivate further investigation. To push the limits of the method further, in the next chapter it is studied on a water potential derived from path integral formalism: q-TIP4P/F. It was parameterized to give the correct liquid structure, diffusion coefficient and infrared absorption spectrum in quantum (path-integral-based) simulations. Therefore, the next chapter will be the study of the adQTB method compared to the PIMD (and TRPMD for dynamical observables) using the q-TIP4P/F force field.

Chapter 3

Nuclear Quantum Effects in Water via the Adaptive Quantum Thermal Bath

Contents

3.1 Nuclear Quantum Effects in Water

With its major role in physics, chemistry, biology and geophysics, water is one of the most important liquids. It is still a challenge to understand how this molecule can have such a complex and unusual behaviour over a broad range of temperatures and pressures, such as the bell shape of density around ambient temperature and its strong permanent dipole which are some of the anomalous properties of this compound [21, 156]. These anomalies motivate researchers to understand its microscopic structure making water an issue of long-standing interest. Modeling water at a molecular level is central for understanding its behaviour in many physico-chemical processes and to understand its decisive role in life. Nevertheless, given the low mass of hydrogen atoms, classical mechanics is not accurate enough to describe the liquid phase even at ambient temperatures where NQEs can have significant contributions. The spatial delocalization, due to quantum mechanics, is conflicting with the confinement due to the atomic environment. Furthermore, the ZPE in the O-H stretching and bending vibrations is more than ten times larger than the thermal energy at room temperature k_BT. It makes the description of NQEs rather subtle: for instance, a simple classical MD trajectory that would take ZPE into account in its initial energy distribution would rapidly lead the water to boil, because of the ZPE leakage effect that drives the system toward the equipartition of this energy [22].

Furthermore, net NQEs on the hydrogen-bond network of water are relatively weak. This weakness is created by the competition between two opposite trends: the stretching ZPE strengthens hydrogen bonding, while the bending ZPE weakens it as shown in Figure 3.1. Quantum fluctuations allow the protons to be more delocalized between hydrogen-bond pairs of water molecules which strengthens the bond, as illustrated schematically in the right panel of Figure 3.1. However, it also allows the protons to spread in the perpendicular directions which weakens the hydrogen bonds [20, 157–159]. Many experiments have highlighted differences between H_2O and its heavier isotope D_2O for various thermodynamic properties. Since such isotope effects are entirely absent in MD simulations with classical nuclei (equilibrium static properties are independent of the atomic masses in the classical framework), these discrepancies provide a clear indication of the impact of NQEs. In particular, Vega et al. [17] have shown that only a quantum treatment can reproduce the heat capacity of liquid water and ice. Studies also show that the hydrogen bond network in

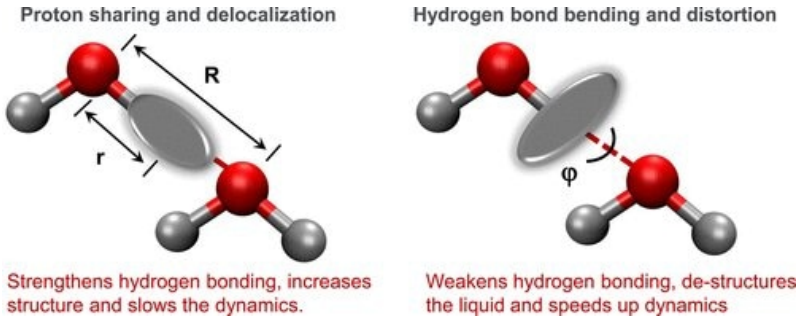


Figure 3.1: Schematic representation of the competing quantum effects in the hydrogen bonds between two water molecules which come from different contributions to the vibrational ZPE. The left panel illustrates the effect of O-H stretching ZPE, whereas the right one comes from the bending vibrational modes. As the distance R between the two oxygens decreases, the stretching contribution decreases while the bending contribution increases. Therefore, the two contributions strengthen and weaken the hydrogen bond respectively. Picture extracted from [20].

D_2O is stronger at 300 K giving a 20-30% lower diffusion coefficient for H_2O [20, 160, 161]. Harmonic or quasiharmonic approximation are often used to model NQEs in molecules or solids. However, in liquids, such as water, anharmonic and entropic effects must be taken into account and hence require a more extensive treatment of NQEs, for instance using the path integral formalism described in Chapter 2.

Needless to say, a realistic model of water is vital to understand its behavior under different conditions and environments, and to characterize its role in physicochemical phenomena. Such modeling should include a proper treatment of the nuclear motion at a quantum mechanical level with an accurate description of the interactions between the water molecules. To do so, numerous water models have been developed, ranging from very simple one, based on atomic partial charges and rigid bonds [162–164], to more sophisticated ones that include molecular flexibility [21, 165, 166] and explicit electronic polarization effects [31, 167–170]. Given the quantum nature of water, PIMD has been used to asses its quantum thermodynamic and structural properties. However, many of the water model which have been studied using quantum methods assume that the water monomers are rigid bodies [171, 172] and describe intramolecular flexibility with simple harmonic potential [166, 173]. Moreover, many studies that have assessed the role of quantum fluctuations in water have used empirical potential models which were parametrized on classical simulations and hence implicitly include NQEs. Therefore, it leads to a 'double counting' of the NQEs on structural and dynamical properties of the liquid. To address this, a new empirical water model, called q-TIP4P/F, was specifically parametrized to reproduce the liquid structure, diffusion coefficient and IR absorption spectra in quantum simulation using RPMD [30]. It is based on the TIP4P/2005 classical rigid water model that was chosen for its satisfactory description of the ice/water phase diagram in classical MC simulations [174].

In this chapter, we present the study of liquid water modeled by the q-TIP4P/F potential, including NQEs via the adQTB method presented in Chapter 2. Before this thesis, the adQTB had been tested on simple model systems, and shown encouraging results [27]. However, a complex and highly anharmonic system such as liquid water presents a considerably more stringent test for the stability and robustness of the method and allows to explore its limits. In particular, we will assess the efficiency of the adQTB to effectively correct the ZPEL that affects standard QTB simulations in this extremely anharmonic system. The different results will be compared to the one obtained with PIMD (and TRPMD for dynamical observables) throughout this Chapter.

All the developments and computations presented throughout this chapter were implemented and carried out using the Tinker-HP massively parallel software [28]. Tinker-HP is a high-performance version of the Tinker package [175] firstly developed by Jay Ponder at Washington University. It was essentially designed for experimenting, testing, creating FF, algorithms and models. However, the initial Tinker implementation was purely sequential and not competitive with other simulation programs such as NAMD [176] and GROMACS [177]. Therefore, Tinker-HP was designed with the most useful features of the original Tinker package to which was added high performance computations thanks to its MPI parallel structure.

3.2 The q-TIP4P/F water model

Most of the popular FF, such as AMBER [178–181] and CHARMM [182] are based on three-site pair-wise additive TIP3P water model [183]. TIP3P model was fitted to the experimental density and heat vaporization at room temperature and 1 atm pressure. However, this model often fails on properties outside

ambient conditions on which it was firstly parameterized [184]. Using long-ranged electrostatics and adding more interaction sites allow the development of new water models such as TIP4P/Ew [185] and TIP4P/2005 [164]. They provide a good agreement for a much broader range of thermodynamical observables outside the reference data used to parameterize them. Considering its correct predictions on various ice polymorphs, the TIP4P/2005 classical rigid water model was used by Habershon et al. as a basis to develop a new potential energy surface, q-TIP4P/F, for which the different thermodynamic properties used to fit the parameters were calculated using the PIMD and RPMD methods [30].

The TIP4P/2005 model has four interaction sites, three of them are placed at the oxygen and hydrogen atom. The last site, called M, is coplanar with the O and H sites and it is located at the bisector of the H-O-H angle. Moreover, the oxygen site carries no charge but contributes to the Lenard-Jones interaction:

$$E_{LJ} = 4\varepsilon \left[\left(\frac{\sigma}{r_{OO}} \right)^{12} - \left(\frac{\sigma}{r_{OO}} \right)^{6} \right] \tag{3.1}$$

with r_{OO} the distance between the oxygen sites of two molecules. The H and M sites are charged but do not contribute to E_{LJ}. The electrostatic potential is defined as:

$$E_{elec} = \frac{e^2}{4\pi\varepsilon_0} \sum_{a \in i} \sum_{b \in j} \frac{q_a q_b}{r_{ab}} \tag{3.2}$$

where e is the electron charge, ε_0 is the dielectric permittivity of the vacuum and a and b are the charged sites of molecules i and j. The distance between partial charge sites in molecules i and j are given by r_{ab}. Two positive charges $q_a/2$ are placed on the hydrogen atoms of each molecule and a negative one is placed at the site M, located at position $\mathbf{r}_a$, a fraction η along the vector connecting the O to the center of mass of the two H:

$$\mathbf{r}_a = \eta \mathbf{r}_O + (1-\eta) \frac{\mathbf{r}_{H_1} + \mathbf{r}_{H_2}}{2} \tag{3.3}$$

A quartic expansion of a Morse potential is used to add intramolecular flexibility to describe the O-H bond stretching. A simple harmonic potential is used for the bond angle:

$$\begin{aligned} V_{intra} &= \sum_i \left[V_{OH}(r_{i1}) + V_{OH}(r_{i2}) + \frac{1}{2} k_\theta (\theta_i - \theta_0)^2 \right] \\ V_{OH}(r) &= d_r \left[\alpha_r^2 (r - r_{eq})^2 - \alpha_r^3 (r - r_{eq})^3 + \frac{7}{12} \alpha_r^4 (r - r_{eq})^4 \right] \end{aligned} \tag{3.4}$$

with r_{i1} and r_{i2} the two O-H distance and θ_i the H-O-H bond angle in molecule i. The interaction potential is parametrized to optimize the different parameters involved in the potential: ε, σ, q_a and η for the intermolecular parameters and d_r, α_r, r_{eq} and θ_{eq} for the intramolecular ones. These parameters were obtained by using TRPMD simulations to fit structural, thermodynamic, and dynamic properties such as O-O, O-H and H-H radial distribution functions, the self-diffusion coefficient and infrared absorption spectra.

3.3 Study of the adQTB method in Liquid Water

To study the adQTB method on a highly harmonic system such as water, the q-TIP4P/F potential was used to model interactomic interactions as it has been broadly used in the literature as a benchmark for NQEs simulation methods. In this section we present some technical points about the adQTB that were added to the method during this thesis. These methodological details are illustrated in the case of liquid water but they are general and would be relevant for the application of the method to any system. The different results obtained were compared to PIMD and classical Langevin dynamics using a small time step $\Delta t = 0.2$ fs to ensure that the 'quantum' dynamics is correctly sampled with 1000 water molecules. All methods were included in the Tinker-HP package to see if the adQTB manages to correct the ZPEL from which the standard QTB suffers. Because static averages are independent of the friction force in both PIMD and classical Langevin dynamics, γ was chosen to $\gamma = 1\,\mathrm{ps}^{-1}$. However, this is not the case for the adQTB method. Therefore, the choice of this parameter will be further studied.

3.3.1 Spectral Deconvolution and Choice of γ

In the adQTB method, the first-kind FDT (equation (2.51)) is used to accordingly adjust the coefficients $\gamma_r(\omega)$. Although in classical Langevin MD and PIMD simulations static properties are independent of the parameter γ, the adQTB requires large friction coefficients to prevent the ZPEL. It causes an excess of energy at low frequency and if γ is too low, the friction forces cannot dissipate the excess energy, even if the corresponding random force coefficients $\gamma_r(\omega)$ reach 0. The friction force γ, however, tends to broaden the spectra. To suppress this broadening of the vibration spectrum, the harmonic reference was used with a deconvolution procedure to eliminate the effect of γ [186]. The deconvolution procedure assumes that the dipole velocity autocorrelation spectrum $\dot{\mathcal{C}}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$ can be expressed as:

$$\dot{\mathcal{C}}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega) = \int \mathrm{d}\omega' \mathcal{C}^{(0)}_{\dot{\mu}\dot{\mu}}(\omega')\mathcal{C}(\omega,\omega') \tag{3.5}$$

where $\mathcal{C}^{(0)}_{\dot{\mu}\dot{\mu}}(\omega')$ is the ideal vibration spectrum which should be obtained in a non-thermostated dynamics. $\mathcal{C}(\omega,\omega')$ is the Langevin convolution kernel derived for a harmonic reference:

$$\mathcal{C}(\omega,\omega') = \frac{2\gamma\omega^2}{\pi\left(\gamma^2\omega^2 + \omega^4 - 2\omega'^2\omega^2 + \omega'^4\right)} \tag{3.6}$$

which is valid in the case of a Markovian friction force with a friction coefficient gamma. One may ask how this relation can hold in real anharmonic systems given that the kernel $\mathcal{C}(\omega,\omega')$ was derived under the assumption of harmonic dynamics. It was shown in Ref. [186] by Rossi et. al that even for strongly anharmonic systems, the deconvolution procedure can be used to recover the unperturbed density of states, which is related to the velocity-velocity autocorrelation function with good accuracy. However, it should be used with caution in the zero-frequency range ($\omega \lessapprox \gamma$). If $\dot{\mathcal{C}}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$ is known, it is possible to compute the deconvolued spectrum $\mathcal{C}^{(0)}_{\dot{\mu}\dot{\mu}}(\omega)$ through an iterative procedure using the following recurrence formula [186]:

$$f_{n+1} = \frac{f_n(\omega)h(\omega)}{\int \mathrm{d}x \mathcal{D}(\omega,x)f_n(x)} \tag{3.7}$$

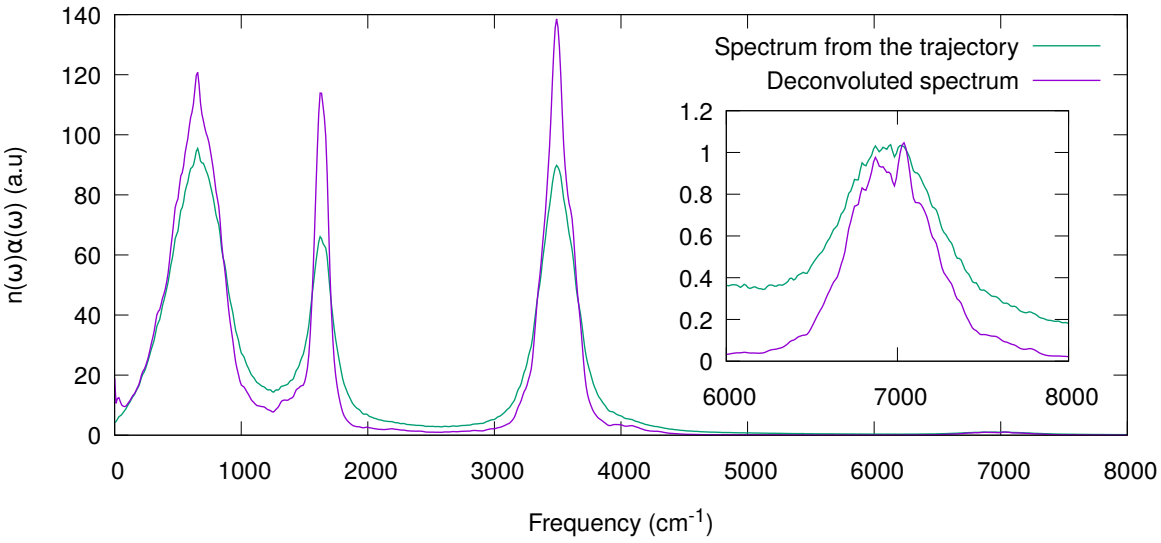


Figure 3.2: Infrared absorption spectrum of water at 300K obtained in adQTB simulation with $\gamma = 20\,\mathrm{ps}^{-1}$ and at fixed volume corresponding to a density $\rho = 0.995\,\mathrm{g.cm}^{-3}$. The figure shows the initial spectrum $\mathcal{C}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$ computed directly from the adQTB trajectory (green) and the deconvoluted IRS (purple) obtained after several iterations of the deconvolution procedure.

with $f_0(\omega) = \mathcal{C}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$ and:

$$\begin{aligned} h(\omega) &= \int \mathrm{d}x\, \mathcal{C}(x,\omega) f_0(x) \\ \mathcal{D}(\omega,x) &= \int \mathrm{d}y\, \mathcal{C}(y,\omega)\mathcal{C}(y,x) \end{aligned} \tag{3.8}$$

The procedure converges to a local solution verifying $\int \mathrm{d}x\, \mathcal{C}(\omega,x) f_\infty(x) = \mathcal{C}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$. Therefore, $f_\infty(\omega)$ can be used as an estimation of $\mathcal{C}^{(0)}_{\dot{\mu}\dot{\mu}}(\omega)$ since it verifies equation (3.5) (in practice, the iterative procedure should be stopped after a few tens of iterations in order to avoid over-fitting issues).

The deconvolution procedure was initially implemented to suppress the broadening due to the large friction coefficient γ in the adQTB and to recover unaffected infrared absorption spectra (IRS). The IRS is the product of the absorption coefficient $\alpha(\omega)$ and the refractive index $n(\omega)$ which is proportional to the Kubo-transformed dipole derivative autocorrelation spectrum [187]. In adQTB, as in linearized semiclassical methods, the autocorrelation spectrum obtained by direct Fourier transform of the adQTB dipole trajectory corresponds to the real part of the standard autocorrelation function given by equation (2.30) [188]. Therefore, it is multiplied by $\frac{\tanh(\beta\hbar\omega/2)}{\beta\hbar\omega/2} = \frac{k_B T}{\theta(\omega,T)}$ to obtain the Kubo autocorrelation that characterizes the IRS. The initial spectrum $\mathcal{C}^{\mathrm{Lang}}_{\dot{\mu}\dot{\mu}}(\omega)$ obtained in adQTB dynamic with $\gamma = 20\,\mathrm{ps}^{-1}$ is shown in Figure 3.2 with its deconvoluted spectrum $\mathcal{C}^{0}_{\dot{\mu}\dot{\mu}}(\omega)$ obtained after 50 iterations. As expected, the deconvoluted spectrum in purple yields sharper peaks compared to its initial spectrum in green.

3.3.2 Finite Gamma Error Correction

An important specificity of QTB (or adQTB) simulations is that the value of the coefficient gamma does not only affect dynamical observables, such as the vibration spectra considered previously, but also static properties, contrary to classical Langevin simulations where the equilibrium probability density distribution corresponds to the Boltzmann canonical density (and hence independent of the value of the friction coefficient). Even for harmonic systems, the average potential and kinetic energies obtained in the QTB do not exactly correspond to their quantum expected values but are subject to a small error that tends to increase with gamma. This error, which is unrelated to ZPEL, can be predicted and quantified for a harmonic oscillator at frequency ω_0 where the average potential and kinetic energies are given by [189, 190]:

$$\begin{aligned} \langle E_{pot}(\omega_0)\rangle &= \int_0^{\omega_{cut}} \frac{\mathrm{d}\omega}{\pi} \frac{\gamma\omega_0^2}{\left(\omega^2-\omega_0^2\right)^2+\gamma^2\omega^2}\theta(\omega,T) \\ \langle E_{kin}(\omega)\rangle &= \int_0^{\omega_{cut}} \frac{\mathrm{d}\omega}{\pi} \frac{\gamma\omega^2}{\left(\omega^2-\omega_0^2\right)^2+\gamma^2\omega^2}\theta(\omega,T) \end{aligned} \tag{3.9}$$

with ω_{cut} a frequency cutoff introduced to avoid the divergence of the kinetic energy. In the limit $\gamma \longrightarrow 0$, the Lorentzian tends to δ-function and both integrals have the same value: $\langle E_{pot}(\omega_0)\rangle \simeq \langle E_{kin}(\omega_0)\rangle \simeq \theta(\omega,T)/2$. However, for finite γ, the spectral broadening of the Lorentzian function modifies these averages and the potential energy tends to be underestimated whereas the kinetic energy is overestimated. To address this inaccuracy of the (ad)QTB methods, the colored random force given by equation (2.49) is modified such that

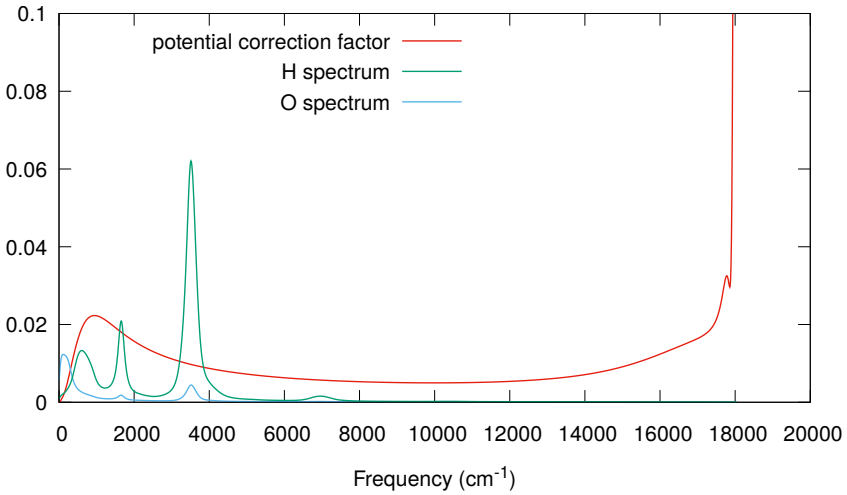


Figure 3.3: Correction factor $\left[\tilde{\theta}(\omega,T)-\theta(\omega,T)\right]/\theta(\omega,T)$ as a function of the frequency ω for $\gamma = 20$ ps^{-1} and $\omega_{cut} = 3400$ ps^{-1}. The average power spectrum $C_{vv}(\omega)$ obtained for each atom type with the q-TIP4/F. water model is also represented in arbitrary units for comparison. The sharp increase at the cutoff frequency ($\simeq 18000$ cm^{-1}) is an artefact of the iterative procedure but it has no relevant impact on the simulation results.

$\theta(\omega, T)$ is replaced by $\tilde{\theta}(\omega, T)$ in order to verify:

$$\begin{aligned} \left\langle \tilde{E}_{pot}(\omega_0) \right\rangle &= \int_0^{\omega_{cut}} \frac{\mathrm{d}\omega}{\pi} \frac{\gamma\omega_0^2}{\left(\omega^2 - \omega_0^2\right)^2 + \gamma^2\omega^2} \tilde{\theta}(\omega, T) \\ &= \frac{\theta(\omega_0, T)}{2} \end{aligned} \tag{3.10}$$

This relation, which defines $\tilde{\theta}(\omega, T)$, has a similar structure as equation (3.5). Therefore, the deconvolution procedure, previously used to correct the IRS, was also used to slightly modify $\theta(\omega, T)$ to compensate the effect of γ on the potential energy. This correction of the potential energy is represented in Figure 3.3 with $\gamma = 20\,\mathrm{ps}^{-1}$ and $\omega_{cut} = 3400\,\mathrm{ps}^{-1}$. It should be noted that the frequency cutoff ω_{cut} should be large enough to enclose all relevant frequencies, including overtones, to obtain good average stretching energies and then compute the potential correction factor. Over the whole relevant frequency range, the potential correction factor is in the order of 1% to 2%.

To also study the choice of γ, the different potential energy terms in the q-TIP4P/F model are plotted for different values of γ ranging from $\gamma = 10\,\mathrm{THz}$ to $44\,\mathrm{THz}$ in Figure 3.4 with the different methods. One may notice the differences between the PIMD - 64 beads and the QTB with $\gamma = 20$ THz. However, thanks to the first-kind FDT used as a criterion (equation (2.51)) in the adQTB, the different energy terms in this method get closer to the references values. As expected the energies depend weakly on γ where the most important variations are on the intermolecular energies. It can be partly explained by the harmonic reference used in equation (3.9) to compute $\tilde{\theta}(\omega, T)$ which is less accurate for the anharmonic low frequency modes. An other possible explanation could be the presence of small residual ZPEL, that slowly vanishes when γ is increased. Indeed, in the adaptation process, Δ_{FDT} is averaged over all equivalent atoms and directions and does not depend on the local molecular modes. To correct the overestimation of the average kinetic energy, the previously defined deconvolution method is applied to the average power spectrum $\mathcal{S}^{\mathrm{Lang}}(\omega)$ as:

$$\mathcal{S}^{\mathrm{Lang}}(\omega) = \sum_i \frac{1}{2} m_i \mathcal{C}_{v_i v_i}(\omega) \tag{3.11}$$

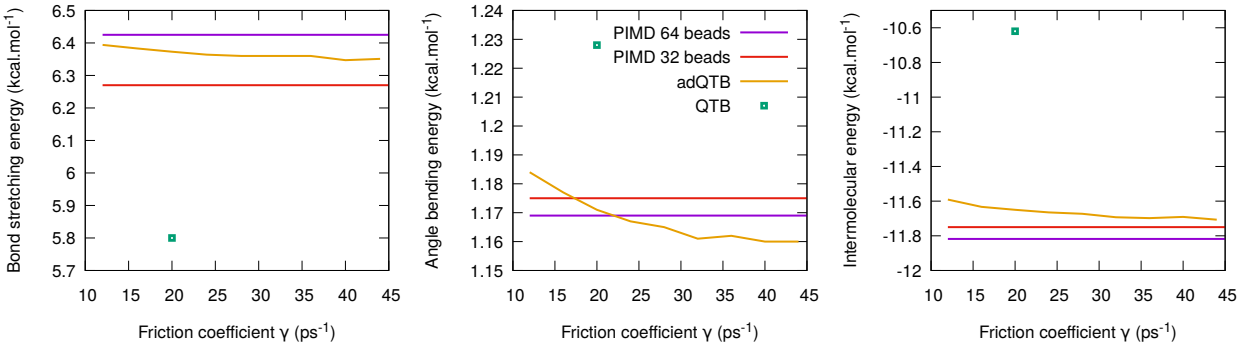


Figure 3.4: Thermodynamical averages of the different energy terms per water molecules of the q-TIP4P/F water model with different value of γ computed in NVT at 300 K and $\omega_{cut} = 3400\,\mathrm{THz}$. The intermolecular energy is the sum of the Coulomb and vdW interactions. In classical and PIMD, the average energies are independent of γ are were obtained with $\gamma = 1\,\mathrm{ps}^{-1}$ and are also represented by horizontal lines.

Method	adQTB										PIMD - 64 beads
ω_{cut} (THz)	1500					3400					
γ (ps^{-1})	12	16	20	24	28	12	16	20	24	28	
E_k (kcal.mol^{-1})	8.52	8.48	8.45	8.43	8.40	8.64	8.61	8.60	8.59	8.58	8.55

Table 3.1: Average kinetic energy with different value of ω_{cut} and γ. Because the PIMD method does not depend on γ and ω_{cut} no value is written.

where index i runs over all atomic degrees of freedom and $\mathcal{C}_{v_i v_i}(\omega)$ is the velocity-velocity autocorrelation function computed from (ad)QTB trajectories via discrete Fourier transform. Before applying the deconvolution procedure, $\mathcal{S}^{\text{Lang}}(\omega)$ is divided by the kernel $\tilde{\theta}(\omega, T)$. Then the deconvoluted spectrum is re-multiplied by the exact density $\theta(\omega, T)$ to yield the corrected power spectrum $\mathcal{S}^0(\omega)$. The corrected kinetic energy is given by the integral of the corrected power spectrum $\mathcal{S}^0(\omega)$.

In equation (3.9) a cutoff frequency was introduced to avoid the divergence of the kinetic energy. Hence, it has an effect on the average kinetic energy as shown in Table 3.1. Including the overtone by increasing the cutoff frequency from $\omega_{cut} = 1500\,\text{THz}$ to $\omega_{cut} = 3400\,\text{THz}$ reduces the absolute difference between the reference PIMD - 64 beads value and adQTB with different values of γ. The values which are obtained with $\omega_{cut} = 1500\,\text{ps}^{-1}$ are generally slightly underestimated because this value of ω_{cut} is almost equal to the frequency of the stretching overtone, whose contribution to the vibrational energy is therefore only partially

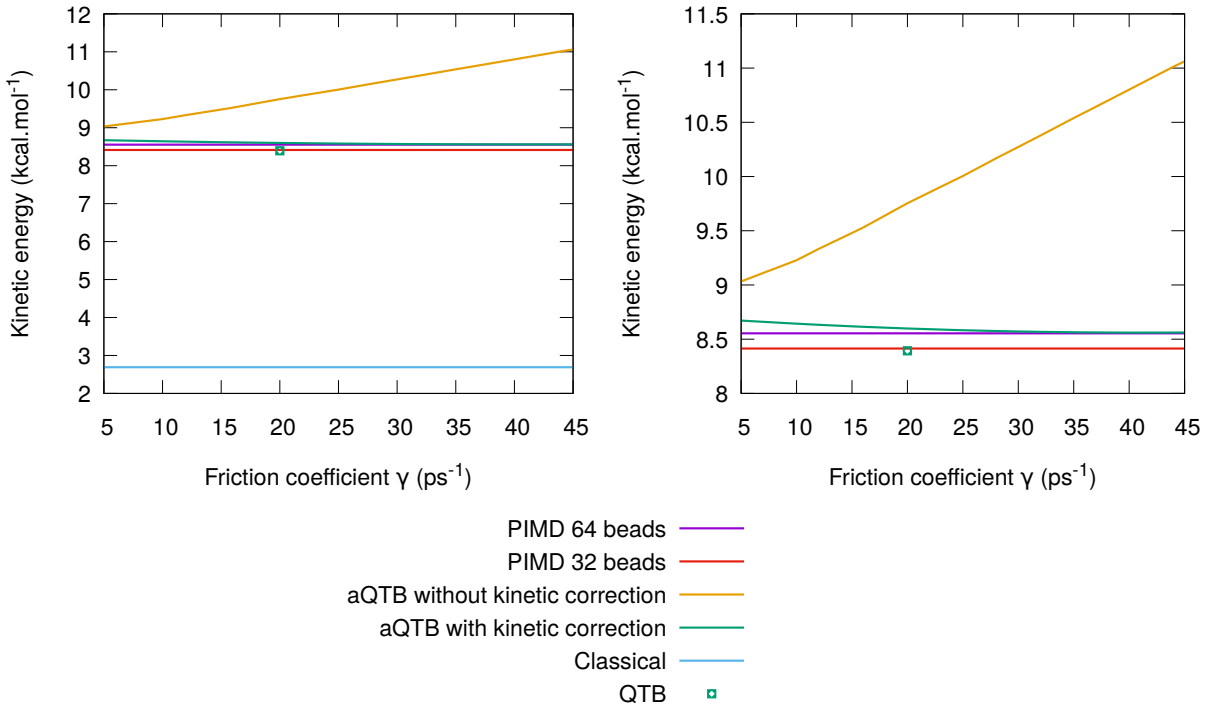


Figure 3.5: Plot of the average kinetic energy as a function of the friction coefficient γ computed in the NVT ensemble at 300K with $\omega_{cut} = 3400\,\text{THz}$. In classical and PIMD methods, the average energies are independent of γ and were obtained with $\gamma = 1\,\text{ps}^{-1}$. Therefore, they are represented by horizontal lines. The left plot is a focus on the PIMD and QTB average kinetic energies.

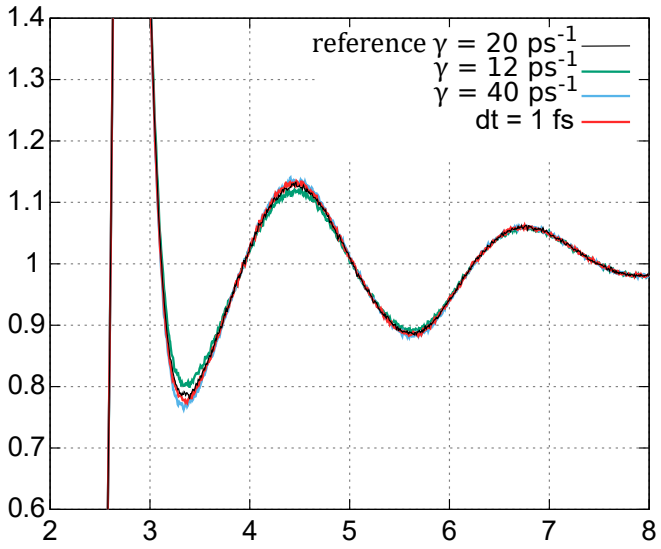


Figure 3.6: Plot of the Oxygen-Oxygen RDF $g_{OO}(r)$ as a function of the atomic distance in Åwith different values of γ and the same cutoff frequency $\omega_{cut} = 3400$ THz. The red curve is obtained using $\gamma = 20$ ps^{-1} but with an increased time step $\Delta t = 1$ fs. The scale is chosen in order to highlight the small differences between the curves.

accounted for in that case. To investigate further the choice of γ, the average kinetic energy was plotted in Figure 3.5 at fixed value of $\omega_{cut} = 3400\,\text{THz}$ within a larger fiction coefficient range with the different methods. As expected, the average kinetic energy in classical MD is dramatically underestimated compared to the PIMD references due to the lack of ZPE. However, including NQEs with the (ad)QTB methods allows to recover an average kinetic energy close to the PIMD references within a large range of γ. The adQTB recovers the PIMD - 64 beads reference, whereas the standard QTB is closer to the PIMD - 32 beads value (note that in both QTB and adQTB, the kinetic energy is corrected for finite gamma errors). Furthermore, without the kinetic correction, the Lorentizian function, defined in equation (3.9), is broadened resulting in an overestimated average kinetic energy (orange curve in Figure 3.5). Moreover, in adQTB, the kinetic correction increases from 8% to 29% when γ varies from 12 ps^{-1} to 44 ps^{-1} showing that the kinetic correction is not negligible. It allows to recover an almost γ-independent estimation of the kinetic energy with less than 1% variation over the whole range of γ.

Looking at the different corrections presented in Figure 3.4 and 3.5, a value of the friction coefficient γ can be set. It is important to choose it high enough so that the adapted value of γ_r does not reach 0, in which case adQTB would not manage to entirely eliminate ZPEL. This is the case for all the values of γ presented in Fig 3.4. Moreover, γ should not also be fixed arbitrary high due to its effect on the dynamics (and more specifically on the self-diffusion coefficient, see below). Therefore, the friction coefficient was set to $\gamma = 20\,\text{THz}$ for all results presented throughout the thesis. It should also be noticed that thanks to the potential and kinetic corrections applied while using the adQTB method, the different results obtained are almost γ-independent and are also robust with respect to changes of this parameter, as it can be seen on the Radial Distribution Function (RDF) O-O in Figure 3.6. The first peak is slightly affected by the changes of the parameters. The sharper peak observed with the increased value of γ is consistent with the trend observed for the intermolecular energies in Figure 3.4.

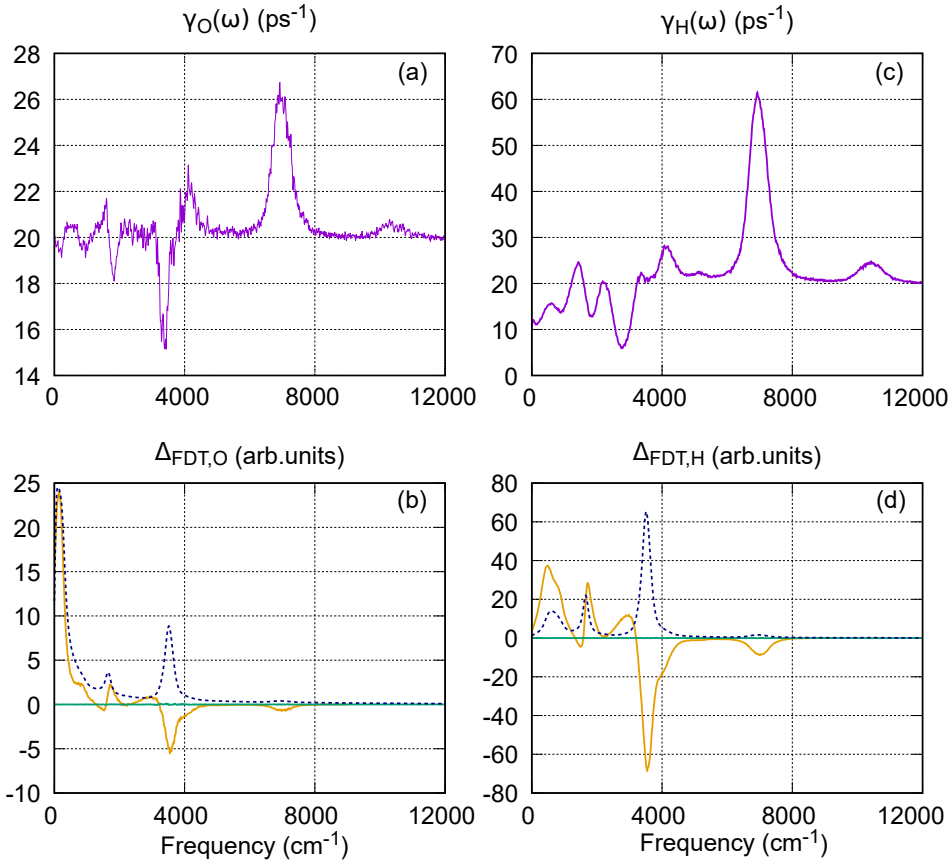


Figure 3.7: Plots (a) and (c) are the converged $\gamma_O(\omega)$ and $\gamma_H(\omega)$ in ps^{-1}. Plots (b) and (d) are the corresponding $\Delta_{FDT}(\omega)$ in arbitrary units using the QTB (orange) and adQTB methods (green). All plots were performed at 300 K at a fixed volume corresponding $\rho = 0.995\,\mathrm{g.cm}^{-3}$. The average power spectrum $\mathcal{C}_{vv}(\omega)$ is represented as a blue dashed line in arbitrary units.

3.3.3 Adaptation Procedure

In the adQTB method, a different $\gamma_{r,\alpha}(\omega)$ coefficient is associated to each atomic species and can be adjusted to enforce the first-kind FDT where α is either O or H. The trajectory is decomposed into equal segments of duration $\tau = 1\,\mathrm{ps}^{-1}$. At the end of each segments, the spectra $\mathcal{C}_{v_i v_i}(\omega)$ and $\mathcal{C}_{v_i F_i}(\omega)$ are estimated from the previous segment for each degree of freedom via a discrete Fourier transform of the trajectory during the past segment. Then, the $\gamma_{r,\alpha}(\omega)$ values are adapted for each frequency and each atom type to reduce $\Delta_{FDT,O}(\omega)$ and $\Delta_{FDT,H}(\omega)$ using equation (2.51) and averaging over the three spatial directions and same atom species. To start our simulations, the initial value of $\gamma_{r,\alpha}(\omega)$ was fixed at $\gamma_{r,\alpha}(\omega) = \gamma$ and they are adapted for all discrete values $\omega_j = 2\pi j/\tau$, inferior to the cutoff angular frequency ω_{cut} as:

$$\gamma_{r,\alpha}^{(n+1)}(\omega) = \gamma_{r,\alpha}^{(n)} + \gamma A_\gamma \tau \frac{\Delta_{FDT,\alpha}}{||\Delta_{FDT,\alpha}||} \tag{3.12}$$

with A_γ the adaptation velocity coefficient. This term allows to ensure that the adaptation should not be too fast such that the adaptation time is much larger than τ (with $\tau \geq \gamma^{-1}$). A too high value can reduce the accuracy of the method by adding excessive noise to $\gamma_{r,\alpha}$. The normalisation factor is given by:

$$||\Delta_{FDT,\alpha}|| = \sqrt{\sum_{\omega_j} \Delta_{FDT,\alpha}(\omega_j)^2} \tag{3.13}$$

The random force $F^{n+1}(t)$ are then generated using the kernel $\gamma_{r,\alpha}^{(n+1)}(\omega)$. The procedure is iterated until $\gamma_{r,\alpha}(\omega)$ converges. The adaptation process is kept active throughout the whole simulations so that $\gamma_{r,\alpha}$ continues to fluctuate smoothly.

The different converged $\gamma_{r,\alpha}$ are shown in the top panel of Figure 3.7. They were obtained at 300 K and at constant volume corresponding to a density $\rho = 0.995\,\mathrm{g.cm^{-3}}$. Due to its low mass, the $\gamma_{r,H}$ coefficient for hydrogen atoms displays a higher variation in its value. Some oscillations around 1000-2000 $\mathrm{cm^{-1}}$ are visible and correspond to the angle bending frequencies. It can be explained by the energies received by the stretching modes but also by the leaking to lower frequency ones. These two phenomenon are both competing, and looking at Δ_{FDT}, the adQTB manages to correct this leakage highlighting the robustness of the method. Moreover, for the intermolecular motions, $\gamma_H(\omega)$ is reduced to compensate for the ZPEL. It is also increased around the stretching frequency $\omega_{str} \approx 3300\,\mathrm{cm^{-1}}$, whereas $\gamma_O(\omega)$ counter-intuitively shows a slight decrease. For both atoms, a sharp peak is noticed at the stretching overtone frequency. Therefore, the ZPEL is corrected thanks to the relation $\Delta_{FDT}(\omega) = 0$, enforcing the quantum distribution of energy for each nuclear degree of freedom as it can be noticed on the bottom panel of Figure 3.7 for both atom types. As expected, to correct the ZPEL, $\gamma_{r,\alpha}(\omega)$ is reduced where an excess of energy occurs whereas it is increased for negative deviations. Then, $\Delta_{FDT}(\omega)$ vanishes on average.

3.3.4 Adaptation Velocity

In equation (3.12) an adaptation velocity coefficient A_γ appears for the random forces kernel. The duration of the adaptation procedure in the adQTB method can be adjusted thanks to this parameter. The adaptation time, however, should be greater than τ (with $\tau > \gamma^{-1}$) in order to avoid large fluctuations in $\gamma_{r,\alpha}(\omega)$ which can reduce the accuracy of the adQTB method. The different results presented in this thesis were obtained with a small $A_\gamma = 0.1\,\mathrm{ps^{-1}}$ to reduce the fluctuations of $\gamma_\alpha(\omega)$. With this value, the coefficient $\gamma_{r,\alpha}(\omega)$ converged over most of the frequency range after 100 ps of adaption process. However, the lowest frequency (below 100 $\mathrm{cm^{-1}}$) and the stretching overtones required a longer adaptation time. Figure 3.8 shows that

	Bond stretching	Angle bending	Intermolecular energies	Kinetic energy
Standard QTB	5.81	1.2	-10.67	8.39
adQTB	6.37	1.17	-11.65	8.60
adQTB, Δt=0.5	6.34	1.17	-11.66	8.50
Fast adaptation	6.38	1.16	-11.66	8.69
average FDT	6.35	1.19	-11.47	8.58
PIMD - 64 beads	6.43	1.18	-11.75	8.55

Table 3.2: Average potential and kinetic energy terms. Fast adaptation is obtained using $A_{\gamma,O} = 1\ \mathrm{ps^{-1}}$, $A_{\gamma,H} = 10\ \mathrm{ps^{-1}}$ and adaptation segments of 100 fs, while average FDT procedure is obtained when $\Delta_{FDT}(\omega)$ is averaged over both atomic types.

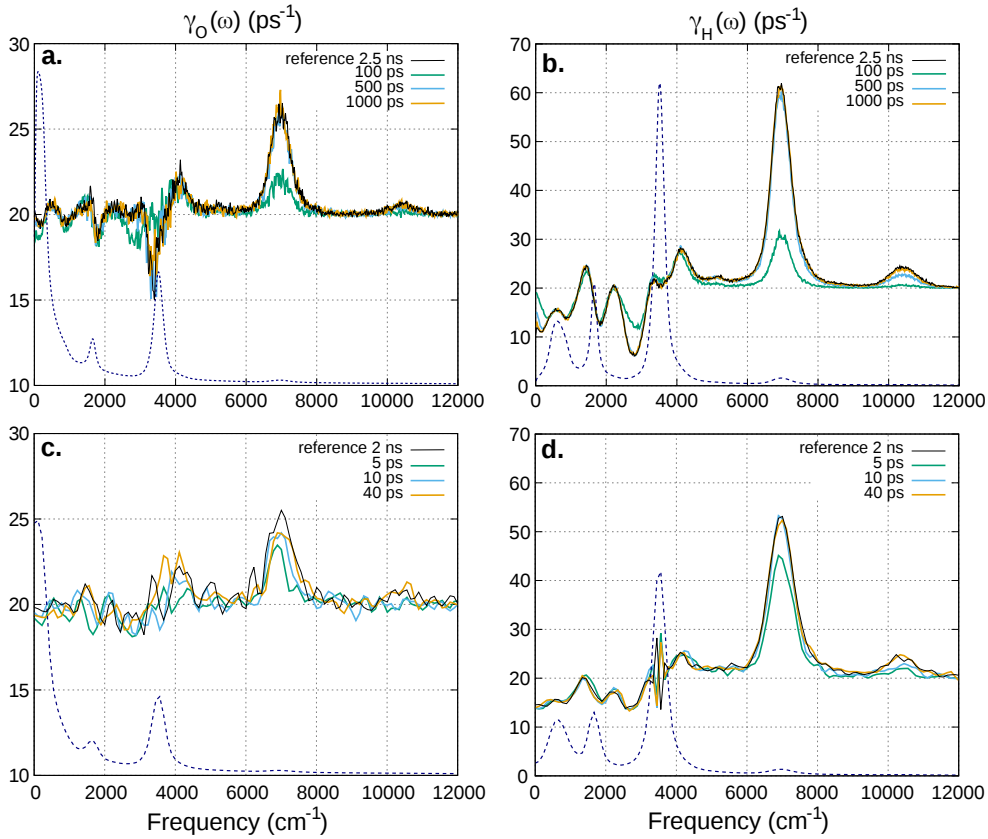


Figure 3.8: Plot of the adapted $\gamma_{O,H}(\omega)$ coefficients at different instants of the simulation, showing the convergence of the adaptation process. In each case a fully converged reference curve is represented in black. For the top panels **a** and **b**, we use an adaptation velocity coefficient $A_\gamma = 0.1$ ps^{-1} and a segment length of 1 ps, while in the bottom panels **c** and **d**, we boosted the adaptation process by using $A_{\gamma,O} = 1$ ps^{-1}, $A_{\gamma,H} = 10$ ps^{-1} and a segment length $\tau = 100$ ps. In each figure, the corresponding average power spectrum $C_{vv}(\omega)$ is represented (in arbitrary units) as a dashed blue line for comparison. Note that in the bottom panels, the spectral features are slightly broadened with respect to the top panels due to the use of a smaller segment length τ.

the coefficients converged after 500 ps. It is also possible to reduce the length of the adaption procedure by changing $\tau = 1$ ps to $\tau =$100 fs and using $A_{\gamma,O} = 1\,\text{ps}^{-1}$ and $A_{\gamma,H} = 10\,\text{ps}^{-1}$. The convergence is then reached in about 10 ps. Although using a higher value of $A_{\gamma,\alpha}$ increases the fluctuations of the corresponding $\gamma_{r,\alpha}(\omega)$, it does not affect the average potential energy contributions as it can be noticed in Table 3.2. The kinetic energy is slightly overestimated but it can be explained by the smaller τ which decreases the spectral resolution $2\,\pi/\tau$ and therefore affects the accuracy of the deconvolution procedure.

One may ask why different γ_r were used for each atom type. To answer this question, simulations were performed in which Δ_{FDT} is averaged over both atomic types so that $\gamma_O(\omega) = \gamma_H(\omega)$. Although $\gamma(\omega)$ converged, some ZPEL remains. Moreover, the ZPEL has an opposite trend on the converged $\gamma(\omega)$ of O and H, showing that the mean $\Delta_{FDT}(\omega)$ is equal to zero even if some leakage is still present. As a consequence of this remaining ZPEL, the RDFs of water is less structured than when different friction coefficients where used for O and H. This is illustrated by the O-O RDFs in Figure 3.9. The results obtained while using the average adaptation method indicated as *average FDT* in Table 3.2 also present a slightly higher average intermolecular energy compared to the reference adQTB simulations. Hence, a less structured RDF O-O is

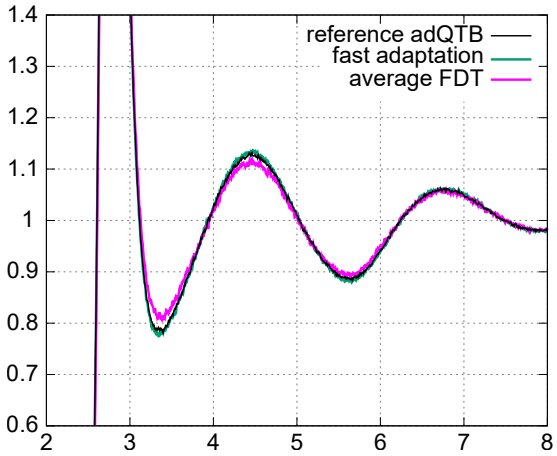


Figure 3.9: Oxygen-oxygen RDF $g_{OO}(r)$ as a function of the atomic distance in Å with γ=20 ps^{-1} and $\omega_{cut} = 3400$ Thz. The green curve labeled *fast adaptation* is obtained using $A_{\gamma,O} = 1$ ps^{-1}, $A_{\gamma,H} = 10$ ps^{-1} and adaptation segments of 100 fs, while the purple curve corresponds to the *average FDT* procedure, where $\Delta_{FDT}(\omega)$ is averaged over both atomic types. The vertical scale is chosen in order to highlight the small differences between the curves.

expected as noticed in Figure 3.9. Although the difference is small and the averaged Δ_{FDT} still allows to eliminate the major part of the leakage, the figure illustrates the improvement achieved via atom-dependent $\gamma_\alpha(\omega)$ coefficients to better correct the ZPEL. Therefore, while using this method on inhomogeneous systems (such as molecules in solution or structural interface), an independent set of $\gamma_\alpha(\omega)$ for each atom type in the system should be used to ensure the efficient compensation of the ZPEL.

3.3.5 Pressure Estimation and Langevin Barostat

One may be interested to perform simulation in the NPT ensemble to obtain different observables such as the density ρ or the enthalpy of vaporization ΔH_{vap}. To do so, a Langevin piston method was implemented

adQTB			
240 K		300 K	
Kinetic energy	Pressure	Kinetic energy	Pressure
9.81/8.88	1231/-138.2	9.38/8.50	1253/-92.89
PIMD - 32 beads			
240 K		300 K	
Kinetic energy	Pressure	Kinetic energy	Pressure
8.06	32.51	8.41	-10.8

Table 3.3: Average energies and pressure estimator while using the adQTB and PIMD methods at two different temperatures and constant volume corresponding to the density $\rho = 0.995\,\mathrm{g.cm^{-3}}$. In the adQTB the first value in each observable does not use the kinetic correction whereas the second value used the kinetic correction. The kinetic energy is given in kcal.mol^{-1} and the pressure in atm.

adQTB			
240 K		300 K	
Kinetic pressure	Potential term	Kinetic pressure	Potential term
-13.99/12.58	-12.76/-12.72	14.26/12.91	-13.01/-13.00
PIMD - 32 beads			
240 K		300 K	
Kinetic pressure	Potential term	Kinetic pressure	Potential term
12.27	-12.23	12.78	-12.79

Table 3.4: Values of the kinetic pressure and potential term part of equation (3.14) at two different temperatures. The potential term refers to the second part of the equation. In the adQTB the first value in each observable does not use the kinetic correction whereas the second one is given with it. All values are given in kcal.mol^{-1}.Ang^{-3}.

in Tinker-HP [191–193] for which the isotropic pressure estimator is given by:

$$P_{int} = \frac{2K}{3V} - \frac{\mathrm{d}U}{\mathrm{d}V} \tag{3.14}$$

with V the simulation box volume and U the interatomic potential energy whose derivative with respect to V is computed via finite difference. The first part of the equation is designated as the kinetic pressure whereas the second part is the potential term. To compute the kinetic energy, the centroid virial estimator was used in PIMD while in (ad)QTB the estimator used is defined as:

$$K(v_1, \ldots, v_{3N}) = \eta^{-1} \sum_1 \frac{1}{2} m_i v_i^2 \tag{3.15}$$

where η is the kinetic correction factor obtained through the deconvolution procedure and defined as the ratio of the raw kinetic energy to the corrected one. Neglecting the kinetic correction would result in an overestimation of the pressure of about 1000 atm at 240 K and 300 K as it can be seen on Table 3.3. It would translate in an error of almost 5% on the evaluation of the density ρ. This sensitivity to the correction arises because both terms of equation (3.14) almost cancel each other (Table 3.4) so that even a small error on one of these terms can cause a non negligible error on the density calculation. Thus, the pressure depends critically on the kinetic pressure estimation and therefore the kinetic correction factor η which is computed via the deconvolution procedure.

In the Langevin piston method, the simulation box volume varies according to the following equation [192]:

$$\begin{aligned} \dot{V} &= 3V\alpha \\ \dot{\alpha} &= \frac{3}{\mu}\left[V\left(P_{int} - P_{ext}\right) + k_B T\right] - \gamma_V \alpha + \sqrt{\frac{2\gamma_V k_B T}{\mu}} R(t) \end{aligned} \tag{3.16}$$

with α the piston velocity, μ the piston mass and $R(t)$ a gaussian white noise. This is a Langevin dynamic for the volume aiming at imposing the appropriate constant pressure with a probability density proportional to $e^{-\beta\left(P_{ext}V + \frac{1}{2}\mu\alpha^2\right)}$. In (ad)QTB dynamics, the volume degree of freedom couples to the atomic degrees

240K				
μ (g.Å^2)	10^5	10^6	2.5^6	10^7
T_{piston} (K)	346.76	264.45	254.13	243.62
ρ (g.cm^{-3})	0.999	1.000	1.000	1.000
300K				
μ (g.Å^2)	10^5	10^6	2.5^6	
T_{piston} (K)	389.39	313.38	310.51	
ρ (g.cm^{-3})	0.997	0.997	0.997	

Table 3.5: Values of the density and piston temperatures with different values of the piston mass μ.

of freedom which have an effective temperature higher than the physical temperature T. Therefore, the piston temperature $\langle \alpha^2/k_B\mu \rangle$ can be higher than T which is not physical. To avoid the overestimation of the piston temperature, a large mass μ and a small piston friction γ_V can be used. It slows down the volume dynamics and reduces the coupling to the high frequency modes. As a consequence, the volume couples essentially with the slow molecular motion whose effective temperature in adQTB is close to k_BT as: $\theta(\omega \longrightarrow 0, T) = k_BT$. Different values of μ were used to study the effect of the overestimation of the piston temperature on the density and are summarized in Table 3.5. Increasing the piston mass allows to recover the correct piston temperature which becomes higher of only a few K than T without affecting the values obtained for the density. However, increasing μ slows down the piston dynamics and because it does not have a significant effect on the density, the results presented throughout this chapter were obtained with $\mu = 10^5\,\text{g.Å}^2$. Nonetheless, it should be noted that increasing μ allow to enforce not only the correct average volume, but also the correct volume fluctuations which can be important for other observables such as the isothermal compressibility. Despite this, the volume fluctuates fast enough for an efficient sampling of the density considering that only a few ns are required to reach statistical convergence. It is also in principle possible to guarantee that the piston temperature is equal to T by changing the white noise $R(t)$ of equation (3.16) for a colored noise by adapting the adQTB procedure to enforce the classical FDT on α (though this approach was not necessary here and was not implemented).

3.4 Results on the q-TIP4P/F Potential

The role of NQEs in liquid water has been a long work interest both experimentally and theoretically [194–198]. The ability of the adQTB to capture the subtle balance between the stretching and bending ZPE highlights the robustness of the method. Therefore, after the study of the influence of the different parameters and corrections of the adQTB method presented in Section 3.3, we present here the main results of this study with a choice of relevant thermodynamical observables. Since static averages are independent of the parameter γ, it was chosen to be equal to $\gamma = 1\,\text{ps}^{-1}$ in classical Langevin MD and PIMD (and TRPMD) simulations, whereas for the (ad)QTB methods it was chosen to be equal to $\gamma = 20\,\text{ps}^{-1}$. Simulations were essentially converged with 32 beads and with a small time step of Δt =0.2 fs (cf Table 3.2).

	E_k	E_p	AB	BS	vdW	Coulomb	r_{OH} (Å)	θ_{HOH}(deg)
classical	2.69	-10.22	0.41	1.18	2.20	-14.00	0.96	104.8
QTB	8.39	-3.63	1.23	5.81	1.72	-12.38	0.98	104.6
adQTB	8.60	-4.11	1.17	6.37	2.11	-13.76	0.98	104.7
PIMD	8.41	-4.29	1.17	6.26	2.15	-13.87	0.98	104.7

Table 3.6: Physical observables obtained with the different methods at 300 K. E_k and E_p denote the kinetic and the total potential energy, while AB, BS and VdW and Coulomb refer to the different energy terms in the q-TIP4P/F potential, respectively for molecular Angular Bending, Bond Stretching, van der Waals interaction and direct Coulomb interaction. All energies are given in kcal.mol^{-1} per water molecule and obtained from 1 ns simulations (preceded by 1.5 ns adaptation for the adQTB) at constant volume corresponding to the density $\rho = 0.995$ g.cm^{-3}. The standard errors on the energy values are all inferior to 0.01 kcal.mol^{-1} per molecule. The table also shows the averages of the oxygen-hydrogen distance r_{OH} and of the molecular angle θ_{HOH}.

3.4.1 Structural Properties of Water

Radial Distribution Function

In Figure 3.10 the different RDFs obtained with all the methods are presented. Comparing classical MD with their quantum counterparts (PIMD and (ad)QTB curves) it appears that quantum RDFs are less structured. This is a consequence of the inclusion of NQEs that slightly modifies the local water structure, which in turn affects other properties such as the density. The intramolecular peaks are much broader in quantum simulations due to strong ZPE effects on intramolecular modes. The ZPEL affecting these intramolecular modes destabilizes the hydrogen bond network in QTB simulations. Hence, the different intermolecular peaks are broadened. In contrast, the adQTB suppresses the leakage and the corresponding curves almost superimpose with the PIMD reference thanks to the first-kind FDT used as a criteria.

Average Energies Terms

The impact of NQEs on the different equilibrium properties and energies of water at ambient conditions were further analyzed and are reported in Table 3.6. As expected, the intermolecular energies (vdW and

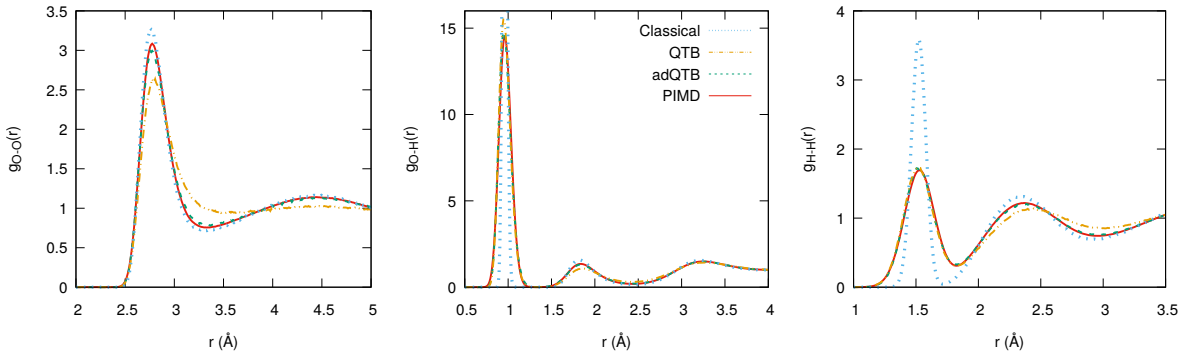


Figure 3.10: Radial distribution function at 300 K and constant volume corresponding to the density $\rho = 0.995\,\mathrm{g.cm^{-3}}$ using the different methods.

Coulomb) are only slightly affected by NQEs and classical and PIMD values are close. However, in QTB simulation, the intermolecular energies are overestimated by more than 1 kcal.mol^{-1} due to the ZPEL. This error is well corrected by the adQTB that recovers values comparable to the PIMD ones. The adQTB is very accurate on the intramolecular energy terms (angle bending and bond stretching), but also on the kinetic (which comprises large amounts of ZPE) and potential energies thanks to the different correction procedures added to the adQTB method and presented in Section 3.3. The adQTB method also manages to capture the elongation of the OH distance induced by NQEs while the angle is unaffected.

3.4.2 Infrared Absorption Spectra

Although PIMD gives a numerically exact reference for static quantum properties, the computation of dynamical observables, such as the infrared absorption spectra (IRS), represents a theoretical challenge and it is a subject of intense research [135, 141, 199–201]. There is no reference method to compute IRS exactly while accounting for NQEs in large systems but several approximations have been developed [127, 128, 202]. Benson et al. have recently compared different state-of-the-art approximate methods to calculate IRS in liquid water and ice [187]. They showed that the linearized semiclassical initial value representation (LSC-IVR) gives the most accurate IRS whereas TRPMD is presented as the cheapest available approach to yield reliable result. Although not formally derivable from first principles, the QTB method has already been used to compute IRS and provided relevant results in comparison with experiments [15, 112]. Moreover, the short time dynamics is only slightly affected by the thermostat and it is essentially classical, but the sampling of the quantum space allows to explore configurations which would not be possible in classical MD. Therefore, much like LSC-IVR, the QTB combines classical dynamics with approximate quantum initial value sampling. Moreover, with the deconvolution procedure presented in Section 3.3.1 the main effect of the thermostat is efficiently suppressed.

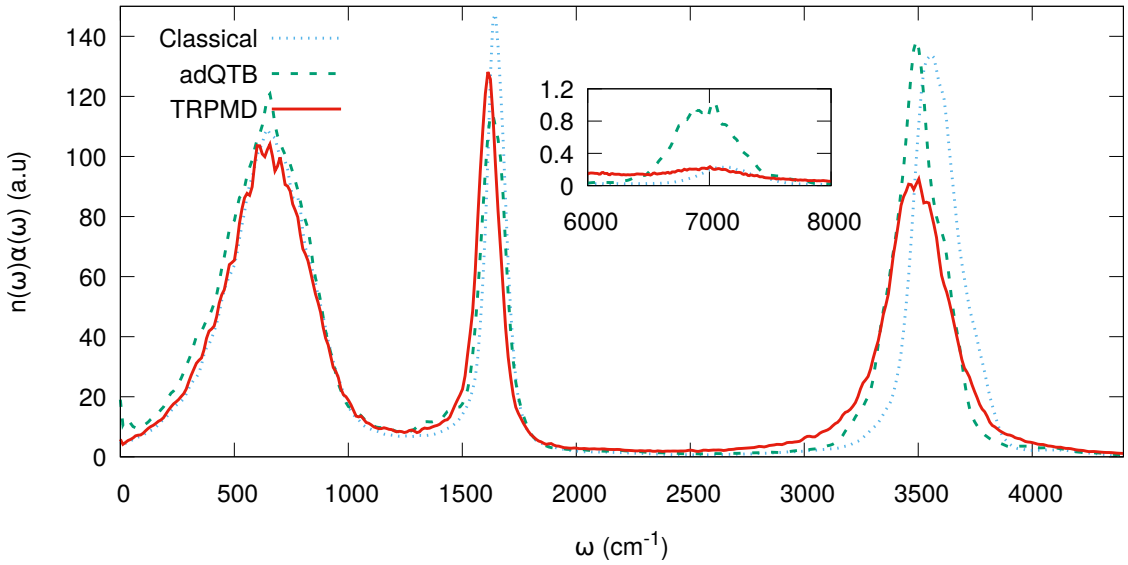


Figure 3.11: Infrared absorption spectra in arbitrary units at 300 K using adQTB and TRPMD compared to classical MD. The quantum spectra are obtained from the Kubo-transformed dipole correlation functions given by equation (2.30).

Indeed, adQTB is able to effectively compensate for the ZPEL in water molecule and yields a IRS which is very close to that of the LSC-IVR (not shown here) approach, even for spectral features corresponding to anharmonic resonances. [188]. The IRS computed from adQTB simulation is compared to those obtained with TRPMD and classical MD in Figure 3.11. Compared to TRPMD, the low-frequency absorption band computed with adQTB is slightly more intense and the bending peak around 1500 cm^{-1} is blue-shifted and broadened. The O-H stretching peak at 3500 cm^{-1} is sharper in adQTB than in TRPMD. Compared to classical MD, stretching peaks of adQTB and TRPMD are equally red-shifted, which is the most prominent effect of the inclusion of NQEs. Moreover, the overtone at 7000 cm^{-1} has a higher intensity in adQTB which is in agreement with LSC-IVR. These differences are mostly in favour of the adQTB approach, since TRPMD has been shown to have spurious broadening and to strongly underestimate anharmonic resonances [187].

3.4.3 Thermodynamic Properties

Density and enthalpy of vaporization

Thanks to the study of the pressure estimation and the Langevin barostat in Section 3.3.5, it is possible to explore fixed-pressure simulations using the Langevin piston barostat defined in equation (3.16). In the adQTB framework, the pressure estimator is challenging because it is defined by the difference of two terms which almost cancel each other. If small inaccuracies remain on either of these two contributions, it can cause non negligible errors. The results obtained in Figure 3.12 were computed as a function of temperature at $P = 1$ atm.

The competition between the stretching and bending ZPE on the hydrogen bonding explains why classical MD and PIMD provide very similar results over the all range of temperatures. Moreover, both curves have the characteristic bell-shape with a maximum around 280 K. This bell-shape is counter intuitive in water which is one of the anomalous properties of this compound [156]. Then, it makes an interesting property on which to study NQEs. The standard QTB fails to capture this temperature dependence: it gives values that decrease monotonically and strongly overestimates the variations over the temperature range. Compensating for the leakage in adQTB allows to recover the overall bell shape and a good agreement with the PIMD reference. Moreover, the adQTB is very accurate within the temperature range interesting for biological systems. However, the curvature of the density is slightly underestimated, leading to errors of the order of 0.005 $\mathrm{g.cm}^{-3}$ in the low and high temperature limits.

The enthalpy of vaporization, $\Delta \mathrm{H}_{vap}$, was also computed in the same NPT simulations. It has been shown that NQEs play an important role on this observable and need to be explicitly included for accurate predictions [203, 204]. It is defined as:

$$\Delta \mathrm{H}_{vap} = U_g - U_l + P(V_g - V_l) \tag{3.17}$$

where U_g and U_l are the internal energies in the gas and liquid phases and V_g and V_l are the corresponding average volumes at pressure P. The classical $\Delta \mathrm{H}_{vap}$ is always overestimated compared to the corresponding PIMD values. By including NQEs with the standard QTB method, $\Delta \mathrm{H}_{vap}$ decreases and becomes underestimated due to ZPEL. Nonetheless, the adQTB allows recovering an almost perfect agreement with the PIMD reference highlighting that the results are unaffected by the inaccuracy in the determination of the density in the low and high temperature ranges.

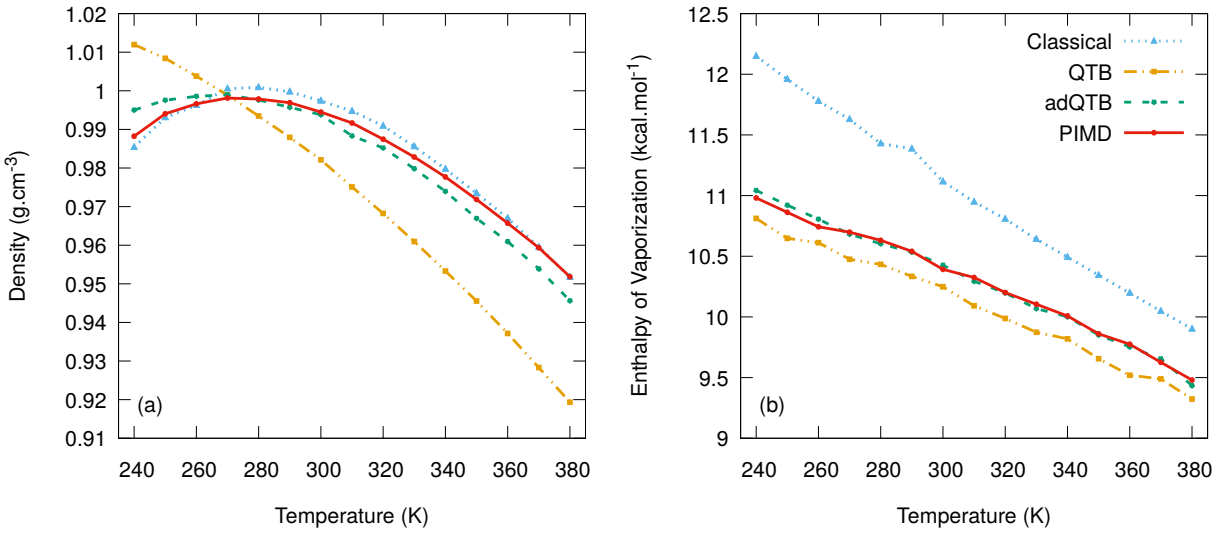


Figure 3.12: Infrared absorption spectra in arbitrary units at 300 K using adQTB and TRPMD compared to classical MD.

Self-diffusion Coefficient

The dynamical properties related to the slow molecular motion of the dynamics cannot be assessed quantitatively in adQTB simulations because of the need for relatively large friction coefficient ($\gamma = 20\,\mathrm{ps}^{-1}$ in (ad)QTB methods). Comparing TRPMD/classical self-diffusion coefficient D to (ad)QTB ones in Table 3.7, a significant drop is observed which shows the higher viscosity induced by the larger friction coefficient γ. On the other hand, the difference between classical MD and TRPMD can be explained by considering that including NQEs increases the energy in the low-frequency modes which influences the most the diffusion coefficient. In adQTB, the diffusion coefficient D ≈ 0.8 cm^2.s^{-1} is underestimated by a factor 2.5 with respect to TRPMD value. This trend is confirmed by the results obtained in classical MD while using $\gamma = 1\,\mathrm{ps}^{-1}$ and $\gamma = 20\,\mathrm{ps}^{-1}$ which give respectively D ≈ 1.9 cm^2.s^{-1} and D ≈ 0.7 cm^2.s^{-1}. Unfortunately, the deconvolution procedure cannot help here since D corresponds to the zero-frequency component of the vibrational spectrum and the deconvolution procedure does not give reliable results in this spectral region [186]. It is possible to reduce $\gamma = 12\,\mathrm{ps}^{-1}$ to increase the accuracy of D giving D ≈ 1.2 cm^2.s^{-1} in adQTB. However, the γ value cannot be reduced further than $\gamma = 10\,\mathrm{ps}^{-1}$ since the friction force will not be able to dissipate the energy excess caused by the ZPEL. The coefficients γ_r will reach 0 which means that the adQTB would not manage to entirely correct the ZPEL. Moreover, a too high value of γ will induce a less efficient sampling, due to the reduced value of D. However, thanks to the low computational cost of the adQTB method, the low self-diffusion constant is not as problematic as it would be for PIMD simulations. There is also the possibility to use a fast-forward Langevin scheme [205] to improve the self-diffusion estimation although it has not been tested and implemented yet.

	Classical	QTB	adQTB	TRPMD
D (10^5. cm^2. s^{-1})	1.89	1.65	0.81	2.25

Table 3.7: Self-diffusion coefficient obtained in classical MD, (ad)QTB and TRPMD simulations.

3.5 Conclusion

Different methods based on path integrals formalism exist to take NQEs into account in MD simulations. However, they are computationally expensive, reducing their range of applicability. It explains the development of alternative methods to include NQEs such as the QTB, which is faster. However, this method suffers from an unphysical flow of energy which is corrected through an alternative method called adQTB. In this chapter we presented some technical improvements that were made to the adQTB method with respect to the original studies [15, 112]. The adaptation efficiency was enhanced by averaging the γ_r over the three directions but also for all same-type atoms. A deconvolution procedure was also added to correct the broadening caused by the friction forces in the potential and kinetic energy terms which can be used on any systems. The results obtained with adQTB on static properties as well as vibrational spectra presented throughout this chapter are in very good agreement with the reference ones highlighting the efficiency of the method to explicitly include NQEs. Therefore, the adQTB aims to renew with the original promise of the QTB method to provide approximate quantum simulations at an almost classical cost.

The additional cost between classical MD and adQTB comes for the additional data needed. These data are the colored noise of the atom and its velocity for each time step in the current simulation segment. They are used for the dynamics and also to compute $\Delta_{FDT}(\omega)$ at the end of each segments. For the simulation box used and the q-TIP4P/F water model, the communication overhead in our parallel implementation inside TINKER-HP represents around 20% of the adQTB simulation time. However, some scalability tests performed on systems containing over 1 million of atoms show that the overhead remains less than 25 %. Moreover, it should be noted that q-TIP4P/F is a really inexpensive model (since it was firstly designed for TRPMD simulations). Then, the communication overheads should become negligible compared to the total simulation time if more complex and realistic interaction potentials were used.

The adQTB, with its cost comparable to classical simulations, is a powerful method to include NQEs for a broad range of applications, in particular to the large-scale simulations required in the field of biophysics and biochemistry. However, the potential used has an inexpensive description of the atomic forces and one may ask how the different results obtained with q-TIP4P/F are affected if an another model is used. Different and more sophisticated models are available such as AMOEBA [31, 206], SIBFA [207] or GEM [208]. Nonetheless, most of them get some of their parameters from quantum mechanics calculations and other from fitting to experimental data. Thus, using PIMD or adQTB methods to include NQEs on these models will result in a double counting of these effects: one time with the FF parameters and a second time within the dynamic. To avoid this and be able to use the adQTB method on more complex systems and realistic potential, a new water model should be developed where the NQEs are explicitly taken into account on the onset. Thereby, the next chapter will be devoted to the development of a new water model, designed to be used in simulations that explicitly account for NQEs (via PIMD or adQTB) and avoid their double counting.

Chapter 4

Development of a Polarizable New Water Model: Q-AMOEBA

Contents

4.1 The Need for a New Water Model

With the extensive development of computing hardware, such as GPU computing [2–5], and techniques such as enhanced sampling [6, 7], MD is now capable to reach timescales of ms making it possible to study macromolecules [4]. The quality of the different thermodynamical observables obtained through a dynamics simulations relies on both the amount of phase-space sampling and on the accuracy of the PES. Most standard additive FFs, such as AMBER [209, 210], CHARMM [211–214] or OPLS [65, 215, 216] have been widely used thanks to their low computational cost and they have been refined throughout the years. Although they provide reliable results for various properties, their accuracy is still limited [217]. The major limitation of these FFs is the use of fixed partial atomic charges to model electrostatic interactions and their lack of a fine description of many-body polarization which inhibits their transferability when the atomic environment changes. Therefore, to circumvent that, much effort has been made to explicitly include many-body polarization in FFs by using different approaches such as Drude oscillators, fluctuating charges or induced dipoles (see Chapter 1 - Section 1.2.6). Then, different versions of CHARMM [218, 219] and AMBER [219] have been published where polarization was added to the existing non polarizable FF. Nonetheless, it has been shown that FFs should include polarization from scratch [82]. Thus, much effort has been made to design new FFs which include polarization from the outset such as PFF [220], AMOEBA [31, 206] or SIBFA [207]. However, these models were parameterized for classical simulations and do not take into account explicitly NQEs hence reducing their transferability. Usually NQEs are implicitly included by fitting the analytical PES to recover thermodynamic observables in classical MD simulations. Therefore, using adQTB or PIMD methods to include NQEs inside a dynamics will end up in a double counting of these effects: one time in the parameters involved in the FF and one more time with the dynamics. Moreover, many recent models use high quality *ab initio* data as a basis for the PES and it has been shown that with such a high accuracy, NQEs must be taken into account explicitly to accurately reproduce thermodynamic observables [29]. Although q-TIP4P/F water model is fitted with TRPMD simulations to reproduce some properties, its inexpensive description of the atomic forces does not make it suitable to study complex systems, even if it is widely used as a benchmark for NQEs simulation method. Therefore, a new polarizable FF which explicitly takes into account NQEs on the onset is needed.

We developed a new polarizable FF, called Q-AMOEBA, that explicitly include NQEs. The methodology to obtain this new FF, with its different results, are presented in this Chapter. Q-AMOEBA uses the AMOEBA PES as a basis which includes many-body polarization effects in its functional form to go beyond the pairwise approximation. Thanks to the promising results obtained in Chapter 3, we have decided to use the adQTB method to reliably approximate NQEs in liquid water with a cost similar to classical MD to develop our new model. The Q-AMOEBA polarizable water model has been parametrized using the ForceBalance (FB) software [184, 221] in combination with Tinker-HP [28] for MD simulations. This new model enables us to finely study the net effects of NQEs by comparing the properties that are obtained with it but with purely classical dynamics. All the different results obtained are compared with path integrals results to ensure that adQTB correctly simulates NQEs in Q-AMOEBA.

4.2 The AMOEBA Water Model

The AMOEBA FF has been used to develop a model which includes explicitly NQEs. The total potential energy of AMOEBA [31] can be expressed as the sum of bonded and non-bonded energy terms:

$$\begin{aligned} E_{total} &= E_{bonded} + E_{non-bonded} \\ E_{bonded} &= E_{bond} + E_{angle} + E_{b\theta} \\ E_{non-bonded} &= E_{vdW} + E_{ele}^{perm} + E_{ele}^{ind} \end{aligned} \tag{4.1}$$

which is analogous to equation (1.45). The functional forms for bond stretching E_{bond} and angle bending E_{angle} were taken from MM3 FF [222]. They include anharmonicity with the use of higher order deviations from ideal bond lengths and angles. An Urey-Bradley term was added in order to better reproduce splitting of the symmetric and asymmetric stretch vibrations. In the original AMOEBA model, the ideal bond length was set to the experimental value of 0.9572 Å [223]. However, the ideal bond angle was set to 108.5°, which is 4° larger than the experimental gas-phase angle of 104.52° to better reproduce the average experimental angle in liquid water.

The non-bonded terms are composed of the van der Waals (vdW) interactions and the electrostatic contributions from both permanent and induced dipoles. The vdW functional term uses Halgren's buffered 14-7 potential to model the pairwise additive interactions for dispersion at long-range and exchange-repulsion at short range [224]:

$$E_{vdW} = \varepsilon_{ij} \left(\frac{1+\delta}{\sigma_{ij}+\delta} \right)^7 \left(\frac{1+\gamma}{\sigma_{ij}^7+\gamma} - 2 \right) \tag{4.2}$$

with ε_{ij} the potential well depth and $\sigma_{ij} = r_{ij}/r_{ij}^0$ where r_{ij} is the i-j separation and r_{ij}^0 is the minimum energy distance (see Figure 1.2). In AMOEBA, the vdW parameters were set to $\gamma = 0.12$ and $\delta = 0.07$. The combining rules used for r_{ij}^0 and ε_{ij} are given by:

$$\begin{aligned} r_{ij}^0 &= \frac{\left(r_{ii}^0\right)^3 + \left(r_{jj}^0\right)^3}{\left(r_{ii}^0\right)^2 + \left(r_{jj}^0\right)^2} \\ \varepsilon_{ij} &= \frac{4\varepsilon_{ii}\varepsilon_{jj}}{\left(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}}\right)^2} \end{aligned} \tag{4.3}$$

A hydrogen reduction factor is added which moves the hydrogen vdW center towards the oxygen along the O–H bond. Permanent atomic monopole, dipole and quadrupole moments are placed on each atomic center. They were computed via distributed multipole analysis (DMA) at the MP2/aug-cc-pVTZ level of theory and then optimized against a high-level ab-initio PES [225]. Polarization effects are treated via mutual induction of dipoles at atomic centers where a classical point dipole moment is induced at each polarizable atomic site according to the electric field felt by that site. This molecular polarization is done via an induction scheme where an induced dipole produced at any site i will induce dipoles at each site until the scheme reaches convergence by using the Thole's damped interaction method. Thole's approach is damped at very short range to avoid the polarization catastrophe [89]. Then, a smearing function is applied on the atomic multipole

moments in each pair of interaction site. Subsequently, the dipole interaction energy has finite value instead of becoming infinite as the separation distance becomes zero. In AMOEBA, the smearing function is defined as:

$$\rho = \frac{3a}{4\pi} e^{-au^3} \tag{4.4}$$

with $u = r_{ij}/(\alpha_i \alpha_j)^{1/6}$ the effective distance as a function of the linear separation r_{ij} and atomic polarizabilities of site i and j. The factor a is dimensionless width parameter of the smeared charge distribution and it controls the damping strength. In AMOEBA, it was chosen to be $a = 0.39$ which leads to a stronger damping and less short-range polarization than the original value of 0.572 given by Thole [206]. However, the atomic polarizabilities for oxygen (0.837 Å^3) and hydrogen (0.496 Å^3) were kept to the same value given by Thole.

4.3 ForceBalance: A Systematic Force Field Optimization

The main challenge in FF development is to choose functional forms that are computationally efficient and flexible enough to capture the relevant physical interactions involved in the different regions explored by the sampling. The fit of the different parameters involved in the FF is also important and usually requires to use fitting data from experimental results and QM calculations. Fitting the different parameters of the FF is often done manually via a trial and error method or it can be performed automatically using different procedure such as least square optimization. ForceBalance (FB), an automatic optimization framework, has already been used to develop new FFs, such as iAMOEBA [184], and also to revise the parameters of existing FFs, such as TIP3P/TIP4P water models. During this thesis, FB was also used to parameterize a revised AMOEBA model with explicit NQEs by using experimental and *ab initio* data (Figure 4.1).

4.3.1 The Optimization Procedure

FB supports many different optimization schemes but the one used during this thesis is the Newton-Raphson with an adaptive trust radius [226]. The overall objective (cost) function is expressed as a sum of weighted mean-square errors over the experimental and theoretical target data sets. It requires the derivatives of the different estimators of the experimental data used in the data set (for example densities, enthalpy of vaporization, etc ...) with respect to the FF parameters. Although it is possible to use finite differences to numerically perform these derivatives, they are very sensitive to statistical noise on average values (computed over a molecular dynamics simulation). Then, very long simulations have to be performed. Furthermore, multiple simulations with slightly different parameters would have to be performed to compute a single derivative. Since the expression of the different observables depends on the parameters through the potential energy E, it is possible to use the explicit form of the partition function to compute the parametric derivatives as an average value over a single MD simulation with the current parameters. Then, one may express the ensemble average of an observable A which depends explicitly on FF parameters as:

$$\langle A \rangle_\lambda = \frac{1}{\mathcal{Z}(\lambda)} \int A(\mathbf{r}, V) e^{-\beta(E(\mathbf{r},V;\lambda)+PV)} \, \mathrm{d}\mathbf{r}\mathrm{d}V \tag{4.5}$$

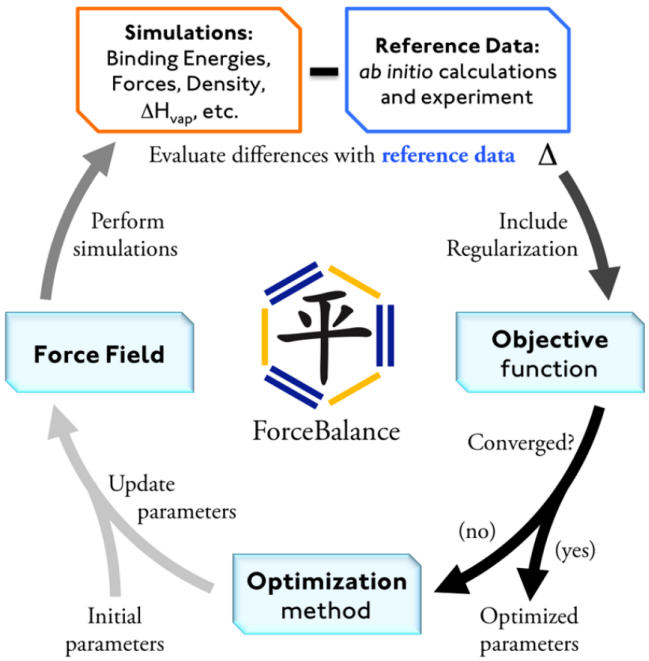


Figure 4.1: ForceBalance procedure. The optimization starts with an initial set of parameters which are used to generate the new FF parameters and perform simulations. The objective function is a weighted sum of squared differences between the simulation results and the reference data, plus a regularization term that prevents overfitting. The optimization method updates the parameters in order to minimize the objective function. Picture extracted from [184, 221].

with $\mathbf{r}$ a molecular configuration, λ the force field parameter, $E(\mathbf{r}, V; \lambda)$ the potential energy and $\mathcal{Z}$ the partition function. Therefore, the derivative can be expressed in classical NPT ensemble as:

$$\frac{\partial \langle A \rangle_\lambda}{\partial \lambda} = \left\langle \frac{\partial A}{\partial \lambda} \right\rangle_\lambda - \beta \left[\left\langle A \frac{\partial E}{\partial \lambda} \right\rangle_\lambda - \langle A \rangle_\lambda \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_\lambda \right] \tag{4.6}$$

where $\langle \cdot \rangle_\lambda$ denotes an expectation value over the distribution $\rho(\mathbf{r}; \lambda) \propto e^{-\beta(E(\mathbf{r},V;\lambda) + PV)}$. This equation can be generalized to path integral simulations using the path integral partition function as:

$$\begin{aligned} \frac{\partial A_{N,\lambda}}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \left[\int \mathrm{d}\mathbf{r}_1 \dots \mathrm{d}\mathbf{r}_N \overline{A}(\lambda) \rho_N(\mathbf{r}_1, \dots, \mathbf{r}_N; \lambda) \right] \\ &= \left\langle \frac{\partial \overline{A}}{\partial \lambda} \right\rangle_{N,\lambda} - \beta \left\langle \overline{A}(\lambda) \left(\frac{1}{N} \sum_{i=1}^{N} \frac{\partial E(\mathbf{r}_i; V; \lambda)}{\partial \lambda} \right) \right\rangle_{N,\lambda} - \langle \overline{A}(\lambda) \rangle_{N,\lambda} \frac{1}{\mathcal{Z}_N(\lambda)} \frac{\partial \mathcal{Z}_N}{\partial \lambda} \\ &= \left\langle \frac{\partial \overline{A}}{\partial \lambda} \right\rangle_{N,\lambda} - \beta \left[\left\langle \overline{A} \frac{\partial \overline{E}}{\partial \lambda} \right\rangle_{N,\lambda} - \langle \overline{A} \rangle_{N,\lambda} \left\langle \frac{\partial \overline{E}}{\partial \lambda} \right\rangle_{N,\lambda} \right] \end{aligned} \tag{4.7}$$

with the average value over the beads for any function O is denoted as $\overline{O} = \sum_{i=1}^{N} O(\mathbf{r}_i, V; \lambda)/N$ and

$$\rho_N(\mathbf{r}_1, \dots, \mathbf{r}_N; \lambda) = \frac{e^{-\beta \left[\sum_{i=1}^{N} \frac{E(\mathbf{r}_i; V; \lambda)}{N} + K(\mathbf{r}_1, \dots, \mathbf{r}_N) + PV \right]}}{\mathcal{Z}_N(\lambda)} \tag{4.8}$$

$K(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the path integral harmonic energy that couples the different beads and which is independent of λ and $\mathcal{Z}_N(\lambda)$ is the partition function which normalizes ρ_N.

To start the FB optimization procedure, we have decided to use the initial parameters of the AMOEBA water model (denoted as AMOEBA03). Because most of its parameters (such as atomic permanent multipolar moments) are obtained directly from *ab initio* calculations, only the vdW parameters were modified. More specifically, the optimization was performed on both vdW radii and epsilon values of the Halgren 14-7 potential associated to oxygen and hydrogen atoms, as well as the buffer radius of the hydrogen atom (equation (4.3)).

4.3.2 Results using Quantum Chemistry Data Only

One may want to obtain a new reliable FF using only *ab initio* data and be able to reproduce thermodynamical observables. To do so, the *ab initio* references include properties for systems ranging from the monomer to clusters of 22 water molecules. The charges, dipole, quadrupoles, polarizability, vibrations and optimized geometries were used for the monomer. The *ab initio* energies and optimized geometries for the ground state dimer, Smith dimer set (10 total) [227], trimer, tetramer, pentamer, eight hexamers [228], two octamers [229], five 11-mers [230], five 16-mers, two 17-mers and four 20-mers [231] were also added in our target *ab inito* QM data set. Using the previous work by Wang et al. [184], 42 000 cluster geometries obtained with AMOEBA for temperatures ranging from 249 to 373 K were used where energies and gradients were computed for the different clusters via RI-MP2/heavy-aug-cc-pVTZ [232, 233] with the Q-Chem 4.0 software [234]. These data were already available in the FB software in the 'studies/015_amoeba_tinker' directory. Then, they were used with FB to obtain a new set of vdW parameters fitted only on *ab initio* data.

Firstly we computed the binding energy (BE) of our new set of parameters to ensure that the FB procedure worked. In *ab initio* calculations, the BE is defined as:

$$\Delta E_{binding} = E_{cluster} - nE_{monomer} \tag{4.9}$$

with $E_{cluster}$ the total energy of the studied cluster, n the number of water molecules and $E_{monomer}$ the energy of an isolated optimized water molecule. Then, $\Delta E_{binding}$ includes the intramolecular deformation interaction energy of the cluster. Following this, it is possible to define the cluster binding energy computed with our new set of parameters obtained through a FB procedure as:

$$\Delta E^{FF}_{binding} = (E_{intra} + E_{inter})_{cluster} - n(E_{intra})_{monomer} \tag{4.10}$$

Moreover, the energy of an optimized water monomer $n(E_{intra})_{monomer}$ is equal to 0 kcal.mol^{-1}. Then, it is possible to re-write equation (4.10) as:

$$\Delta E^{FF}_{binding} = (E_{intra} + E_{inter})_{cluster} \tag{4.11}$$

Therefore, the binding energies of different clusters ranging from dimer to 20-mers were computed and compared with *ab initio* references as well as AMOEBA03. To do so, the Root Mean Square Error (RMSE) were computed. It is equal to 2.21 kcal.mol^{-1} for our new set of parameters whereas AMOEBA03 gives 3.33 kcal.mol^{-1}. Then, this new set of parameters was used with PIMD and adQTB simulations with

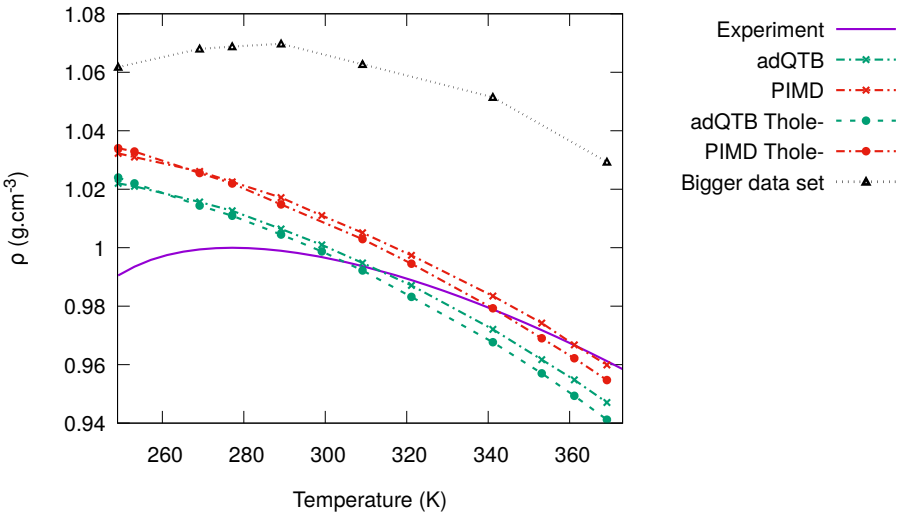


Figure 4.2: Density obtained with the FB optimization using only *ab initio* chemistry data. The red and green curve correspond to the PIMD and adQTB methods respectively. Because the first FB procedure which changed only vdW parameters does not give a reliable shape, the Thole parameter was reduced to decrease the RMSE. The corresponding density with both methods are plotted with a dot and labelled 'Thole-'. The black curve corresponds to the density obtained while adding 26-mer [235] and 32-mer [236] at MP2 level using aug-cc-pvDZ basis in the data set.

Tinker-HP to see if *ab initio* polarizable force fields and quantum simulations are able to reproduce different observables such as the density. The adQTB and PIMD methods were used to compute the density in Figure 4.2 (cross curves). Both methods fail to reproduce the bell-shape of the density. Moreover, adQTB tends to systematically slightly underestimate the density by roughly 1% with respect to the PIMD reference, but this discrepancy is small compared to the difference between the experimental data. Consequently, the main problem in our FF is the parameters obtained after the FB procedure and not the method used to include NQEs.

So far, these results were obtained by modifying only the vdW parameters with respect to the initial set of parameters of AMOEBA03. However, it is also possible to change the Thole parameter and see if the results previously obtained improve. To do so, we have decided to change the width parameter a. The corresponding RMSE obtained is equal to 1.99 kcal.mol^{-1} and its density with this new width parameter using adQTB and PIMD methods is shown in Figure 4.2 (because the Thole parameter a was reduced, its corresponding density is referred as Thole-). One may notice that no matter which method is used to compute ρ and how the RMSE is changed, the over-all shape does not correspond to the one given by the experiment (purple curve). Moreover, changing the parameter a has the only effect to shift the curves.

To reduce the number of parameters in the FB procedure, we have decided to keep the Thole parameter a to its initial value. One may think that having a data set which includes more conformations will help to recover the density since it gets closer to the bulk phase. Therefore, more *ab initio* data were also added to study how the new vdW parameters obtained with this new target data set affect the density. This new data set includes 26-mer [235] and 32-mer [236] at MP2 level using aug-cc-pvDZ basis set. Although the RMSE is reduced (2.06 kcal.mol^{-1} while the previous parameter set gave 2.21 kcal.mol^{-1}), its corresponding density is strongly overestimated compared to the one previously obtained (black curve). Moreover, it is really close to the density obtained using *ab initio*-based polarizable TTM2.1-F force field [237]. Therefore, we have also

System	Reference data	Data type	Data point	Weight	
Clusters	Gas Phase Dipole-Quadrupole	CCSD(T)		1.0	
	Gas Phase Vibrational modes	CCSD(T)		1.0	
	Smith Dimer	CCSD(T) BE	10	1.0	
	Small Gas Phase Cluster	CCSD(T) BE	21	1.0	
	Large Gas Phase Clusters	MP2 BE	18	1.0	
	PE and AF	MP2	42.000	1.0	
Liquid	ρ	Expt.	10	1.0	0.6
	ΔH_{vap}	Expt.	10		0.4

Table 4.1: Reference data used in FB to derive the new sets of parameters. BE refers to Binding Energies, PE to Potential Energies and AF to Atomistic Forces. The table also shows the weights that were applied to the different target properties. The Clusters were only used in the objective function on the results discussed in Section 4.3.2.

decided to only use the initial *ab initio* set of data already available in the FB software.

4.3.3 Results using thermodynamic observables

Although the RMSE is decreased while using only quantum chemistry data in FB, the density computed with both methods are overestimated and the bell-shape is not recovered. Changing the Thole parameter as well as the vdW ones did not give the expected results. Therefore, thermodynamic observables have to be added to our objective function in FB to ensure that the new FF parameters obtained are capable to reproduce different properties. The density ρ and enthalpy of vaporization ΔH_{vap} were added to the cost function (Table 4.1) and then evaluated at different temperatures ranging from 249.15 K to 369.15 K under 1 atm using with the TINKER-HP multi-GPUs software. All simulations were carried out for N=4000 molecules in periodic boundary conditions with a cubic box. For each FB iterations, the MD simulation includes 1 ns of equilibration and 4 ns of production.

Thanks to its low computational cost, the FB procedure done with the adQTB method was the first to converge. The density obtained with this new set of parameters is shown in Figure 4.3. Including the density in the objective function allows to recover the correct bell-shape. However, it is overall underestimated compared to the experimental data. Therefore, we have decided to also modify the average O-H-O bond angle and bond length. These new values get closer to theoretical value given in gas phase which makes our fit procedure encouraging. By doing this, the density is improved in the low and high temperature ranges as it can be seen in Table 4.2 and in Figure 4.3 (referred as adQTB finely tuned). The need to also modify these parameters can be explained by the high frequencies of the intramolecular stretching and bending modes sampled by the quantum dynamics, which are associated with high zero point energies, and then impact the average value of the corresponding bond angle and bond length. Then, we have decided to use this set of parameters with the PIMD method to ensure that the density is correctly sampled no matter which method is used to include NQEs. However, at 298.15K, the corresponding density is $\rho = 1.006\,\mathrm{g.cm^{-3}}$ whereas the same parameters in adQTB gives $\rho = 0.999\,\mathrm{g.cm^{-3}}$. These differences can be explained by how the pressure estimator is calculated between adQTB and PIMD. As explained in Chapter 3, the pressure estimator is composed of

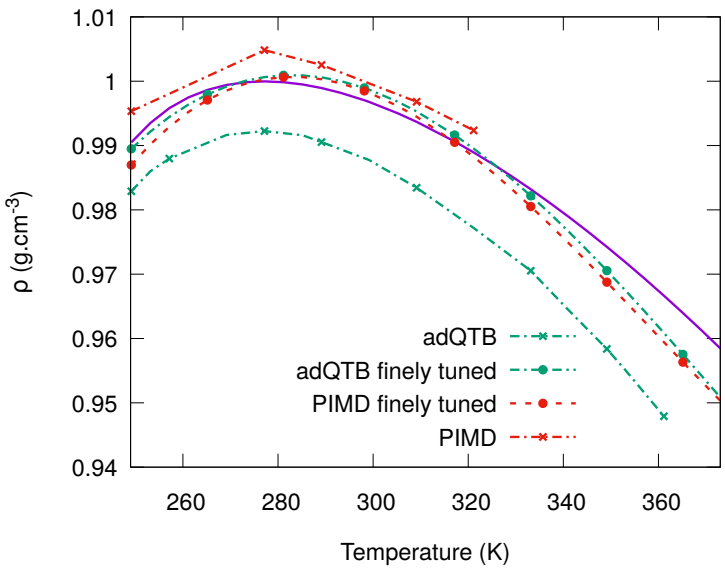


Figure 4.3: Density obtained with the FB procedure using thermodynamical observables in the cost function. When only the vdW parameters were changed, the corresponding curve is labelled with a cross. The PIMD curve uses the vdW adQTB parameters. Because at 298.15K, its density is overestimated, no calculations at temperatures higher than 321.15K were done. The finely tuned referred to the intramolecular terms changed which is labeled with a dot. Because PIMD and adQTB finely tuned give the best results, they will be denoted as Q-AMOEBA in the rest of the Chapter. The purple curve is the experimental data.

two terms which almost cancel each other (equation (3.14)). In the PIMD framework, the kinetic energy (from which depends the pressure estimator), is given by the centroid virial estimator whereas in adQTB it is proportional to a correction factor obtained through the deconvolution procedure. Although the differences between both terms in the adQTB method are really small, some differences remain with respect to the PIMD calculations which result in some differences for the pressure estimation. These discrepancies cause an error on the density estimation in adQTB. Then, by using the adQTB with FB, the corresponding objective function will target the experimental density while having this error on the pressure estimation. Because FB uses a series of thermodynamical fluctuation equations to obtain the converged parameters, the error on the pressure in adQTB will have an impact on how the parameters are optimized. Therefore, two different set of

	249.15 K	298.15 K	361.15 K
adQTB	0.983	0.988	0.945
adQTB finely tuned	0.989	0.999	0.961
PIMD	0.995	0.999	N/A
PIMD finely tuned	0.987	0.998	0.959
experimental data	0.991	0.997	0.967

Table 4.2: Density in g.cm^{-3} obtained with the new set of parameters while doing FB optimizations only on the vdW parameters compared to experimental data. The 'finely tuned' refers to the set of parameters where intramolecular terms were also modified to reproduce the correct density over the all range of temperatures with adQTB and PIMD methods.

term	parameter	unit	initial	Q-AMOEBA (adQTB)	Q-AMOEBA (PIMD)
vdW	O vdW diameter	Å	3.4050	3.405099	3.405099
	O vdW epsilon	kcal.mol^{-1}	0.1100	0.120139	0.120190
	H vdW diameter	Å	2.6550	2.643027	2.642967
	H vdW epsilon	kcal.mol^{-1}	0.0135	0.009943	0.010287
	H vdW reduction	none	0.910	0.93429	0.937788
bonded	O-H bond length	Å	0.9572	0.9472	0.9572
	H-O-H angle	Å	108.50	106.20	107.20

Table 4.3: Value of the initial/AMOEBA03 [31] water parameters compared to the ones obtained with FB [184] with the adQTB and PIMD methods.

parameters have to be developed.

To obtain a reliable FF with FB while using PIMD, we have decided to start our optimization procedure by using the set of vdW parameters previously derived with the adQTB method. However, the initial value of AMOEBA03 for the bond length and bond angle were used. Therefore, a new set of parameters were obtained and its corresponding density is shown in Figure 4.3 (denoted as PIMD). Contrary to what was observed with the adQTB, the PIMD density is overestimated. Because of that, no calculation at temperatures higher than 321.15 K were done. The bond length is already set to its experimental value (0.9572 Å), then only the H-O-H bond angle was reduced to recover the correct density over the all range of temperatures (see Table 4.2 and Figure 4.3). The need to change the average bond and angle length in adQTB can be explained by the discrepancies observed in the pressure estimator. It should also be noted that increasing the weight of ΔH_{vap} in the cost function does not change the optimized vdW parameters in both methods. By using the density and the enthalpy of vaporization in FB procedure, our new sets of parameters are now able to reproduce some thermodynamical properties. Therefore, they will be denoted as Q-AMOEBA (PIMD/adQTB) in the rest of the chapter. Both set of parameters and the initial AMOEBA03 values given by are shown in Table 4.3.

4.4 A new quantum polarizable water model: Q-AMOEBA

With the results obtained on the density, we have decided to use these two set of parameters (PIMD and adQTB) and see if Q-AMOEBA is able to reproduce both gas phase and condensed phase properties. As explained earlier, both set of data were used in FB. However, the parameters for the intramolecular energy terms were not optimized automatically but adjusted manually. Therefore, to assess the accuracy of our fitting procedure, the RMSE of the different conformations used as a reference data were computed. Because the functional form and the initial set of parameters were taken from AMOEBA03, all the results will be compared to this model.

$(H_2O)_2$	CCSD(T)	AMOEBA03	Q-AMOEBA (PIMD)	Q-AMOEBA (adQTB)
Smith01	-4.97	-4.58	-4,96	-4,95
Smith02	-4.45	-3.98	-4.35	-4,34
Smith03	-4.42	-3.94	-4.3	-4.3
Smith04	-4.25	-3.54	-3.49	-3.43
Smith05	-4.00	-2.69	-3.06	-3.00
Smith06	-3.96	-2.59	-2,95	-2.90
Smith07	-3.26	-2.55	-2,81	-2.73
Smith08	-1.30	-0.8	-1.04	-0.95
Smith09	-3.05	-2.69	-2,97	-2.90
Smith10	-2.18	-1.89	-2.14	-2.07
RMSE		0.81	0.53	0.57

Table 4.4: Binding Energies in kcal.mol^{-1} of the 10 Smith's water dimers using Q-AMOEBA compared to the initial set of parameters AMOEBA03 and *ab initio* references. The table distinguishes Q-AMOEBA (PIMD) and Q-AMOEBA (adQTB) depending on which method was used to include NQEs in the fit procedure.

4.4.1 Binding energies of water clusters

Smith's water dimers

The BE of the Smith's water dimer are presented in Table 4.4. They are representative of the most favorable (oxygen-hydrogen) and unfavorable (oxygen-oxygen and hydrogen-hydrogen) interactions that could occur between two water molecules during a MD simulation. Therefore, they have been widely used as a benchmark for modern potentials but also for *ab initio* methods. Both sets of parameters yield the same most stable Smith dimer. Moreover, both set of parameters give a smaller RMSE (0.57 kcal.mol^{-1} and 0.53 kcal.mol^{-1} for adQTB and PIMD respectively) compared to the initial set of parameters (AMOEBA03 yields 0.81 kcal.mol^{-1}).

Water Clusters

To go towards the bulk condensed-phase structure, BE of larger water clusters were also computed in Table 4.5. In the intermediate size of water clusters ($3 \leq n \leq 8$), Q-AMOEBA slightly overestimates the BE while remaining within a 0.5 kcal.mol^{-1} per monomer range. For larger cluster, the accuracy of Q-AMOEBA is improved, giving an overall smaller RMSE than AMOEBA03. It is approximately improved by 0.9 and 0.2 kcal.mol^{-1} for the PIMD and adQTB sets of parameters respectively. Moreover, it should be noted that none of the two sets of parameters give a better RMSE compared to the one obtained while fitting only on QM calculations in Section 4.3.2 (2.21 kcal.mol^{-1}). It highlights that our *ab initio* references used were not accurate enough to reproduce thermodynamical properties. Moreover, adding some properties in our data set increase the RMSE. Therefore, it highlights that our CCSD(T) reference is not accurate enough to fully reflect the experimental complexity of thermodynamical observables. Then, using explicit inclusion of higher level of Coupled Cluster excitations could lead to better predictions of condensed-phase properties while reducing the RMSE [238].

$(H_2O)_n$	geometry	QM	AMOEBA03	Q-AMOEBA (PIMD)	Q-AMOEBA (adQTB)
n=3	cyclic	-15.74	-15.03	-16.05	-16.11
n=4	cyclic	-27.40	-27.63	-28.98	-29.32
n=5	cyclic	-35.93	-36.38	-38.19	-38.72
n=6	prism	-45.92	-45.71	-47.88	-48.36
	cage	-45.67	-45.82	-47.95	-48.39
	bag	-44.30	-44.79	-46.97	-47.53
	cyclic chair	-44.12	-44.62	-46.88	-47.60
	book1	-45.20	-45.6	-47.8	-48.39
	book2	-44.90	-45.37	-47.55	-48.11
	cyclic boat1	-43.13	-43.78	-45.79	-46.65
	cyclic boat2	-43.07	-43.84	-46.04	-46.73
n=8	S4	-72.70	-72.34	-75.71	-76.46
	D2d	-72.70	-72.39	-75.75	-76.50
n=11	434	-105.72	-101.74	-106.40	-107.52
	515	-105.18	-102.08	-106.85	-108.11
	551	-104.92	-101.85	-106.61	-107.86
	443	-104.76	-101.86	-106.63	-107.78
	4412	-103.97	-101.41	-106.26	-107.50
n=16	boat-a	-170.80	-161.21	-168.83	-170.57
	boat-b	-170.63	-161.54	-169.12	-170.92
	antiboat	-170.54	-162.26	-169.98	-171.95
	ABAB	-171.05	-161.18	-168.69	-170.32
	AABB	-170.51	-160.98	-168.51	-170.13
n=17	sphere	-182.54	-173.02	-181.21	-183.32
	5525	-181.83	-172.32	-180.42	-182.46
n=20	dodecahedron	-200.10	-197.03	-206.1	-208.54
	fused cubes	-212.10	-205.76	-215.35	-217.42
	face sharing prisms	-215.20	-206.13	-215.74	-218.05
	edge sharing prisms	-218.10	-208.46	-218.40	-220.85
RMSE			3.33	2.36	3.12

Table 4.5: Binding energies in kcal.mol^{-1} from trimer to 20-mer water clusters using Q-AMOEBA compared to the initial set of parameters AMOEBA03 and *ab initio* references. The table distinguishes Q-AMOEBA (PIMD) and Q-AMOEBA (adQTB) depending on which method was used to include NQEs in the fit procedure.

With these encouraging results obtained, other thermodynamical properties have been computed from temperatures ranging from 249.15 K to 369.15 K under 1 atm. To reduce the computation time and be able to do dynamics on big boxes (and get closer to the bulk-phase representation), both methods were implemented in the TINKER-HP multi GPUs software. Then, a box of 4000 molecules with a RESPA integrator was used with timesteps of 0.2 and 2 fs for the bonded/non-bonded part. PIMD simulations used 32 beads with a mild Langevin thermostat on the centroid where the friction coefficient was set to $\gamma = 1\,\mathrm{ps}^{-1}$. In the adQTB method, the friction was set to $20\,\mathrm{ps}^{-1}$.

4.4.2 Structural Properties of Water

Static Equilibrium Properties

In Figure 4.4, the different RDFs obtained with Q-AMOEBA (PIMD and adQTB) are shown. The experimental RDF given by neutron diffraction (1986) [239] and X-ray scattering (2013) [240] were not included in our parametrization. Then, this property will be used as a first test for our Q-AMOEBA model. Both Q-AMOEBA (PIMD and adQTB) RDFs are almost indistinguishable and in close agreement with experiments. Moreover, the differences between our results and the experiments are of the same order as the discrepancies within experimental data. The first peak of the O-O RDFs is located at 2.73 Å and 2.80 Å in X-ray and neutron diffraction respectively. This feature is well reproduced by Q-AMOEBA where this peak is located at 2.80 Å (PIMD) and 2.79 Å (adQTB).

Properties of the ensemble-averaged water monomer are also given in Table 4.6 where the O-H bond length and H-O-H angle are shown with Q-AMOEBA set of parameters. The average value of the H-O-H angle in the liquid phase compared to its value in gas phase is known to be problematic with classical FF. Then, it requires to artificially increase the equilibrium parameter [31] as it can been seen in Table 4.3. In

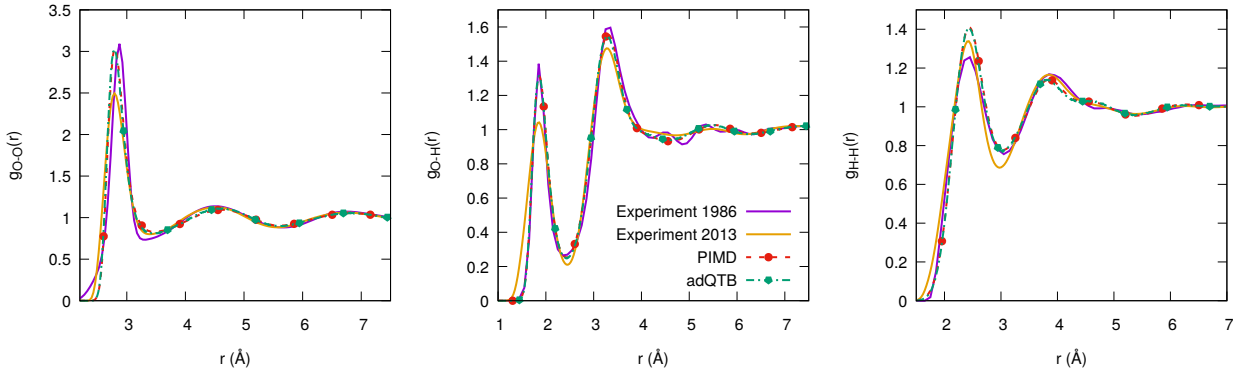


Figure 4.4: Radial distribution function at 298.15 K and constant volume obtained with Q-AMOEBA compared to X-ray (2013) [240] and neutron diffraction (1986) [239] experiments.

	PIMD (32 beads)	PIMD (64 beads)	adQTB	Expt.
r_{OH}(Å)	0.985	0.985	0.975	0.97
θ_{HOH}(deg)	106.08	106.10	104.95	105.1

Table 4.6: Average O-H bond length and H-O-H angle while using Q-AMOEBA water model compared to experimental values in liquid phase.

	Bond stretching	Angle bending	Urey-Bradley	Van der Waals	Atomic Multipoles	Polarization	Non-Bonded
Classical	0.75	0.43	-0.03	5.48	-11.43	-5.19	-11.14
adQTB	5.55	1.22	-0.11	4.37	-9.95	-4.43	-10.01
PIMD (32 beads)	5.43	1.19	-0.12	4.32	-9.87	-4.35	-9.9
PIMD (64 beads)	5.57	1.20	-0.12	4.33	-9.88	-4.38	-9.93

Table 4.7: Average energy terms per water molecules in kcal.mol^{-1} obtained with the different methods at 298.15 K obtained from 1 ns simulations at constant volume corresponding to the density $\rho = 0.997\,\mathrm{g.cm^{-3}}$. The classical row corresponds to the adQTB set of parameters.

AMOEBA03 it is equal to 108.2°. However, both set of Q-AMOEBA parameters have a reduced ideal angle. It is equal to 106.2° and 107.2° with adQTB and PIMD respectively. This is because NQEs tend to slightly increase the average H-O-H angle of water molecules in the liquid phase. Hence, it has to be reduced in the parameters set. Furthermore, NQEs also slightly increase the average O-H distance. Then, this effect was compensated in both set of parameters. However, it was decreased further in adQTB. It went from 0.9572 to 0.9472 Å. It highlights the net effects of NQEs and the need to explicitly include them to design accurate models of water, although adQTB water model tends to slightly overestimate their effects.

Average Energy Terms

The impact of NQEs on the different energy terms of water at ambient conditions were further analyzed and are reported in Table 4.7. The Q-AMOEBA model also allows to finely study the practical effects of NQEs by comparing the properties that are obtained with it but with purely classical dynamics. Therefore, all results labelled 'Classical' are obtained with classical MD simulations using Q-AMOEBA (adQTB) set of parameters. All intramolecular energies obtained with classical MD are underestimated compared to their quantum counterparts (adQTB and PIMD). However, the difference in the non-bonded part between classical and quantum simulations are less significant (although still present) which highlights some error compensation in the functional form of AMOEBA03.

4.4.3 Infrared Absorption Spectra

In AMOEBA, the total dipole moment is estimated as:

$$\mu = \sum_{i=1}^{N_{at}} (q_i \mathbf{r}_i + \mu_i^0 + \mu_i^{ind}) \tag{4.12}$$

with q_i the permanent charges located on the atom's position $\mathbf{r}_i$, μ_i^0 the permanent dipoles and μ_i^{ind} the induced dipoles that are obtained at each step of the dynamics via a minimization procedure [28]. Because no analytical form for the time derivative of the induced dipole is available, finite differences were used to estimate the total dipole moment over the trajectory. As seen in Chapter 3, the IRS is directly evaluated in TRPMD dynamics, however it is multiplied by $\frac{\tanh(\beta\hbar\omega/2)}{\beta\hbar\omega} = \frac{k_B T}{\theta(\omega,T)}$ in the adQTB framework to obtain the Kubo-transformed spectrum. Moreover, in adQTB simulations, the high friction coefficient needed to compensate ZPEL tends to broaden the IRS. Then, the deconvolution procedure is needed to recover the unaffected spectrum. The different IRS are shown in Figure 4.5 using TRPMD and adQTB compared to the experimental spectrum, classical AMOEBA03 and the classical spectrum obtained from a simulation using the adQTB set of parameters. The spectra obtained from both quantum simulations (TRPMD and adQTB)

are very similar in their peak positions and are in good agreement with the experimental reference. Although relative intensities are similar between quantum and classical simulations, TRPMD spectrum is broadened for the stretching peak which is a well-known discrepancy of this method. Moreover, including NQEs shifts the intramolecular peaks, composed by the bending around 1600 cm^{-1} and stretching around 3500 cm^{-1}, towards the lower frequencies. The low-frequency area (0 to 1000 cm^{-1}) is very similar for AMOEBA03, TRPMD and adQTB which are red-shifted compared to the results obtained in classical simulation with Q-AMOEBA (light blue). Then, it highlights the impact of NQEs even at those low frequencies. Moreover, the general shape of this low-frequency area is in good agreement with the experiment and even display some sub-structure at around 200 cm^{-1} due to the slow induced-dipole dynamics [30, 241]. This feature is absent for non-polarizable models, even when including NQEs [30]. Moreover, the IRS were also computed using only the linear dipole while including NQEs and this sub-structure were not on the spectrum. It should also be noted that q-TIP4P/F water model only includes permanent charges and its corresponding spectrum presented in Chapter 3 (Figure 3.11) does not have this feature showing that NQEs with permanent and induced dipoles should be used to correctly recover IRS.

4.4.4 Thermodynamic Properties

Properties Used in the Objective Function: Density and Enthalpy

In Section 4.3.3 (Figure 4.2) results obtained using only Q-AMOEBA (adQTB and PIMD) set of parameters in the density were shown. However, the enthalpy of vaporization ΔH_{vap} was also added to our objective function. Then, to extend the study previously done, we also compute this property with both sets of parameters and compared them to classical MD and AMOEBA03. Results are shown in Figure 4.6 for both properties.

Compared to the experimental curve (purple), the deviations of the density at ambient temperature for

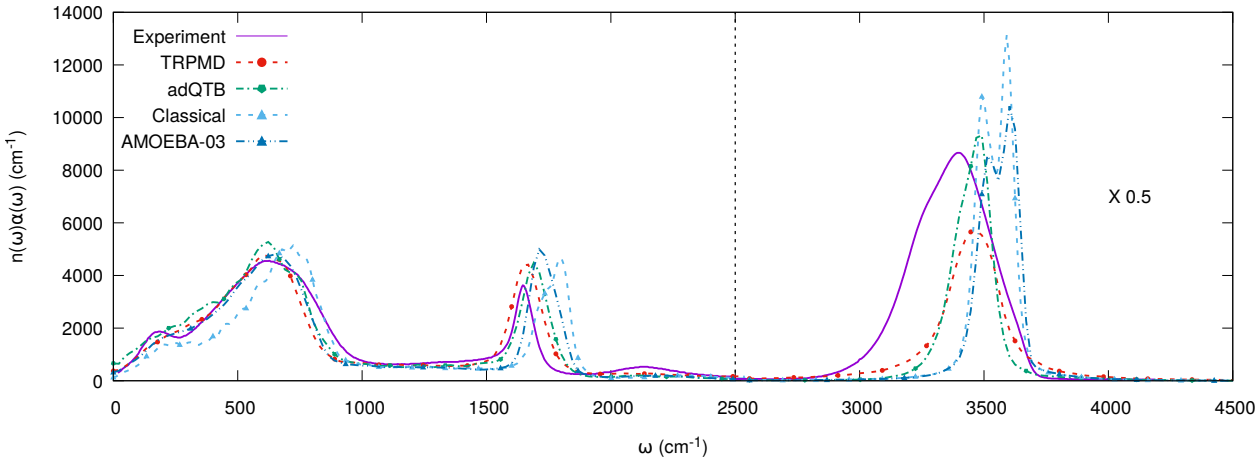


Figure 4.5: IRS computed at 300 K and $\rho = 0.997\,g.cm^{-3}$ with Q-AMOEBA water model using adQTB and PIMD methods with their respective set of parameters. The classical curve is obtained from classical MD simulation with the Q-AMOEBA (adQTB) model. They are compared with AMOEBA03 (classical MD) and experimental data. The right part of the plot is multiplied by 0.5 to put forward the stretching mode. The quantum spectra are obtained from the Kubo-transformed dipole correlation functions given by equation (2.30).

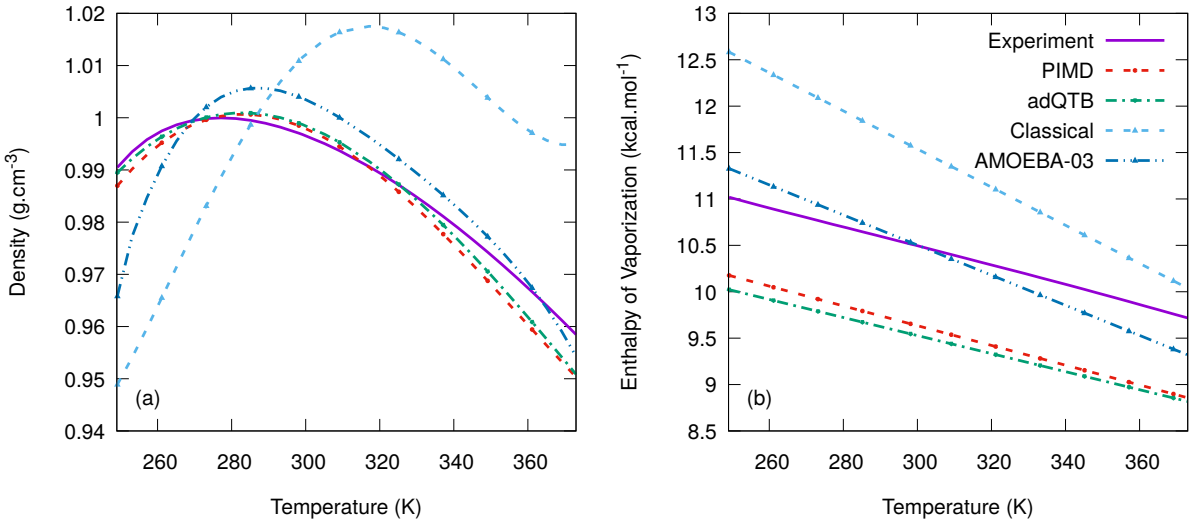


Figure 4.6: Density ρ (left) and enthalpy of vaporization ΔH_{vap} (right) of liquid water at $P = 1$ atm as a function of the temperature for the different AMOEBA model. The dashed line represents the results obtained with the Q-AMOEBA (PIMD) method whereas the dot-dashed lines are obtained with Q-AMOEBA (adQTB) model.

the PIMD and adQTB methods are less than 0.2 % difference (see Table 4.2). In both low and high ranges of temperatures, the density is slightly underestimated by a maximum of 0.85 % at 369.15 K. One may notice that classical MD simulation using Q-AMOEBA (adQTB) set of parameters yields a significantly different curve from both quantum (adQTB and PIMD) and AMOEBA03 density. This result highlights that the different changes made in the FF have a non-negligible impact on the density. Moreover, the temperature of maximum density, T_{MD}, in the classical MD curve is shifted to higher values while the density is reduced at low temperature and increased at high ones with respect to the results obtained including NQEs. To explain the impact of NQEs, the peculiarities of liquid water are used to explain the unusual bell-shape of the density. A loosely packed local structure is produced due to the tetrahedral arrangement of the water molecules. Increasing the temperature for $T < T_{MD}$ allows hydrogen bond to strain and break more easily which enables larger deviations from the tetrahedral order with a better packing efficiency and hence increases the density. For $T > T_{MD}$, the H-bonds breaking tends to increase nearest-neighbors distances which leads to the more usual thermal expansion observed at high temperatures [17] and decreases the density. Therefore, NQEs lead to an overall weakening of the hydrogen bonds over the whole temperature range. The study of the different non-bonded energy contributions of the Q-AMOEBA model given in Table 4.7 confirmed this reasoning. The values given by the PIMD parameter set are superior by approximately 1.0 kcal.mol $^{-1}$ per water molecule at 298.15 K compared to the classical counterpart. It should also be noted that the impact of NQEs on the hydrogen bond strength depends on the model. In the q-TIP4P/F water model, densities computed with PIMD or classical MD are very similar (see Figure 3.12). However, in TTM2.1-F, both curves are overestimated and the PIMD density is lower than its classical counterpart over the all temperature range.

The enthalpy of vaporization, ΔH_{vap}, was also computed in the same NPT simulations. As shown in the right plot of Figure 4.6, classical enthalpy of vaporization is overestimated and decreases with the inclusion of NQEs. The Q-AMOEBA classical simulation yields 11.5 kcal.mol^{-1} at 298.15 K which is ~ 0.5 kcal.mol^{-1} higher than the reference. Although Q-AMOEBA (PIMD and adQTB) results underestimate the ΔH_{vap} by approximately ~ 0.8 kcal.mol^{-1} (which is close to other sophisticated models [242]), the slope is correctly

recovered over the whole range of temperatures. Moreover, the accuracy of this observable is strongly related to the PES. Then, using a higher level of *ab initio* data as a reference could better reflect the complexity of this observable.

Properties Outside the Objective Function

The thermodynamical properties presented so far were included in our data set. However, one may be interested to see how Q-AMOEBA performs on properties outside the data used to fit it. Therefore, the validity of the Q-AMOEBA water model was further studied by computing other equilibrium properties which were not included in the objective function of FB.

The isobaric heat capacity (left) and thermal expansion coefficient (right) are shown in Figure 4.7 with their classical counterpart to study the net effects of NQEs on this thermodynamic observable. The isobaric heat capacity is defined as:

$$c_p = \left(\frac{\partial \mathrm{H}}{\partial T}\right)_P \tag{4.13}$$

with H and T the enthalpy and the temperature at a given pressure P. A fourth-order polynomial interpolation procedure of the temperature derivative of H at different temperatures was used to compute c_p in Figure 4.7 (left panel). As the enthalpy of vaporization $\Delta \mathrm{H}_{vap}$, the heat capacity c_p is strongly affected by NQEs. The intramolecular modes have a frequency $\hbar\omega \gg k_B T$ which is mainly ZPE (which is approximately 1000 K for the O-H stretching in water). Then, it is independent of the temperature and hence does not contribute to the final isobaric heat capacity c_p. However, in the classical case, intramolecular modes have an average energy $k_B T$ and so contribute to the final c_p. It leads to a significant overestimation in classical MD simulations as it can be seen in Figure 4.7 with the classical and AMOEBA03 curves. On the other hand, including NQEs

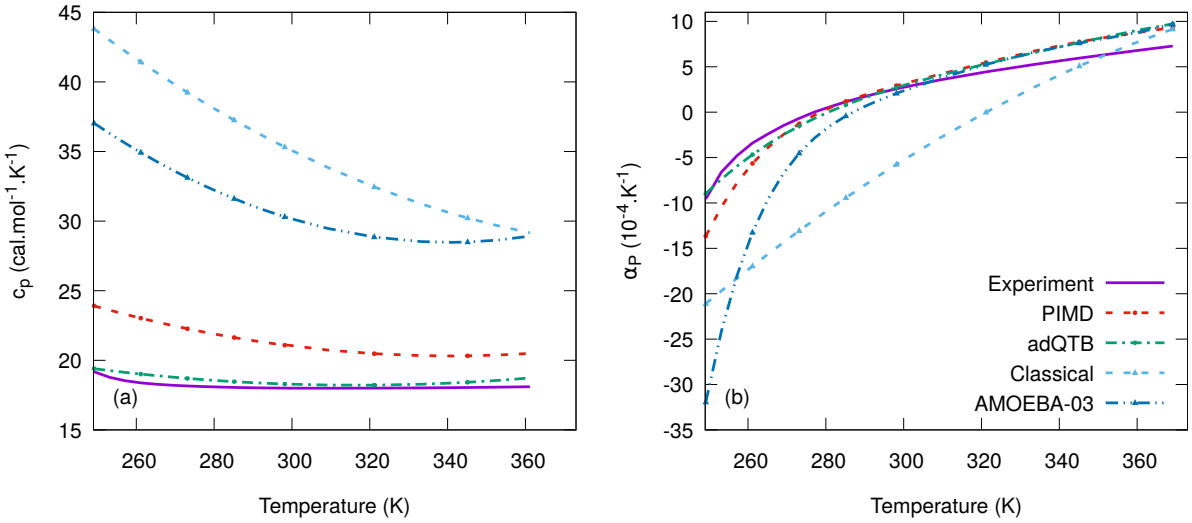


Figure 4.7: Isobaric heat capacity c_p (left) and thermal expansion coefficient α_P (right) of liquid water at $P = 1$ atm as a function of the temperature for the different AMOEBA models. The dashed line represents the results obtained with the Q-AMOEBA (PIMD) method whereas the dot-dashed lines are obtained with Q-AMOEBA (adQTB) model.

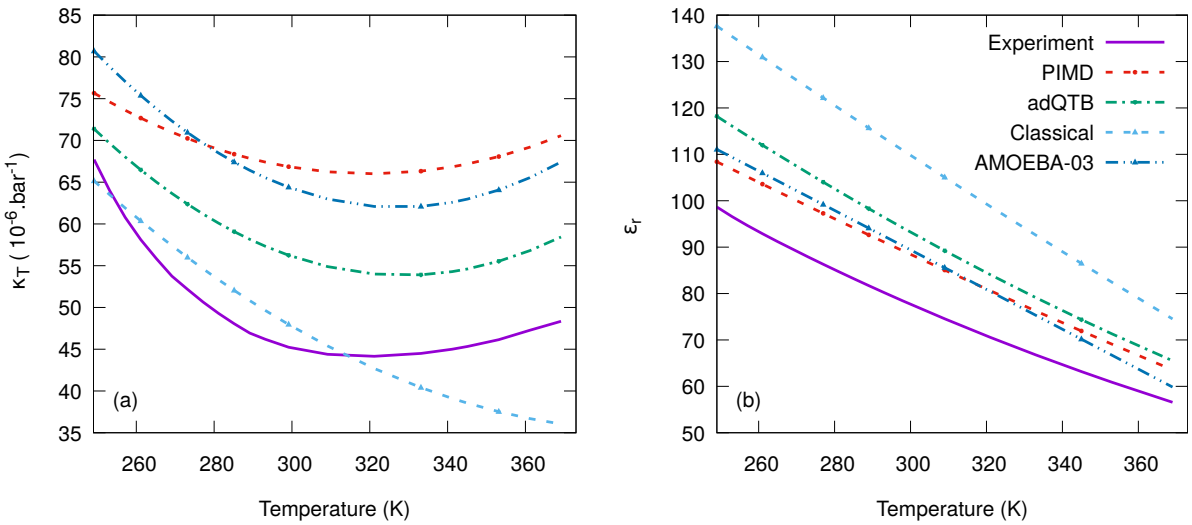


Figure 4.8: Isothermal compressibility κ_T(left) and static dielectric constant ε_r (right) of liquid water at $P = 1$ atm as a function of the temperature for the different AMOEBA models. The dashed line represents the results obtained with the Q-AMOEBA (PIMD) method whereas the dot-dashed lines are obtained with Q-AMOEBA (adQTB) model.

with Q-AMOEBA (PIMD and adQTB) allows to recover lower values which are closer to the experimental data with a smaller slope over the whole range of temperatures. The thermal expansion coefficient α_P is calculated using analytic differentiation of a polynomial fit of the simulated density $\rho(T)$:

$$\alpha_P = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P = -\frac{\mathrm{d}\ln\rho(T)}{\mathrm{d}T} \tag{4.14}$$

As expected the same trend as the density is observed for this thermodynamical observable. Classical MD gives too negative results at low temperatures which is corrected by the quantum simulations. At high temperatures, all methods behave similarly.

Because the static dielectric constant, the isothermal compressibility and the self-diffusion coefficient converge very slowly, longer NPT simulations were necessary. The isothermal compressibility characterizes the volume change as a response to an applied pressure. It can be related to the volume fluctuations in an NPT simulation:

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T = \frac{1}{k_B T}\frac{\langle V^2\rangle - \langle V\rangle^2}{\langle V\rangle} \tag{4.15}$$

Classical MD simulations with Q-AMOEBA (adQTB) amplify the variations with T and underestimate its value at high temperature in Figure 4.8 (left panel). Explicitly including NQEs with Q-AMOEBA (PIMD and adQTB) reduces its variations over the all range of temperatures. Although Q-AMOEBA (adQTB) gives slightly higher values (overestimated by approximately 6×10^6.bar^{-1} at 298.15 K compared to experiment), both Q-AMOEBA (adQTB and PIMD) yield a shape closer to the experimental curve. Moreover, the results obtained with Q-AMOEBA (adQTB) model are significantly improved with respect to the initial set of parameters given by AMOEBA03. Furthermore, AMOEBA03 and Q-AMOEBA (PIMD) yield similar results, even in the low-temperatures range which highlights the implicit inclusion of NQEs in the original parameter set.

Method	Classical	adQTB	TRPMD	Experiment
D ± 0.01 ($10^5 . \mathrm{cm}^2 . \mathrm{s}^{-1}$)	0.91	0.78	2.29	2.30

Table 4.8: Self-diffusion coefficient obtained with Q-AMOEBA (adQTB and TRPMD) with classical MD simulations.

The static dielectric constant is calculated from the fluctuations of the total dipole moment as:

$$\varepsilon_r = 1 + \frac{4\pi}{3k_B T \langle V \rangle} \left(\langle \mu^2 \rangle - \langle \mu \rangle \cdot \langle \mu \rangle \right) \tag{4.16}$$

with $\langle \mu \rangle$ the average total dipole moment defined in equation (4.12) and $\langle V \rangle$ the average volume of the simulation box. Comparing classical MD (light blue) with quantum simulations with Q-AMOEBA shows the importance of NQEs on this property, although NQEs have a smaller influence when T increases.

The self-diffusion coefficient was evaluated at 298.15 K under 1 atm using the Einstein equation:

$$\mathrm{D} = \lim_{t \to \infty} \frac{\mathrm{d}}{\mathrm{d}t} \langle |r(t) - r(t_0)|^2 \rangle \tag{4.17}$$

A mild Langevin thermostat ($\gamma = 1\,\mathrm{ps}^{-1}$) was applied in classical and TRPMD simulations. However, in adQTB, a larger friction coefficient was used ($\gamma = 20\,\mathrm{ps}^{-1}$) to compensate the ZPEL. This affects the diffusion as it was shown in Chapter 3 (Section 3.4.2) and by Maginn et. al [243]. As shown in Table 4.8, the TRPMD value is $2.29 \pm 0.01 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ which is in excellent agreement with the experimental value of $2.30 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$. Moreover, classical dynamics gives $0.91 \pm 0.01 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ which is in accordance of other studies which gave $1.5 \pm 0.1 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ [196, 237]. On the other hand, Q-AMOEBA (adQTB) value is around three times lower than the expected value. Although this slower diffusion implies a less efficient sampling [162], the computational speedup of the adQTB method compared to PIMD compensate it. It should also be noted that q-TIP4P/F water model has a classical self-diffusion coefficient around 1.5 lower than its TRPMD counterparts (see Chapter 3 Section 3.7). Moreover, the *ab-initio* based polarizable TTM2.1-F water model has a classical value of $1.5 \pm 0.1 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ and its TRPMD value gives $2.2 \pm 0.1 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ [237] which is really close to the value obtained with Q-AMOEBA. Then, it highlights that the net effect of NQEs can not be generalized to all models. It should also be noted that the self-diffusion coefficient goes from 2.2 to 1.8 $\pm 0.1 \times 10^5\,\mathrm{cm}^2 . \mathrm{s}^{-1}$ under deuteration which shows that this dynamical observables is strongly impacted by NQEs as they tend to weaken H-bonds and hence increase the diffusion. Although Q-AMOEBA slightly overestimates the net effect of NQEs, it recovers qualitatively the good behaviour when going from quantum to classical MD simulations.

To conclude, the molecular dipole moments were also studied to confirm the interpretation of NQEs and hydrogen bond strength. Figure 4.9 shows that classical MD simulations with Q-AMOEBA (adQTB) parameters give a larger dipole moment than the original AMOEBA03 water model. Moreover, at 298.15 K, classical simulations with Q-AMOEBA (adQTB) give 2.90 D whereas explicit inclusion of NQEs with Q-AMOEBA (PIMD or adQTB) give 2.76 D and 2.78 D respectively and weaken the H-bonds. The values obtained are close to those obtained in AMOEBA03 (showing again that it implicitly includes NQEs) and also with *ab initio* simulations and experiments [244, 245].

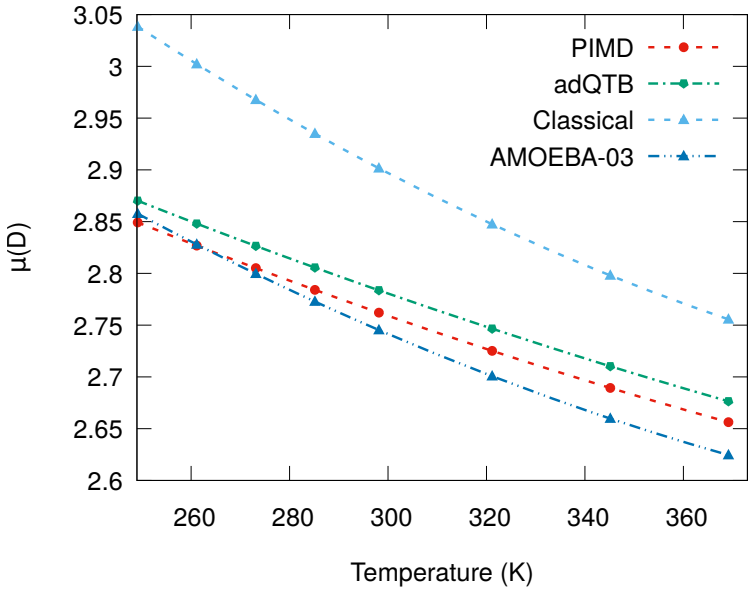


Figure 4.9: Molecular dipole moments of liquid water at $P = 1\,\text{atm}$ as a function of the temperature for the different AMOEBA model. The dashed line represents the results obtained with the Q-AMOEBA (PIMD) method whereas the dot-dashed lines are obtained with Q-AMOEBA (adQTB) model.

4.4.5 Transferability Study

Q-AMOEBA was fitted with FB using thermodynamical properties which belongs only to the liquid phase. One may be interested to see how Q-AMOEBA (PIMD and adQTB) perform on a different area of the phase diagram. Then, to study the transferability, we have decided to investigate the density of ice phase I_c but also isotopic effects with heavy water D_2O.

Heavy Water: D_2O

Isotope effects are an important maker of the relevance of NQEs, since classical MD predicts different isotopes to have identical thermal equilibrium properties. Although deuterated water is often viewed as a 'classical' analog of water, NQEs are visible while substituting H for D (or even T) [20], although D_2O (or T_2O) have smaller NQEs compared to H_2O. With the explicit treatment of NQEs, it is possible to further validate the Q-AMOEBA approach. Different thermodynamic properties are presented in Table 4.9, focusing on those known to be the most impacted by the isotopic substitution.

Both set of Q-AMOEBA parameters (PIMD and adQTB) yield the correct trends for the isotopic substitution of H to D which reduces the heat capacity c_p and increases ΔH_{vap}. The enthalpy of vaporization has an isotopic substitution amplitude overestimation of approximately 2, with a change of $\sim 0.6\,\text{kcal.mol}^{-1}$ for Q-AMOEBA with respect to the experimental value of $\sim 0.3\,\text{kcal.mol}^{-1}$. Isotopic effects increase the isobaric heat capacity by $\sim 3.6\,\text{kcal.mol}^{-1}.\text{K}^{-1}$ for Q-AMOEBA (adQTB) whereas the experimental amplitude is $\sim 1.9\,\text{kcal.mol}^{-1}.\text{K}^{-1}$. One may notice that Q-AMOEBA (PIMD) captures more accurately the increase of c_p under deuteration, although its absolute value is slightly overestimated. These differences can be explained by the intramolecular parameters of the two models which is related to the pressure estimator discussion done in Section 4.3.3. Even though the change of the properties with isotope substitution are

	H_2O			D_2O		
	PIMD	adQTB	Experiments	PIMD	adQTB	Experiments
ρ (g.cm^{-3})	0.998	0.999	0.997	1.109	1.106	1.104
ΔH_{vap} (kcal.mol^{-1})	9.65	9.55	10.52	10.25	10.28	10.85
c_p (cal.mol^{-1}.K^{-1})	21.297	18.297	18.002	22.724	21.885	20.148

Table 4.9: Thermodynamic properties computed at 298.15 K under 1 atm pressure for both Q-AMOEBA models for H_2O and D_2O compared to experimental values [246, 247].

sometimes overestimated, the different results obtained with Q-AMOEBA (PIMD and adQTB) are really promising. Since weak H-bonds tend to be further weakened by NQEs [20, 157], an overestimation of the isotopic effect is expected. Then, the discrepancies observed highlight that FB optimization leads to slightly too weak hydrogen bonds which is consistent with the slight underestimation of ΔH_{vap}. Finally, the experimental T_{MD} shifts from 277.15 K for H_2O to 284.32 K for D_2O [20]. This trend is also qualitatively observed with the shift between the quantum and classical MD curves with Q-AMOEBA (see Figure 4.6 left panel).

Ice I_c

The transferability study was further extended by studying the density of ice phase I_c. The simulations were performed at 78.0 K under 0 atm. 128 beads were used to converge Q-AMOEBA (PIMD) results. The density with both set of parameters is 0.893 and 0.906 g.cm^{-3} in PIMD and adQTB respectively whereas the experimental value is 0.931 g.cm^{-3} [60]. Classical MD simulations with AMOEBA03 yields a density $\sim$ 2% smaller than the experiment value, whereas PIMD with the same model underestimates the density by $\sim$ 6.8%. Although Q-AMOEBA (PIMD) tends to slightly overemphasize NQEs, its corresponding density is still improved compared to PIMD simulations with the original AMOEBA03 parameters which highlights some error compensations and an implicit inclusion of NQEs in the FF parametrization. Then, it reduces the transferability of AMOEBA03 and one may expect Q-AMOEBA to offer more robust predictions in this low temperature range.

4.5 Conclusion

With the extensive development of hardware and software, the different models used in MD become more and more complex and go beyond the first pair-wise approximation to model electrostatic interactions. A major limitation of the basic FFs is the lack of a fine description of many-body polarization which reduces their transferability. To circumvent that, much efforts have been made to include it. Then, new FFs were developed which include polarization from the onset, AMOEBA being one of them. Although the description becomes more precise, it lacks an explicit inclusion of NQEs. Usually, they are included implicitly by fitting the different parameters used in the model to obtain different thermodynamical properties (such as the density ρ or the enthalpy of vaporization ΔH_{vap}). Then, using PIMD or adQTB methods would lead poor results due to the double counting of NQEs. Therefore, we have decided to develop a new water model, Q-AMOEBA, which is based on the AMOEBA functional form and it is optimized to accurately reproduce energies in the gas phase as well as condensed phase properties with a quantum description of the nuclei using PIMD

and adQTB. The final set of Q-AMOEBA parameters obtained while including NQEs led to an improved intramolecular potential with the O-H bond length and H-O-H angle parameters closer to their gas phase values compared to the initial set of parameters AMOEBA03. With the different thermodynamical properties shown through the Chapter, our new quantum water model is very robust even on properties outside the reference data used in FB during the fitting procedure such as the thermal expansion coefficient α_P. The self-diffusion constant was also studied and, although Q-AMOEBA (adQTB) tends to underestimate its value, this dynamical observable is strongly affected by NQEs and the good trend is captured by our new quantum water model. In order to study further the robustness of Q-AMOEBA, isotopic effects and heavy water were also studied. Because they are not present in classical MD simulations, isotope effects are a valuable indicator of NQEs. Q-AMOEBA is able to capture the experimental trends associated to the deuteration of water, although it slightly overestimates the impact of NQEs. We interpreted it as a sign that hydrogen bonds strength is slightly underestimated with Q-AMOEBA. These already promising results should be improved with more advanced functional forms such as AMOEBA+ (which includes charge flux, charge transfer and charge penetration effects [248, 249]) and by using better *ab initio* references computations including higher levels of coupled cluster excitations. It should also be noted that the functional form for the intramolecular terms of AMOEBA03 (and hence Q-AMOEBA) were taken from MM3 FF. Therefore, changing these parameters can improve our Q-AMOEBA model and then Q-AMOEBA+. Moreover, the inclusion of the many-body polarization effects lead to a non-trivial interplay with NQEs and need to be further studied with more extended model of dispersion or polarization [250]. Since adQTB simulations can be performed at near classical cost, the improved Q-AMOEBA model can easily be extended to organic molecules and proteins opening the possibility for large scale study of the importance of NQEs in biophysics.

Conclusion and Perspectives

Throughout this thesis, NQEs have been studied using different frameworks. The Feynman path integrals provide an elegant framework to include them explicitly in MD simulations [23, 24]. However, due to the number of beads required to provide enough energy to account for ZPE, this method is computationally expensive reducing its range of applicability. To circumvent that, large efforts have been devoted to develop different methods which aim to include NQEs at lower computational cost. Following this idea, the QTB approach was introduced. It allows to include both thermal and quantum fluctuations in the thermostat, via a stochastic force [25, 26]. Because it aims to impose the second-kind FDT, it suffers from ZPE leakage, where some energy is unphysically transferred from the high frequency modes to the low frequency ones. [146, 147] Such unphysical energy flow comes from the competition between classical Newton's equation of motion which drives the system towards equipartition of energy and the thermostat that tries to impose the quantum distribution. Therefore, the adQTB [27] was developed which uses the first-kind FDT as a criterion to correct the leakage. The adQTB has already provided accurate results on model systems which motivated the extended study of this method on more realistic ones. NQEs are ubiquitous in water with its ZPE which subtly influence hydrogen bonding due to a competition between delocalization over O-H stretching and bending modes [20, 21]. Furthermore, given the low mass of hydrogen NQEs cannot be neglected. It has been shown that they have a strong influence on the structure and the dynamics of this compound. Moreover, it is an important benchmark for methods which aim to include NQEs. It is also an important step toward biological systems Thus, at the beginning of this thesis, the adQTB seemed to be a promising alternative to the computationally heavy PIMD to study NQEs in highly anharmonic systems such as water. Therefore, the adQTB and PIMD were implemented in the Tinker-HP software (CPU and GPU) whose purpose is to routinely perform MD simulations [5, 28].

We began our study with the q-TIP4P/F water model which was specifically parameterized using PIMD simulations [30]. Because this potential is broadly used in the literature as a benchmark for NQEs simulation method, the different results obtained with adQTB method are presented in Chapter 3. The different parameters involved in the adQTB method, such as γ and ω_{cut} were studied in details to know how they influence the results and how to choose them to reduce their impact on the dynamics. Compared to the previous studies [15, 112], $\gamma_r(\omega)$ coefficients were averaged on the three directions of space and also over same-type atoms. On top of that, a deconvolution procedure [186] has been used to correct the broadening of the spectrum caused by the friction force. It also aims to correct the overestimation and underestimation of the average kinetic and potential energies terms respectively. Altought water is highly anharmonic and QTB suffers from a major ZPE leakage, adQTB manages to recover almost identical results to PIMD on various properties such as the density and infrared absorption spectra with a computational cost comparable

to classical MD simulations. However, some questions remain open such as the possibility to improve the results obtained for the self-diffusion coefficient with a fast-forward Langevin scheme [205]. The fact that the adQTB method is not formally derivable from first principles and also the need to study how the method behaves in inhomogeneous systems. Nonetheless, with its low computational cost, the different improvements made to correct the kinetic and pressure estimators, and the results almost identical to the PIMD reference, the adQTB method has become a promising method to include NQEs explicitly in MD simulation.

Although the results obtained in Chapter 3 were really encouraging, they were obtained with a very inexpensive and simple water model. To extend the study of the adQTB method and NQEs to more complex and realistic systems, a more sophisticated model was needed. More sophisticated water models are available such as AMOEBA [31, 206], GEM [208, 251], HIPPO [252] or SIBFA [207, 253]. However, their parameters were obtained by fitting on both quantum electronic structure calculations and some experimentally measured properties. Then, using the PIMD or adQTB methods on one of these FFs will result in a double counting of NQEs. To circumvent that and be able to study NQEs on more complex potential forms, we have decided to develop a new FF by using the AMOEBA functional and explicitly including NQEs on the onset. The methodology to develop this new polarizable water model, called Q-AMOEBA, is presented in Chapter 4. We firstly intended to obtain our new water model by only using *ab initio* data. However, it showed that the different data used were not accurate enough to correctly reproduce the density, highlighting that PIMD and adQTB methods are as good as the PES is. Therefore, more accurate calculations including higher level of coupled cluster excitations must be performed to be able to only use *ab initio* data in the fit procedure. To circumvent that thermodynamical properties were added to our fit to obtain reliable results. To do so, we have also used the adQTB method to derive the new parameters of Q-AMOEBA. The O-H bond length and H-O-H angle parameters were also reduced which shows that explicitly including NQEs leads to an improved intramolecular potential. We have also used the PIMD method to compute different properties and study the accuracy of our strategy. All comparisons between PIMD and adQTB results were presented in Chapter 4 and they demonstrate an improvement on properties of liquid water with respect to the original AMOEBA model, making our new water model more robust. This study also revealed that many-body polarization effects lead to a non-trivial interplay with NQEs which highlights that the effect of NQEs on different observables are more subtle than usually assumed and can no longer be generalized a priory to all functional forms.

Nonetheless, our Q-AMOEBA water model could be improved by using even more complex FF functional forms such as the one used in AMOEBA+ [248, 249]. This model includes short-range electrostatic Charge Penetration (CP), 2-body Charge Transfer (CT) and Charge Flux (CF). It has been shown that these terms improve the intra and intermolecular potential with respect to AMOEBA. Therefore, including them in the fit used to derive Q-AMOEBA+ will certainly lead to even better results. Although this thesis presents an extensive study of NQEs in pure liquid water (and these effects on hydrogen bond networks) by using different potentials, it is crucial in the future to extend this study to inhomogeneous systems. Studying how the adQTB behaves in such systems can lead to the development of biological models, such as AMOEBA BIO [254–256], by explicitly including NQEs. Having a low computational cost method to include NQEs in complex and big systems (several millions of atoms) could lead to the discovery of new phenomenon in such complex systems. It could also be interesting to use Machine Learning on the intramolecular parts of Q-AMOEBA to see how the different results are impacted. Moreover, the study performed in Chapter 4 highlights that the inclusion of many body polarization effects leads to non-trivial interplay with NQEs. Then, it could be interesting to study Q-AMOEBA with a more developed polarization and dispersion model

such as the Bond Capacity Charge Polarization Model [250] or the recently developed many-body dispersion model presented in Ref [257].

Simulating NQEs for large and complex systems is still an issue that adQTB tries to address thanks to its low computational requirements. Therefore, with the implementation of this method in the multi-CPU and multi-GPU Tinker-HP software, it is now possible to study NQEs on a broader range of systems. However, if one may want to study NQEs with PIMD and compared the different results obtained with adQTB, it is also possible to do it with the implementation of PIMD in Tinker-HP. Therefore, this thesis provides a proof of concept to routine production MD simulations while explicitly including NQEs and extend their study within large scale simulations of complex systems.

Publications

The following works have been published:

- *Nuclear Quantum Effects in Liquid Water at Near Classical Computational Cost Using the Adaptive Quantum Thermal Bath*
  Nastasia Mauger, Thomas Plé, Louis Lagardère, Sara Bonella, Étienne Mangaud, Jean-Philip Piquemal, and Simon Huppert, J. Phys. Chem. Lett. 2021, **12**, 34, 8285–8291

- *Improving Condensed Phase Water Dynamics with Explicit Nuclear Quantum Effects: the Polarizable Q-AMOEBA Force Field*
  Nastasia Mauger, Thomas Plé, Louis Lagardère, Simon Huppert, Jean-Philip Piquemal, J. Phys. Chem. B, **126**, 43, 8813-8826

The following work has to be submitted shortly:

- *Routine Molecular Dynamics Simulations Including Nuclear Quantum Effects with Tinker-HP on GPUs*
  Thomas Plé, Nastasia Mauger, Louis Lagardère, Simon Huppert, and Jean-Philip Piquemal

THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL Letter

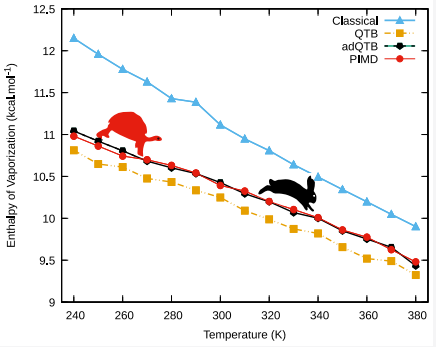
Nuclear Quantum Effects in Liquid Water at Near Classical Computational Cost Using the Adaptive Quantum Thermal Bath

Nastasia Mauger,[#] Thomas Plé,[#] Louis Lagardère,[*] Sara Bonella, Étienne Mangaud, Jean-Philip Piquemal,[*] and Simon Huppert[*]

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 8285−8291 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We demonstrate the accuracy and efficiency of a recently introduced approach to account for nuclear quantum effects (NQEs) in molecular simulations: the adaptive quantum thermal bath (adQTB). In this method, zero-point energy is introduced through a generalized Langevin thermostat designed to precisely enforce the quantum fluctuation−dissipation theorem. We propose a refined adQTB algorithm with improved accuracy and report adQTB simulations of liquid water. Through extensive comparison with reference path integral calculations, we demonstrate that it provides excellent accuracy for a broad range of structural and thermodynamic observables as well as infrared vibrational spectra. The adQTB has a computational cost comparable to that of classical molecular dynamics, enabling simulations of up to millions of degrees of freedom.

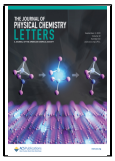


Nuclear quantum effects (NQEs) play a major role in a wide range of physical and chemical processes where light atoms, especially hydrogen, are involved.[6,49,50,66] In particular, a few studies point to their importance in biological systems,[2,58,71] where hydrogen bonding is ubiquitous, but realistic atomic-scale simulations in that area remain scarce. For such large and complex systems, the most common approach has been to include NQEs implicitly by fitting analytical potential energy surface models in order to recover experimental thermodynamic properties when performing simulations with classical nuclei.[24,53] This strategy potentially limits transferability, and its ability to make predictions outside the fitting data set is questionable. Furthermore, the recent developments of new-generation polarizable force fields[24,42,43,47,62] and machine learning (ML) potentials[24,51,68,69,72] have opened perspectives for atomistic simulations of condensed matter systems. These approaches enable high-fidelity modeling of the Born−Oppenheimer (BO) energy and reproduce advanced quantum-chemical calculations at a fraction of their computational cost. When reaching such precision on the BO energy, it becomes crucial to account for NQEs explicitly in order to accurately reproduce experimental observations and take full advantage of the high accuracy achieved.[22,27,55,56]

The conceptual and computational complexity of the methods that account for NQEs explicitly has hindered their spread to a broad community. Reliable results can be obtained in the imaginary-time path integral (PI) framework[21,29] by simulation of multiple classical replicas of the system (also called beads). PI provides a numerically exact reference for static properties (approximations have also been derived for dynamical observables, as discussed below), but its numerical cost increases linearly with the number of replicas and can become very large compared with classical molecular dynamics (MD). Several solutions have been proposed to mitigate this cost, such as multiple timestepping in real and imaginary time.[23,38,45,46] However, this method is based on a decomposition of the energy as a sum of cheap high-frequency and expensive low-frequency terms that is not always feasible (in particular in ML approaches). Other developments, such as high-order PI[37,57] or PI perturbation theory,[60,61] allow the number of necessary replicas to be decreased, but the computational overhead remains important, typically increasing the simulation load by an order of magnitude for hydrogen-bonded systems at room temperature.

Recently, a different approach was introduced for the explicit treatment of NQEs with the quantum thermal bath (QTB)[12,26] and the related quantum thermostat,[16,17] relying on generalized Langevin thermostats to approximate the zero-point motion of the nuclei. Although elegant and inexpensive, these methods suffer from zero-point energy (ZPE) leakage from high- to low-frequency modes, which can lead to massive errors.[9,34] One possible workaround is to combine the

Received: May 31, 2021
Accepted: August 17, 2021
Published: August 24, 2021



generalized thermostat approach with path integrals.[10,19] Even though the number of required replicas is reduced compared with standard PIMD simulations, the computational cost remains significant (at least six beads are needed for water under ambient conditions[18]). In this Letter, we focus on an alternative approach, the adaptive QTB (adQTB), which completely avoids resorting to PI.

In adQTB, a quantitative criterion derived from the fluctuation−dissipation theorem (FDT) is used to compensate for the ZPE leakage directly. The method was successfully tested on model systems,[44] but its applicability to more realistic problems remained to be demonstrated. In the following, we report the main theoretical aspects of the QTB and adQTB methodologies and introduce two refinements to the adQTB algorithm that improve its efficiency and accuracy and broaden the range of its possible applications, in particular enabling reliable constant-pressure simulations. We then apply the method to liquid water. Careful comparison with PI references for structural and thermodynamic properties as well as infrared absorption spectra (IRS) shows that, contrary to standard QTB, which is plagued by massive ZPE leakage, adQTB is able to capture NQEs with remarkable accuracy, while its computational overhead remains limited to less than 25% compared with classical MD, allowing the system size to be scaled up to over a million atoms.

In (ad)QTB simulations, each nuclear degree of freedom follows a Langevin equation:[26]

$$m_i \frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2}(t) = -\frac{\partial V}{\partial r_i} - m_i \gamma \frac{\mathrm{d}r_i}{\mathrm{d}t}(t) + R_i(t) \quad i = 1, ..., 3N \tag{1}$$

where $V(r_1, ..., r_i, ..., r_{3N})$ is the interatomic potential (i denotes both the atom number and the direction x, y, or z). Equation 1 comprises a dissipative force (with friction coefficient γ) balanced by a random force $R_i(t)$ that injects energy into the system. In classical Langevin dynamics, $R_i(t)$ is a white-noise function whose amplitude is proportional to the temperature. In QTB, the random force is colored with the following correlation spectrum:

$$C_{R_i R_j}(\omega) = 2m_i \gamma_i(\omega) \theta(\omega, T) \delta_i^j \tag{2}$$

where $\gamma_i(\omega)$ is the random force amplitude, δ_{ij} is the Kronecker delta, and

$$\theta(\omega, T) = \hbar\omega \left[\frac{1}{2} + \frac{1}{\mathrm{e}^{\hbar\omega/k_\mathrm{B}T} - 1} \right] \tag{3}$$

corresponds to the average thermal energy in a quantum harmonic oscillator at frequency ω and temperature T. Therefore, the aim of the QTB is to account for ZPE contributions in an otherwise classical dynamics by thermalizing each vibrational mode with an effective energy $\theta(\omega, T)$ instead of the classical thermal energy $k_\mathrm{B}T$. However, in the original formulation of the QTB (where $\gamma_i(\omega) = \gamma \ \forall\omega$), the ZPE provided to high-frequency modes leaks toward low-frequency modes, which leads to an incorrect energy distribution and can dramatically alter the results. In adQTB, this leakage is quantified precisely using a general result of linear response theory: the quantum FDT.[39] For each degree of freedom i, we define the deviation from the FDT as

$$\Delta_{\mathrm{FDT},i}(\omega) = \mathrm{Re}[C_{v_i R_i}(\omega)] - m_i \gamma_i(\omega) C_{v_i v_i}(\omega) \tag{4}$$

in which $v_i = \frac{\mathrm{d}r_i}{\mathrm{d}t}$ is the velocity and $C_{v_i v_i}(\omega)$ and $C_{v_i R_i}(\omega)$ are its autocorrelation and its cross-correlation spectrum with the random force R_i, respectively. The FDT characterizes the frequency-dependent distribution of energy in a quantum system at thermal equilibrium. It implies that $\Delta_{\mathrm{FDT},i}(\omega)$ should be zero for any ω, a condition violated in standard QTB because of the ZPE leakage. In adQTB, $\Delta_{\mathrm{FDT},i}(\omega)$ is estimated at regular intervals, and the coefficients $\gamma_i(\omega)$ are adjusted on the fly via first-order dynamics to correct for this violation: a negative $\Delta_{\mathrm{FDT},i}(\omega)$ reveals an excess of energy at frequency ω, so $\gamma_i(\omega)$ is reduced, and conversely for positive deviations. The adQTB results are produced once the $\gamma_i(\omega)$ are adapted and $\Delta_{\mathrm{FDT},i}(\omega)$ vanishes on average.

Here we introduce two refinements with respect to ref 44, both of which are presented in full detail in the Supporting Information. First, to improve the adaptation efficiency, the coefficients $\gamma_i(\omega)$ are adjusted according to the mean FDT deviation averaged over all equivalent degrees of freedom (i.e., over the three directions and over all same-type atoms). Second, we account for the fact that because of the spectral broadening induced by the friction force, the QTB (and adQTB) tends to slightly underestimate the average potential energy and overestimate the kinetic energy. This error (unrelated to ZPE leakage) can be predicted and quantified for a harmonic oscillator.[3,4] We use this harmonic reference and the deconvolution procedure of ref 67 to correct for this inaccuracy: we slightly modify $\theta(\omega, T)$ to compensate for the effect of γ on the potential energy, while the kinetic energy is corrected a posteriori. The kinetic energy correction is significant (more than 10%) and essential to enable reliable isobaric simulations, as its neglect causes large errors in the pressure estimation.

The role of NQEs in liquid water has been extensively investigated both experimentally and theoretically.[36,52,54,63,64] It also represents a major challenge for the adQTB, as massive ZPE leakage takes place from the high-frequency intramolecular vibrations (O−H stretching and H−O−H bending modes) to the slow intermolecular motion.[31,34] Moreover, net NQEs on the structural properties of water are relatively weak because of the competition between two opposite trends: the stretching ZPE strengthens hydrogen bonding, while the bending ZPE weakens it.[18,41] The ability of the adQTB to capture this subtle balance is an important indication of its robustness that opens perspectives for its broader application.

Interatomic interactions were modeled by the q-TIP4P/F potential,[32] which was included in a local version of the Tinker-HP massively parallel package,[40] where we also implemented PIMD and (ad)QTB. Simulations were performed with 1000 water molecules. The PIMD simulations were essentially converged with 32 beads (the number typically reported in the literature) and required short time steps; we used a time step of 0.2 fs for all of the methods and checked that increasing it to 1 fs had only a limited effect on the accuracy of the adQTB results. In classical Langevin MD and PIMD simulations, static averages are independent of the parameter γ, and we used $\gamma = 1$ ps^{-1} in both cases to limit its effect on the dynamical properties. On the other hand, adQTB requires relatively large friction coefficients γ to prevent vanishing of $\gamma_i(\omega)$ during adaptation (which would result in incorrect compensation for the ZPE leakage[44]). We used $\gamma = 20$ ps^{-1} for all of the QTB and adQTB simulations (the influence of these parameters and the scalabilty of the

https://doi.org/10.1021/acs.jpclett.1c01722
J. Phys. Chem. Lett. 2021, 12, 8285−8291

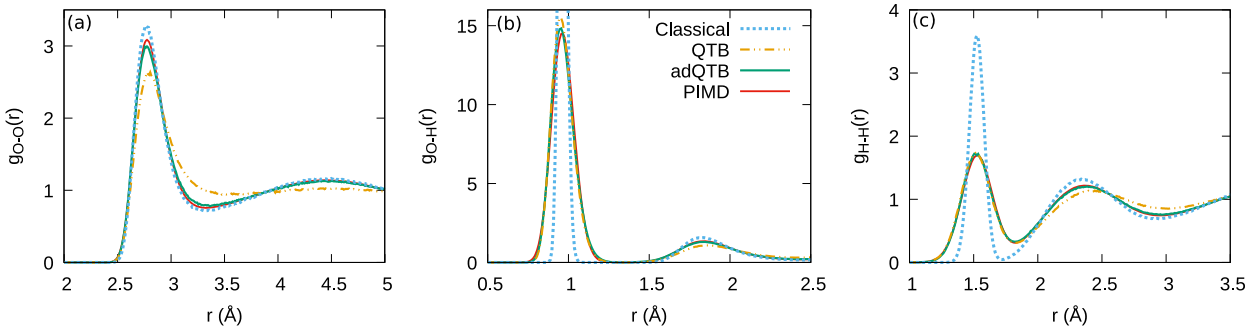


Figure 1. Radial distribution functions at 300 K and constant volume corresponding to the density ρ = 0.995 g·cm^{-3}.

algorithm for large systems are assessed in the Supporting Information).

In Figure 1, the QTB and adQTB radial distribution functions (RDFs) are compared with their classical and PIMD counterparts. The most salient NQE for this observable is the strong broadening of the intramolecular peaks caused by ZPE in the O−H and H−H RDFs. This effect is very well captured by the adQTB simulations, whereas it is slightly underestimated by the standard QTB as a result of ZPE leakage. Apart from this, the classical and quantum RDFs are very similar because of the aforementioned competition of NQEs. In standard QTB simulations, the leakage of the intramolecular ZPE destabilizes the hydrogen-bonding network completely, and the intermolecular peaks are excessively broadened. In contrast, the adQTB procedure efficiently suppresses the leakage, and the corresponding curves almost superimpose with the PIMD reference.

This analysis is further confirmed by Table I, which reports the averages of the different q-TIP4P/F energy terms.

Table I. Observables at 300 K: The Kinetic Energy (E_k), Angular Bending (AB), Bond Stretching (BS), Van der Waals (VdW), and Coulomb (Coul.) Energy Terms of the q-TIP4P/F Potential[32] (All Are Reported in kcal·mol^{-1} per Water Molecule; the Standard Error Is <0.01 kcal·mol^{-1}), the Average Oxygen−Hydrogen Distance (r_{OH}), and the Average HOH Molecular Angle (θ_{HOH})

	E_k	AB	BS	VdW	Coul.	r_{OH} (Å)	θ_{HOH} (deg)
classical	2.69	0.41	1.18	2.20	−14.00	0.96	104.8
QTB	8.39	1.23	5.81	1.72	−12.38	0.98	104.6
adQTB	8.60	1.17	6.37	2.11	−13.76	0.98	104.7
PIMD	8.41	1.17	6.26	2.15	−13.87	0.98	104.7

Intermolecular interactions (labeled Coul. and VdW) are only slightly affected by NQEs, and their classical and PIMD values are close. In standard QTB, the total intermolecular energy is overestimated by more than 1 kcal·mol^{-1} as a result of ZPE leakage, but this is well-corrected in adQTB, where accurate values are recovered. The adQTB is remarkably precise for intramolecular energies (labeled AB and BS) and for the kinetic energy (which comprise large amounts of ZPE). It also captures the elongation of the OH distance induced by NQEs, while the molecular angle is essentially unaffected. The dielectric constant computed from the adQTB simulations at 300 K is 57, in good agreement with our PIMD estimation of 58 and with the value in ref 32 given the relatively large statistical uncertainties.

Although PIMD provides a numerically exact reference for static quantum properties, the computation of dynamical observables, such as infrared absorption spectra, represents a much steeper theoretical challenge and is a subject of intense research.[5,8,15,33,59,65,70] There is no reference method to compute IRS exactly while accounting for NQEs in large systems, but various approximations have been developed.[13,14,25,48] Recently, Benson et al. compared different state-of-the-art approximate methods for the calculation of IRS in liquid water and ice.[7] They showed that the linearized semiclassical initial value representation (LSC-IVR) method,[48] in which time-correlation functions are computed from short classical trajectories initialized from an approximate sampling of the Wigner distribution, provides the most accurate IRS within their broad set of approaches, while PI-based thermostated ring-polymer MD (TRPMD)[65] is presented as the cheapest available approach yielding reliable results. QTB has formerly been used with some success as an empirical method to compute approximate IRS.[11,12] Although not formally derivable from first principles except for the harmonic

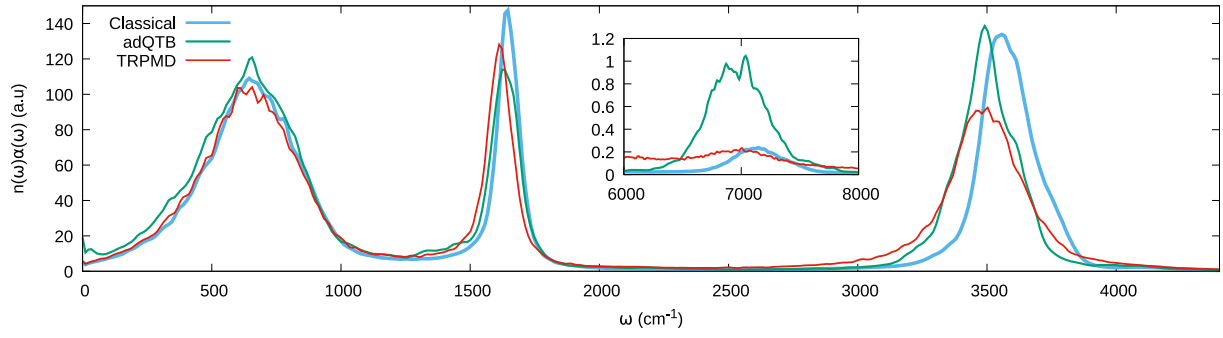


Figure 2. Infrared absorption spectra at 300 K (in arbitrary units).

https://doi.org/10.1021/acs.jpclett.1c01722
J. Phys. Chem. Lett. 2021, 12, 8285−8291

oscillator case, the use of QTB and adQTB for IRS calculations can be justified qualitatively by noting that the short-time dynamics is affected only a little by the thermostat and thus is essentially classical. Therefore, much like LSC-IVR, the QTB combines classical dynamics with approximate quantum initial value sampling. Furthermore, the deconvolution procedure of ref 67 efficiently eliminates the main effect of the thermostat, namely, the broadening of the spectral peaks.

Figure 2 compares IRS computed with adQTB to those obtained using classical MD and TRPMD (for which a mild Langevin thermostat with $\gamma = 1$ ps^{-1} was applied). Compared with TRPMD, the low-frequency absorption band computed with adQTB is slightly more intense, and the bending peak (around 1500 cm^{-1}) is a little blue-shifted and broadened. The OH stretching peak at 3500 cm^{-1} is sharper in adQTB than in TRPMD, and its overtone at 7000 cm^{-1} has a much larger intensity. These two discrepancies are in favor of the adQTB approach since TRPMD has been shown to cause spurious broadening of the spectral features and to strongly underestimate anharmonic resonances.[7] Overall, the adQTB IRS are very similar to the LSC-IVR results reported in ref 7. This should be further confirmed by studies on different systems, but it is extremely promising given the almost classical computational cost of adQTB.

The dynamical properties related to slow molecular motions, on the other hand, cannot be quantitatively assessed in our present adQTB implementation because of the need for relatively large friction coefficients. The diffusion coefficient $D \approx 0.8$ cm^2·s^{-1} is underestimated by almost a factor of 3 with respect to its RPMD value[32] (a similar decrease in D is observed in classical Langevin MD using $\gamma = 20$ ps^{-1}). The deconvolution procedure is of no help here, since D corresponds to the zero-frequency component of the vibrational spectrum, and the deconvolution does not provide reliable results in that spectral region.[67] Improved diffusion estimates might be obtained in future works by decreasing γ selectively at low frequencies using a generalized friction force or by appropriately redesigning the adQTB algorithm, for example using the recently introduced fast-forward Langevin method.[35]

We now explore the use of adQTB to perform fixed-pressure simulations using a Langevin piston barostat.[20,28] Pressure is a challenging quantity to evaluate in the (ad)QTB framework: its estimator is a difference between two large terms that almost cancel (a potential term and a kinetic term, of the order of 10^5 atm each). Therefore, even small inaccuracies in either of these contributions can result in non-negligible errors (see the Supporting Information). The results obtained for the density as a function of temperature at $P = 1$ atm are shown in Figure 3.

Because of the competition between NQEs, the classical and PIMD results are very similar, both showing a characteristic bell shape with a maximum at around 280 K. NQEs are responsible for only a small decrease in the density in the intermediate temperature range (270−330 K). The standard QTB completely fails to capture this temperature dependence: it gives values that decrease monotonously and strongly overestimates the variations of the density. Compensating for the leakage in adQTB results in recovery of the overall bell shape and good agreement with the PIMD reference. In the intermediate temperature range (most relevant for biological systems), adQTB is very accurate. The curvature of the density curve is only slightly underestimated, leading to small errors on

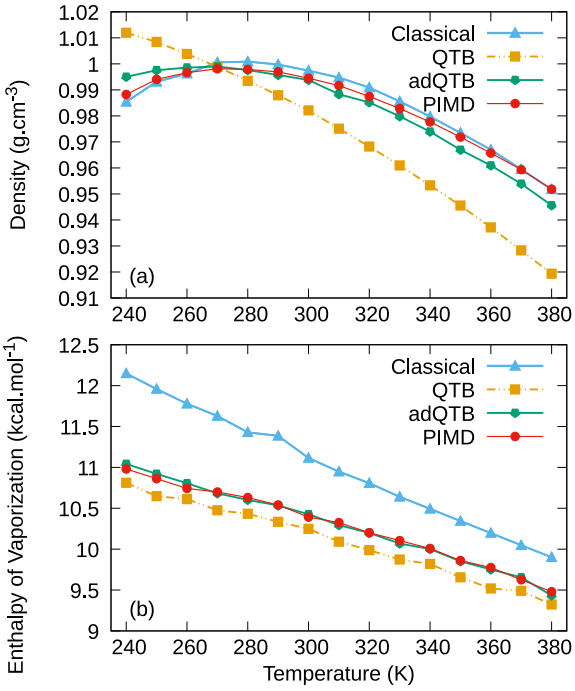


Figure 3. (a) Density of liquid water and (b) enthalpy of vaporization at constant pressure ($P = 1$ atm).

the order of 0.005 g·cm^{-3} in the low-temperature and high-temperature limits. It should be noted that in barostated simulations, ZPE leakage can take place from the atomic system toward the fictitious piston degree of freedom, but this leakage can easily be avoided by appropriate choices of the piston mass and friction parameters (see the Supporting Information).

These results show that adQTB can be a useful and inexpensive tool for constant-pressure simulations of physical and chemical properties. As an illustration, we present in Figure 3b the enthalpy of vaporization (ΔH_{vap}) computed from the same isobaric simulations. The classical ΔH_{vap} is systematically overestimated compared with the corresponding PIMD values.[30,62] When NQEs are included with the standard QTB, ΔH_{vap} decreases markedly and becomes even underestimated, but this is due to ZPE leakage, and the adQTB recovers an almost perfect agreement with the PIMD reference.

Finally, we discuss the computational overhead of the adQTB simulations with respect to classical Langevin MD. A first additional cost comes from the generation of the colored random forces and the adaptation of the $\gamma_i(\omega)$ coefficients. It represents approximately 20% of the total simulation time, and the scalability tests provided in the Supporting Information show that even for systems containing over 1 million atoms it remains less than 25% in our present implementation—which will be further accelerated using graphics processing units (GPUs).[1] The q-TIP4P/F water model is particularly inexpensive, and we expect this overhead to become negligible in comparison with atomic force calculations with more realistic models. A second additional cost comes from the adaptation procedure, which requires time for the $\gamma_i(\omega)$ to converge. This necessary time can vary from one system to another. In the Supporting Information we show that in our liquid water simulations, with an appropriate choice of

https://doi.org/10.1021/acs.jpclett.1c01722
J. Phys. Chem. Lett. 2021, 12, 8285−8291

adaptation parameters, the $\gamma_i(\omega)$ coefficients can converge in about 10 ps. The minimum adaptation time is thus small compared with the several nanoseconds required to reach statistical convergence on some of the physical observables, such as the density and the dielectric constant.

The adQTB renews the original promise of the QTB method to provide approximate quantum simulations at an almost classical cost, but with much-improved reliability. It is a promising alternative to PI methods to account for NQEs explicitely in the calculation of static properties as well as vibrational spectra. Combined with accurate ML potentials or polarizable force fields, it should provide a powerful tool with broad applications, in particular for the large-scale simulations required in biophysics and biochemistry.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c01722.

Implementation details about the adQTB adaptation procedure, description of how the effect of the friction force is deconvoluted to recover more accurate results, specificity of barostated simulations in adQTB, and scalability tests of the method up to over 1 million atoms (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Louis Lagardère − *Sorbonne Université, LCT, UMR 7616 CNRS, F-75005 Paris, France*; Email: louis.lagardere@sorbonne-universite.fr

Jean-Philip Piquemal − *Sorbonne Université, LCT, UMR 7616 CNRS, F-75005 Paris, France; Institut Universitaire de France, 75005 Paris, France; Department of Biomedical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-6615-9426; Email: jean-philip.piquemal@sorbonne-universite.fr

Simon Huppert − *CNRS, Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588, F-75005 Paris, France;* orcid.org/0000-0002-7666-559X; Email: simon.huppert@sorbonne-universite.fr

Authors

Nastasia Mauger − *Sorbonne Université, LCT, UMR 7616 CNRS, F-75005 Paris, France*

Thomas Plé − *CNRS, Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588, F-75005 Paris, France*

Sara Bonella − *CECAM Centre Européen de Calcul Atomique et Moléculaire, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland;* orcid.org/0000-0003-4131-2513

Étienne Mangaud − *CNRS, Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588, F-75005 Paris, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c01722

Author Contributions

#N.M. and T.P. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was made possible thanks to funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement 810367), Project EMC2. Computations were performed at CINES on the Occigen machine on Grant A0070707671. The authors are grateful to Fabio Finocchi and Philippe Depondt for many interesting discussions.

■ REFERENCES

(1) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-hp: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using gpus and multi-gpu systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034−2053. PMID: 33755446.

(2) Agarwal, P. K.; Billeter, S. R.; Rajagopalan, P. R.; Benkovic, S. J.; Hammes-Schiffer, S. Network of coupled promoting motions in enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2794−2799.

(3) Barrat, J.-L.; Rodney, D. Portable implementation of a quantum thermal bath for molecular dynamics simulations. *J. Stat. Phys.* **2011**, *144*, 679−689.

(4) Basire, M.; Borgis, D.; Vuilleumier, R. Computing Wigner distributions and time correlation functions using the quantum thermal bath method: application to proton transfer spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12591−12601.

(5) Basire, M.; Mouhat, F.; Fraux, G.; Bordage, A.; Hazemann, J.-L.; Louvel, M.; Spezia, R.; Bonella, S.; Vuilleumier, R. Fermi resonance in CO_2: Mode assignment and quantum nuclear effects from first principles molecular dynamics. *J. Chem. Phys.* **2017**, *146*, 134102.

(6) Benoit, M.; Marx, D.; Parrinello, M. Tunnelling and zero-point motion in high-pressure ice. *Nature* **1998**, *392*, 258−261.

(7) Benson, R. L.; Trenins, G.; Althorpe, S. C. Which quantum statistics-classical dynamics method is best for water? *Faraday Discuss.* **2020**, *221*, 350−366.

(8) Beutier, J.; Vuilleumier, R.; Bonella, S.; Ciccotti, G. Gas phase infrared spectra from quasi-classical kubo time correlation functions. *Mol. Phys.* **2015**, *113*, 2894−2904.

(9) Brieuc, F.; Bronstein, Y.; Dammak, H.; Depondt, P.; Finocchi, F.; Hayoun, M. Zero-point energy leakage in quantum thermal bath molecular dynamics simulations. *J. Chem. Theory Comput.* **2016**, *12*, 5688−5697.

(10) Brieuc, F.; Dammak, H.; Hayoun, M. Quantum thermal bath for path integral molecular dynamics simulation. *J. Chem. Theory Comput.* **2016**, *12*, 1351−1359.

(11) 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. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 024104.

(12) Bronstein, Y.; Depondt, P.; Finocchi, F.; Saitta, A. M. Quantum-driven phase transition in ice described via an efficient Langevin approach. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 214101.

(13) Cao, J.; Voth, G. A. A new perspective on quantum time correlation functions. *J. Chem. Phys.* **1993**, *99*, 10070−10073.

(14) Cao, J.; Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties. *J. Chem. Phys.* **1994**, *100*, 5093−5105.

(15) Ceotto, M.; Di Liberto, G.; Conte, R. Semiclassical "divide-and-conquer" method for spectroscopic calculations of high dimensional molecular systems. *Phys. Rev. Lett.* **2017**, *119*, 010401.

(16) Ceriotti, M.; Bussi, G.; Parrinello, M. Nuclear quantum effects in solids using a colored-noise thermostat. *Phys. Rev. Lett.* **2009**, *103*, 030603.

(17) Ceriotti, M.; Bussi, G.; Parrinello, M. Colored-noise thermostats à la carte. *J. Chem. Theory Comput.* **2010**, *6*, 1170−1180.

(18) 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. *Chem. Rev.* **2016**, *116*, 7529−7550.
(19) Ceriotti, M.; Manolopoulos, D. E.; Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *J. Chem. Phys.* **2011**, *134*, 084104.
(20) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-pi: A python interface for ab initio path integral molecular dynamics simulations. *Comput. Phys. Commun.* **2014**, *185*, 1019−1026.
(21) Chandler, D.; Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.* **1981**, *74*, 4078−4095.
(22) Cheng, B.; Engel, E. A.; Behler, J.; Dellago, C.; Ceriotti, M. Ab initio thermodynamics of liquid and solid water. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 1110−1115.
(23) Cheng, X.; Herr, J. D.; Steele, R. P. Accelerating ab initio path integral simulations via imaginary multiple-timestepping. *J. Chem. Theory Comput.* **2016**, *12*, 1627−1638.
(24) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling molecular interactions in water: From pairwise to many-body potential energy functions. *Chem. Rev.* **2016**, *116*, 7501−7528.
(25) Craig, I. R.; Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* **2004**, *121*, 3368−3373.
(26) Dammak, H.; Chalopin, Y.; Laroche, M.; Hayoun, M.; Greffet, J.-J. Quantum thermal bath for molecular dynamics simulation. *Phys. Rev. Lett.* **2009**, *103*, 190601.
(27) Fanourgakis, G. S.; Xantheas, S. S. Development of transferable interaction potentials for water. v. extension of the flexible, polarizable, thole-type model potential (ttm3-f, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *J. Chem. Phys.* **2008**, *128*, 074506.
(28) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(29) Feynman, R. P.; Hibbs, A. R.; Styer, D. F. *Quantum Mechanics and Path Integrals*; Courier Corporation, 2010.
(30) Guillot, B.; Guissani, Y. Quantum effects in simulated water by the feynman−hibbs approach. *J. Chem. Phys.* **1998**, *108*, 10162−10174.
(31) Habershon, S.; Manolopoulos, D. E. Zero point energy leakage in condensed phase dynamics: An assessment of quantum simulation methods for liquid water. *J. Chem. Phys.* **2009**, *131*, 244518.
(32) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **2009**, *131*, 024501.
(33) Hele, T. J.; Willatt, M. J.; Muolo, A.; Althorpe, S. C. Boltzmann-conserving classical dynamics in quantum time-correlation functions: "Matsubara dynamics". *J. Chem. Phys.* **2015**, *142*, 134103.
(34) Hernández-Rojas, J.; Calvo, F.; Noya, E. G. Applicability of quantum thermal baths to complex many-body systems with various degrees of anharmonicity. *J. Chem. Theory Comput.* **2015**, *11*, 861−870.
(35) Hijazi, M.; Wilkins, D. M.; Ceriotti, M. Fast-forward Langevin dynamics with momentum flips. *J. Chem. Phys.* **2018**, *148*, 184109.
(36) Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. A high-quality x-ray scattering experiment on liquid water at ambient conditions. *J. Chem. Phys.* **2000**, *113*, 9140−9148.
(37) Kapil, V.; Behler, J.; Ceriotti, M. High order path integrals made easy. *J. Chem. Phys.* **2016**, *145*, 234103.
(38) Kapil, V.; VandeVondele, J.; Ceriotti, M. Accurate molecular dynamics and nuclear quantum effects at low cost by multiple steps in real and imaginary time: Using density functional theory to accelerate wavefunction methods. *J. Chem. Phys.* **2016**, *144*, 054111.
(39) Kubo, R. The fluctuation−dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255.
(40) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956−972.
(41) Li, X.-Z.; Walker, B.; Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 6369−6373.
(42) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ classical potential for modeling molecular interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122−4139.
(43) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of geometry-dependent charge flux into the polarizable AMOEBA+ potential. *J. Phys. Chem. Lett.* **2020**, *11*, 419−426.
(44) 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. *J. Chem. Theory Comput.* **2019**, *15*, 2863−2880.
(45) Markland, T. E.; Manolopoulos, D. E. An efficient ring polymer contraction scheme for imaginary time path integral simulations. *J. Chem. Phys.* **2008**, *129*, 024105.
(46) Marsalek, O.; Markland, T. E. Ab initio molecular dynamics with nuclear quantum effects at classical cost: Ring polymer contraction for density functional theory. *J. Chem. Phys.* **2016**, *144*, 054112.
(47) Melcr, J.; Piquemal, J.-P. Accurate biomolecular simulations account for electronic polarization. *Front. Mol. Biosci.* **2019**, *6*, 143.
(48) Miller, W. H. The semiclassical initial value representation: A potentially practical way for adding quantum effects to classical molecular dynamics simulations. *J. Phys. Chem. A* **2001**, *105*, 2942−2955.
(49) Miura, S.; Tuckerman, M. E.; Klein, M. L. An ab initio path integral molecular dynamics study of double proton transfer in the formic acid dimer. *J. Chem. Phys.* **1998**, *109*, 5290−5299.
(50) Monacelli, L.; Errea, I.; Calandra, M.; Mauri, F. Black metal hydrogen above 360 GPa driven by proton quantum fluctuations. *Nat. Phys.* **2021**, *17*, 63−67.
(51) Morawietz, T.; Behler, J. A density-functional theory-based neural network potential for water clusters including van der waals corrections. *J. Phys. Chem. A* **2013**, *117*, 7356−7366.
(52) Morrone, J. A.; Car, R. Nuclear quantum effects in water. *Phys. Rev. Lett.* **2008**, *101*, 017801.
(53) Onufriev, A. V.; Izadi, S. Water models for biomolecular simulations. *WIREs Computational Molecular Science* **2018**, *8*, e1347.
(54) Paesani, F.; Voth, G. A. The properties of water: Insights from quantum simulations. *J. Phys. Chem. B* **2009**, *113*, 5702−5719.
(55) Paesani, F.; Yoo, S.; Bakker, H. J.; Xantheas, S. S. Nuclear quantum effects in the reorientation of water. *J. Phys. Chem. Lett.* **2010**, *1*, 2316−2321.
(56) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R. D.; Fain, B. On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8878−8882.
(57) Pérez, A.; Tuckerman, M. E. Improving the convergence of closed and open path integral molecular dynamics via higher order trotter factorization schemes. *J. Chem. Phys.* **2011**, *135*, 064104.
(58) Pérez, A.; Tuckerman, M. E.; Hjalmarson, H. P.; von Lilienfeld, O. A. Enol tautomers of watson- crick base pair models are metastable because of nuclear quantum effects. *J. Am. Chem. Soc.* **2010**, *132*, 11510−11515.
(59) Plé, T.; Huppert, S.; Finocchi, F.; Depondt, P.; Bonella, S. Sampling the thermal wigner density via a generalized Langevin dynamics. *J. Chem. Phys.* **2019**, *151*, 114114.
(60) Poltavsky, I.; Kapil, V.; Ceriotti, M.; Kim, K. S.; Tkatchenko, A. Accurate description of nuclear quantum effects with high-order perturbed path integrals (hoppi). *J. Chem. Theory Comput.* **2020**, *16*, 1128−1135.
(61) Poltavsky, I.; Tkatchenko, A. Modeling quantum nuclei with perturbed path integral molecular dynamics. *Chem. Sci.* **2016**, *7*, 1368−1372.

https://doi.org/10.1021/acs.jpclett.1c01722
J. Phys. Chem. Lett. 2021, 12, 8285−8291

(62) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the mb-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.
(63) Reiter, G.; Li, J. C.; Mayers, J.; Abdul-Redah, T.; Platzman, P. The proton momentum distribution in water and ice. *Braz. J. Phys.* **2004**, *34*, 142−147.
(64) Romanelli, G.; Ceriotti, M.; Manolopoulos, D. E.; Pantalei, C.; Senesi, R.; Andreani, C. Direct measurement of competing quantum effects on the kinetic energy of heavy water upon melting. *J. Phys. Chem. Lett.* **2013**, *4*, 3251−3256.
(65) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* **2014**, *140*, 234116.
(66) Rossi, M.; Gasparotto, P.; Ceriotti, M. Anharmonic and quantum fluctuations in molecular crystals: A first-principles study of the stability of paracetamol. *Phys. Rev. Lett.* **2016**, *117*, 115702.
(67) Rossi, M.; Kapil, V.; Ceriotti, M. Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation. *J. Chem. Phys.* **2018**, *148*, 102301.
(68) Singraber, A.; Behler, J.; Dellago, C. Library-based lammps implementation of high-dimensional neural network potentials. *J. Chem. Theory Comput.* **2019**, *15*, 1827−1840.
(69) Smith, J. S.; Isayev, O.; Roitberg, A. E. Ani-1: an extensible neural network potential with dft accuracy at force field computational cost. *Chem. Sci.* **2017**, *8*, 3192−3203.
(70) Trenins, G.; Willatt, M. J.; Althorpe, S. C. Path-integral dynamics of water using curvilinear centroids. *J. Chem. Phys.* **2019**, *151*, 054109.
(71) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18454−18459.
(72) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.

https://doi.org/10.1021/acs.jpclett.1c01722
J. Phys. Chem. Lett. 2021, 12, 8285−8291

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Improving Condensed-Phase Water Dynamics with Explicit Nuclear Quantum Effects: The Polarizable Q-AMOEBA Force Field

Published as part of The Journal of Physical Chemistry virtual special issue "Biomolecular Electrostatic Phenomena".

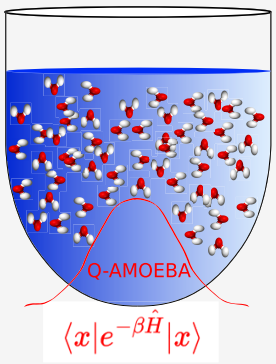
Nastasia Mauger, Thomas Plé, Louis Lagardère,* Simon Huppert,* and Jean-Philip Piquemal*

Cite This: *J. Phys. Chem. B* 2022, 126, 8813−8826

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We introduce a new parametrization of the AMOEBA polarizable force field for water denoted Q-AMOEBA, for use in simulations that explicitly account for nuclear quantum effects (NQEs). This study is made possible thanks to the recently introduced adaptive Quantum Thermal Bath (adQTB) simulation technique which computational cost is comparable to classical molecular dynamics. The flexible Q-AMOEBA model conserves the initial AMOEBA functional form, with an intermolecular potential including an atomic multipole description of electrostatic interactions (up to quadrupole), a polarization contribution based on the Thole interaction model and a buffered 14−7 potential to model van der Waals interactions. It has been obtained by using a ForceBalance fitting strategy including high-level quantum chemistry reference energies and selected condensed-phase properties targets. The final Q-AMOEBA model is shown to accurately reproduce both gas-phase and condensed-phase properties, notably improving the original AMOEBA water model. This development allows the fine study of NQEs on water liquid phase properties such as the average H−O−H angle compared to its gas-phase equilibrium value, isotope effects, and so on. Q-AMOEBA also provides improved infrared spectroscopy prediction capabilities compared to AMOEBA03. Overall, we show that the impact of NQEs depends on the underlying model functional form and on the associated strength of hydrogen bonds. Since adQTB simulations can be performed at near classical computational cost using the Tinker-HP package, Q-AMOEBA can be extended to organic molecules, proteins, and nucleic acids opening the possibility for the large-scale study of the importance of NQEs in biophysics.



INTRODUCTION

Classical molecular dynamics (MD) has become a powerful valuable tool to study the properties of complex systems, such as condensed matter, organic molecules, and proteins, but also to design new molecules and drugs.[1−6] It provides access to most of the relevant thermodynamic observables of a given system. The quality of these observables relies not only on the accuracy of the potential energy surface (PES), usually called a force field (FF), but also on the amount of the phase-space sampling. Thanks to the different advances in computing hardware, such as GPU computing,[7−10] and techniques such as enhanced sampling,[11,12] MD is now capable of reaching time scales of milliseconds making it possible the extensive study of macromolecules.[13]

Most standard additive FFs, such as AMBER,[14,15] CHARMM,[16−19] OPLS,[20−22] and COMPASS,[23] have been widely used in MD thanks to their low computational cost due to their relatively simple functional form and they have been refined throughout the years. However, they lack a proper description of many-body interactions, and although they provide good results for various properties compared to experiments, it appears that their accuracy is limited in other contexts.[24] The major limitation of these FFs is the use of fixed partial atomic charges to model electrostatic interactions and their lack of a fine description of many-body polarization which hinders their transferability when the atomic environment changes. To circumvent that, much effort has been made to include explicitly many-body polarization in FFs by using either Drude oscillators, fluctuating charges, or induced dipoles.[24−26] Versions of CHARMM[27,28] and AMBER[29] have been published where polarization was added to the existing nonpolarizable FF and new ones have been developed such as PFF.[30] Moreover, it has been shown that FFs should include the polarization from scratch.[31] Hence, more modern FFs such as AMOEBA,[32,33] AMOEBA+,[34,35] or SIBFA[36,37] have been designed with polarization included on the outset.

Received: June 27, 2022
Revised: October 5, 2022
Published: October 21, 2022

Due to its major role in many fields, water has been a central focus of FF development. Over the past decades, many FFs have been developed, such as TIP3P,[38] SPC,[39] and their variants.[40,41] Although widely used, these models struggle to describe a wide part of the phase diagram of water and ice. For example, the gas-phase binding of the water dimer is overestimated by approximately 30% in the TIP5P model.[41] Based on this, other water models have been developed such as BK3,[42] TIP4P/Ew,[43] and TIP4P/2005,[44] which use implicit many-body effects through classical polarization to go beyond the pairwise approximation. However, these models were parametrized for classical simulations and do not take into account explicitly Nuclear Quantum Effects (NQEs) which reduce their transferability. Indeed, such effects are important for systems involving light atoms such as hydrogen,[45] and some studies have highlighted their importance in biological systems.[46,47] In practice, such effects are usually implicitly added by fitting the analytical PES to recover thermodynamic observables. Moreover, many recent water models use high-quality *ab initio* data as a basis for the PES, and it has been shown that with such a high precision, NQEs must be taken into account explicitly to accurately reproduce thermodynamic observables.[48]

A rigorous and asymptotically exact approach to include NQEs explicitly in MD simulations is provided by the path integral molecular dynamics (PIMD) formalism where the dynamics is performed on an extended classical system consisting in several replicas (also called beads) of the physical system.[49,50] This method allows for a systematically improvable treatment of NQEs but at a computational cost that is considerably larger than that of classical MD. This increased cost has so far limited the development of FF models based on PIMD estimations of thermodynamic properties, with a few notable exceptions such as TIP4PQ/2005,[51] q-SPC/Fw,[52] q-TIP4P/F,[53] and more recently ArrowFF.[48] In parallel, different methods have been developed to reduce the computational cost of PIMD simulations, such as high-order PI,[54,55] PI perturbation theory,[56,57] or path integral generalized Langevin methods.[58,59] Though these methods have enabled significant progress, PIMD simulations remain significantly costlier than classical MD (typically requiring 1 order of magnitude more resources at room temperature). Other approaches, such as ring-polymer contractions[60,61] or imaginary multiple time-stepping,[62] allow reducing the number of replicas used in the evaluation of long-range interaction terms, thereby achieving almost classical computational cost in some applications.[63] However, the splitting of the PES between long-range and short-range terms is not always available, in particular in the context of *ab initio* MD or when using machine learning potentials. In this paper, we introduce a new polarizable force field, Q-AMOEBA, based on the AMOEBA functional form that is designed to be used with the explicit inclusion of NQEs. Indeed, since experiments naturally contain "quantumness", the initial AMOEBA model intended to implicitly include quantum nuclear effects. When adding explicitly NQEs, one needs to reparameterize the force field potential to propose a version of it that only tends to reproduce the Born–Oppenheimer energy in order to avoid double counting when NQEs are explicitly included through the dynamics. To do so, we make extensive use of the quantum thermal bath (QTB) a method based on a generalized Langevin thermostat to approximate the zero-point motion of the nuclei. More precisely, we use the adaptive QTB method (adQTB) that relies on the quantum fluctuation–dissipation theorem to systematically compensate the effects of zero-point energy leakage (ZPEL).[64] This approach was recently shown to reliably approximate NQEs in liquid water at a computational cost that remains similar to that of classical MD.[65] The Q-AMOEBA force field parameters are optimized with the ForceBalance software (FB)[66] in combination with the Tinker-HP[10,67] package for molecular dynamics simulations, in order to accurately reproduce energies of various water systems in the gas phase as well as a few condensed-phase properties obtained with a quantum description of the nuclei. This model will then make the extensive description of more complex systems involving water with explicit NQEs possible. Q-AMOEBA gives overall better results on energies and on key observables such as the isobaric heat capacity or the thermal expansion coefficient than the original AMOEBA model. Such a model also allows us to finely study the practical effects of NQEs by comparing the properties that are obtained with it but with purely classical dynamics. We first describe our parametrization strategy in detail and then test the accuracy and transferability of our model on various thermodynamic and structural properties as well as on the IR spectra, showing the validity of our approach and its future applicability to the development of new generation models.

METHODS

AMOEBA Model. The total potential energy of the AMOEBA[68] water model can be expressed as the sum of bonded and nonbonded energy terms:

$$\begin{aligned} E_{\text{total}} &= E_{\text{bonded}} + E_{\text{nonbonded}} \\ E_{\text{bonded}} &= E_{\text{bond}} + E_{\text{angle}} + E_{\text{b}\theta} \\ E_{\text{nonbonded}} &= E_{\text{vdW}} + E_{\text{ele}}^{\text{perm}} + E_{\text{ele}}^{\text{ind}} \end{aligned} \tag{1}$$

The bonded terms which include the anharmonic bond-stretching and angle-bending terms are similar to the MM3 force field.[69] The water intramolecular geometry and vibrations are described with an Urey–Bradley terms $E_{\text{b}\theta}$. In the original AMOEBA model,[32] the ideal bond length was chosen to be at the experimental value of 0.9572 Å. The ideal bending angle was 108.5°, and the Urey–Bradley ideal distance was set to 1.5326 Å. The angle is larger than the experimental gas-phase angle of 104.52° which was shown to be necessary to reproduce the correct average experimental angle in liquid water.[68]

The nonbonded terms are composed of the vdW interactions and the electrostatic contributions from both permanent and induced dipoles. The vdW functional term uses Halgren's buffered 14–7 potential to model the pairwise additive interactions for dispersion at long-range and exchange-repulsion at short-range:[70]

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

where ε_{ij} is the potential well depth and $\sigma_{ij} = r_{ij}/r_{ij}^{0}$, where r_{ij} is the i–j separation and r_{ij}^{0} is the minimum energy distance (distances are calculated from the oxygen atoms, see ref 32). In AMOEBA, the vdW parameters were set to $\gamma = 0.12$ and $\delta = 0.07$. The combining rules used for r_{ij}^{0} and ε_{ij} are given by

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813–8826

$$r_{ij}^0 = \frac{(r_{ii}^0)^3 + (r_{jj}^0)^3}{(r_{ii}^0)^2 + (r_{jj}^0)^2}$$
$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}})^2} \quad (3)$$

A hydrogen reduction factor is added which moves the hydrogen vdW center toward the oxygen along the O−H bond. To compute the electrostatic interactions the AMOEBA model uses point atomic multipoles truncated at quadrupoles for each atom center. The atomic multipole moments are derived using the distributed multipole analysis (DMA) approach and then optimized against a high-level *ab initio* PES.[71] More details about the functional form of AMOEBA can be found in reference.[33]

Parametrization of Q-AMOEBA. The Q-AMOEBA water model has been parametrized toward liquid phase simulations using the ForceBalance (FB) software.[72−74] FB calculates the derivatives of the target properties with respect to the FF parameters to be optimized (see the "Derivatives of Average Properties with Respect to Parameters in the Path Integrals Formalism" section in the Technical Appendix for more details) and uses a Newton−Raphson procedure to reach a minimum with respect to a given objective function. In this work, the target properties include both *ab initio* configurational energies and experimental observables. The initial parameters were taken from the original AMOEBA water model[32] (denoted as AMOEBA03 in the rest of the text) and the weights of the different data included in the objective function are summarized in Table 1. The different convergence criteria used for FB are given in the Supporting Information.

Table 1. Reference Data Used in FB to Derive the Q-AMOEBA Sets of Parameters[a]

system	reference data	data type	data point	weight	
clusters	gas-phase dipole−quadrupole	CCSD(T)		1.0	
	gas-phase vibrational modes	CCSD(T)		1.0	
	Smith dimer	CCSD(T) BE	10	1.0	
	small gas-phase cluster	CCSD(T) BE	21	1.0	
	large gas-phase clusters	MP2 BE	18	1.0	
	PE and AF	MP2	42.000	1.0	
liquid	ρ	expt.	10	1.0	0.6
	ΔH_{vap}	expt.	10		0.4

[a]BE, binding energies; PE, potential energies; and AF, atomistic forces. The table also shows the weights that were applied to the different target properties.

Most parameters of AMOEBA03 (such as the atomic permanent multipolar moments) are obtained directly from *ab initio* calculations; therefore, only van der Waals parameters and some well-chosen intramolecular terms were modified. More specifically, the optimization was performed on both van der Waals radii and epsilon values of the Halgren 14−7 potential[70] associated with oxygen and hydrogen atoms, as well as the buffer radius of the hydrogen atom and equilibrium distances and angles of the intramolecular stretching and bending terms. Our initial intent was to develop a single set of parameters to be used with both PIMD and adQTB approaches. In practice, we observed that separate FB optimizations with both methods yielded slightly different results. In particular, this is due to the particularities of pressure estimation in adQTB simulations (see the "Pressure Estimator" section in the Appendix), which leads to small inaccuracies in the determination of the density with this method (of the order of 1%, as was already noted in a former study on a different FF model).[65] Considering the importance of this observable, we resorted to two sets of parameters with specific optimization for each method: one for PIMD denoted as (1) and a second for adQTB, denoted as (2). Because the adQTB density appeared systematically slightly underestimated, we decided to further modify the buffer radius of H and the epsilon of O as they displayed the most significant discrepancies compared to the PIMD ones. To do so, we used the PIMD values as a guess and further used FB to finally obtain a satisfactory density with adQTB. This strategy ultimately led to the final adQTB parameters discussed in the rest of the paper. We emphasize that the two final sets of Q-AMOEBA parameters differ only slightly from each other. To perform the simulations of the condensed-phase properties with FB, we used the Tinker-HP software on GPUs for PIMD and adQTB methods.

Geometries extracted from condensed-phase simulations while using the original AMOEBA03 model at temperatures ranging from 249.15 to 373.15 K and including cluster sizes from 2 to 22 molecules were used to obtain *ab initio* references. The calculations were performed using Q-Chem 4.0 and include both energies and gradients calculated using the aug-cc-pVTZ basis set. Optimal geometries and binding energies of 40 small water clusters ranging from 2 to 20 molecules were also used. Each cluster has been used on the highest level of theory available; see Table 1 for details.

The objective function includes two thermodynamic properties with experimental reference: the density ρ and the enthalpy of vaporization ΔH_{vap}. These properties were evaluated at different temperatures ranging from 249.15 to 369.15 K under 1 atm. All simulations were carried out for N = 4000 molecules in periodic boundary conditions with a cubic box. The van der Waals (vdW) cutoff was set to 12 Å, and electrostatic interactions were computed with the SPME method[75] with a real space cutoff of 7 Å and a 60 × 60 × 60 grid. A RESPA integrator using a bonded/nonbonded split with time steps of 0.2 and 2 fs respectively was used. For each FB iteration, the dynamics include 1 ns of equilibration and 4 ns of production. For the PIMD calculations, 32 beads were used with the Thermostated Ring Polymer Molecular Dynamics (TRPMD) algorithm[76] combined with a mild Langevin thermostat on the centroid (friction coefficient γ = 1 ps^{-1}), whereas in the adQTB method the friction was set to 20 ps^{-1}.

Because of the high frequencies of the intramolecular stretching and bending modes, the associated zero point energies are large, impacting the average values of the O−H distance and H−O−H angle. We therefore modified the corresponding parameters in Q-AMOEBA to recover experimental values. The final parameters of our models can be found in the Supporting Information.

RESULTS AND DISCUSSION

Binding Energy of Water Clusters. Tables 2 and 3 show the comparison of the configuration energies obtained with

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

Table 2. Binding Energies of the 10 Smith Dimers Using Q-AMOEBA Compared to the AMOEBA03 Model and *Ab Initio* References[a]

$(H_2O)_2$	CCSD(T)	AMOEBA03	Q-AMOEBA (PIMD)	Q-AMOEBA (adQTB)
Smith01	−4.97	−4.58	−4.96	−4.95
Smith02	−4.45	−3.98	−4.35	−4.34
Smith03	−4.42	−3.94	−4.3	−4.3
Smith04	−4.25	−3.54	−3.49	−3.43
Smith05	−4.00	−2.69	−3.06	−3.00
Smith06	−3.96	−2.59	−2.95	−2.90
Smith07	−3.26	−2.55	−2.81	−2.73
Smith08	−1.30	−0.8	−1.04	−0.95
Smith09	−3.05	−2.69	−2.97	−2.90
Smith10	−2.18	−1.89	−2.14	−2.07
RMSE		0.81	0.53	0.57

[a]The table distinguishes between Q-AMOEBA (PIMD) and Q-AMOEBA (adQTB) depending on which method was used to include NQEs in the calibration of the model. CCSD(T) results come from ref 80.

AMOEBA and Q-AMOEBA compared to accurate quantum references for water clusters of different sizes. For numbers of atoms $n < 6$, the CCSD(T) level of theory has been used whereas larger clusters, such as octamers,[77] 16−17-mers[78] and 20-mers[79] were investigated at the MP2 level. These studies provide the optimized structures and binding energies (BEs) of these clusters that were used as a reference for the Q-AMOEBA parametrization. For a given model, the binding energy is defined as the difference between the optimized cluster energy and the sum of the energies of the optimized monomers. Over all the configurations explored, both Q-AMOEBA models give very similar and accurate results. In particular, Q-AMOEBA consistently yields more accurate BE for the Smith dimers than AMOEBA03 with root mean squared errors (RMSE) of 0.53 and 0.57 kcal·mol^{-1} compared to 0.81 kcal·mol^{-1} for AMOEBA03. For water clusters of intermediate sizes ($3 \leq n \leq 8$), Q-AMOEBA slightly overestimates BEs while remaining within a 0.5 kcal·mol^{-1} per monomer range. The accuracy of Q-AMOEBA improves for the largest clusters, yielding an overall smaller RMSE than AMOEBA03 (overall improvement of approximately 0.9 and 0.2 kcal·mol^{-1}). This agreement should translate into a better representation of condensed-phase properties.

Structural Properties of Water. Radial distribution functions (RDFs) are a good indicator of the local molecular structure in the liquid phase both in terms of the different peak positions and their width. They were computed from *NVT* simulations at 298.15 K and the corresponding equilibrium volume. In Figure 1, the PIMD (dash red curve) and adQTB

Table 3. Binding Energies of Trimer to 20-mer of Water Clusters Computed Using Q-AMOEBA Compared to AMOEBA03 and *Ab Initio*[a]

$(H_2O)_n$	geometry	QM	AMOEBA03	Q-AMOEBA (PIMD)	Q-AMOEBA (adQTB)
$n = 3$	cyclic	−15.74	−15.03	−16.05	−16.11
$n = 4$	cyclic	−27.40	−27.63	−28.98	−29.32
$n = 5$	cyclic	−35.93	−36.38	−38.19	−38.72
$n = 6$	prism	−45.92	−45.71	−47.88	−48.36
	cage	−45.67	−45.82	−47.95	−48.39
	bag	−44.30	−44.79	−46.97	−47.53
	cyclic chair	−44.12	−44.62	−46.88	−47.60
	book1	−45.20	−45.6	−47.8	−48.39
	book2	−44.90	−45.37	−47.55	−48.11
	cyclic boat1	−43.13	−43.78	−45.79	−46.65
	cyclic boat2	−43.07	−43.84	−46.04	−46.73
$n = 8$	S4	−72.70	−72.34	−75.71	−76.46
	D2d	−72.70	−72.39	−75.75	−76.50
$n = 11$	434	−105.72	−101.74	−106.40	−107.52
	515	−105.18	−102.08	−106.85	−108.11
	551	−104.92	−101.85	−106.61	−107.86
	443	−104.76	−101.86	−106.63	−107.78
	4412	−103.97	−101.41	−106.26	−107.50
$n = 16$	boat-a	−170.80	−161.21	−168.83	−170.57
	boat-b	−170.63	−161.54	−169.12	−170.92
	antiboat	−170.54	−162.26	−169.98	−171.95
	ABAB	−171.05	−161.18	−168.69	−170.32
	AABB	−170.51	−160.98	−168.51	−170.13
$n = 17$	sphere	−182.54	−173.02	−181.21	−183.32
	5525	−181.83	−172.32	−180.42	−182.46
$n = 20$	dodecahedron	−200.10	−197.03	−206.1	−208.54
	fused cubes	−212.10	−205.76	−215.35	−217.42
	face sharing prisms	−215.20	−206.13	−215.74	−218.05
	edge sharing prisms	−218.10	−208.46	−218.40	−220.85
RMSE			3.33	2.36	3.12

[a]The table distinguishes between Q-AMOEBA (PIMD) and Q-AMOEBA (adQTB) depending on which method was used to include NQEs in the calibration of the model. QM reference correspond to CCSD(T) for $n \leq 6$[32,81] and to MP2 for $n \geq 8$.[77−79].

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

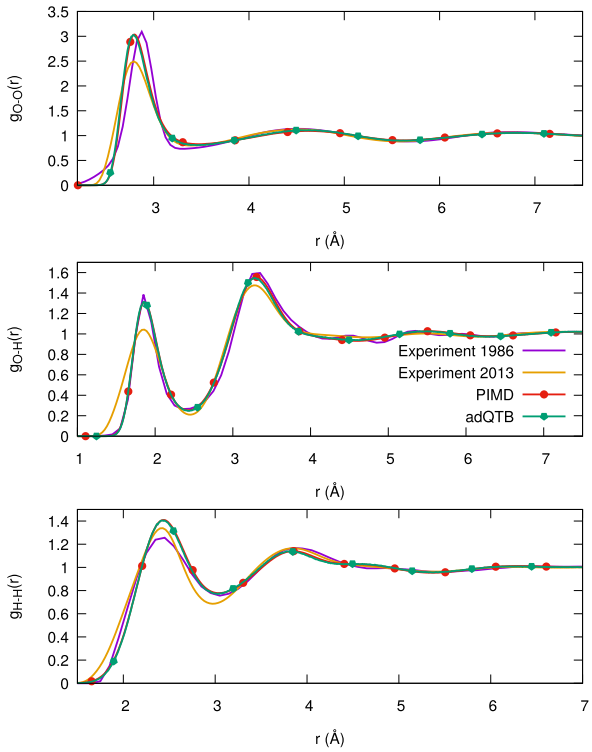


Figure 1. RDFs computed at 298.15 K with Q-AMOEBA compared to the X-ray[82] and neutron[83] diffraction experiments.

(dotted and dashed green curve) radial distributions are compared to the experimental data obtained from X-ray scattering[82] and neutron diffraction.[83] These data were not included in our parametrization and hence were used as a first test.

Both Q-AMOEBA results are almost indistinguishable and overall in close agreement with experiments. The difference between simulation and experimental curves are of the same order as the discrepancies within experimental data. For instance, the first peak of the Oxygen−Oxygen RDF is located at 2.73 Å in the X-ray scattering experiment whereas it is at 2.80 Å in the neutron diffraction one. It is well reproduced by the two Q-AMOEBA models, with peaks located respectively at 2.80 Å (PIMD) and 2.79 Å (adQTB).

Intramolecular structure is illustrated by the average O−H distance and H−O−H angle as shown in Table 4. NQEs tend

Table 4. Average O−H Bond Length and H−O−H Angle while Using the Q-AMOEBA Parameters Compared to the Experimental Results[83]

	PIMD (32 beads)	PIMD (64 beads)	adQTB	expt.
r_{OH} (Å)	0.985	0.985	0.975	0.97
θ_{HOH} (deg)	106.08	106.10	104.95	105.1

to slightly increase the average O−H distance and H−O−H angle of water molecules in the liquid phase. This is why both sets of parameters have a reduced ideal angle compared to the initial set of parameters (106.2° in adQTB and 107.2° in PIMD compared to 108.2° in AMOEBA03). However, the PIMD values of Table 4 are a little overestimated, whereas this effect was compensated in our adQTB model by decreasing further the associated equilibrium parameters (from 0.9572 to 0.9472 Å for the O−H distance and 108.2° to 106.2° for the H−O−H angle). Interestingly, the modeling of the average value of the H−O−H angle in liquid phase compared to its average value in gas phase is known to be problematic with classical force fields and requires to artificially increase the equilibrium parameter.[32] It has been shown that this can be traced back to the absence of charge-flux effect.[35] The former effect of NQEs on these averages highlights the need to explicitly include them to design accurate models of water.

Infrared Spectra. The infrared absorption spectrum can be related to the total dipole derivative autocorrelation function as[85,86]

$$n(\omega)\,\alpha(\omega) = \frac{\pi\beta}{3V\epsilon_0} C_{\dot{\mu}\dot{\mu}}(\omega) \tag{4}$$

where $n(\omega)$ is the refractive index, $\alpha(\omega)$ is the Beer−Lambert absorption coefficient, $\beta = 1/k_B T$ is the inverse thermal energy, V is the system's volume, and $C_{\dot{\mu}\dot{\mu}}(\omega)$ is the Kubo autocorrelation spectrum of the total dipole time derivative $\dot{\mu}$. In the AMOEBA framework, the total dipole moment is estimated as

$$\boldsymbol{\mu} = \sum_{i=1}^{N_{at}} (q_i \mathbf{r}_i + \boldsymbol{\mu}_i^0 + \boldsymbol{\mu}_i^{ind}) \tag{5}$$

where q_i are permanent charges located on the atom's position $\mathbf{r}_i$, $\boldsymbol{\mu}_i^0$ are permanent dipoles (which magnitudes are fixed but rotate with the water molecules), and $\boldsymbol{\mu}_i^{ind}$ are induced dipoles that are obtained at each step of the dynamics via a minimization procedure.[67] Since no analytical form for the time derivative of the induced dipole is available, it is estimated using finite differences of the total dipole moment over the trajectory.

The IR spectrum is directly evaluated in path integrals simulations in the framework of TRPMD,[76] while it is recovered from adQTB simulations through the following relation in the Fourier domain:[87]

$$C_{\dot{\mu}\dot{\mu}}(\omega) \approx \frac{\tanh(\beta\hbar\omega/2)}{\beta\hbar\omega/2} C_{\dot{\mu}\dot{\mu}}^{adQTB}(\omega) \tag{6}$$

In practice, it is directly estimated from the Fourier transform of the dipole derivative trajectory (averaged over the beads in TRPMD simulations) according to the Wiener−Khinchin theorem. Furthermore, in the adQTB method, the relatively high friction coefficient (that is required to compensate ZPEL) tends to broaden spectral lineshapes so that it is necessary to use the deconvolution procedure of ref 88 to improve the spectrum.

Figure 2 shows the IR spectra calculated using TRPMD and adQTB compared to the experimental spectrum (extracted from ref 53), classical AMOEBA03 spectrum and the classical spectrum obtained from a simulation using the parameters optimized for adQTB. The spectra from both quantum simulations are very similar in their peak positions and are in good agreement with the experimental data. As already well-documented,[45,65,86] including NQEs shifts the intramolecular peaks (bending around 1600 cm^{-1} and stretching around 3500 cm^{-1}) toward lower frequencies. Relative intensities between the spectral features are similar in quantum and classical simulations and agree well with experimental results (although, we can note that the TRPMD spectrum for the stretching peak is broadened compared to the other simulations, which is a well-known discrepancy of this method). The low-frequency

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

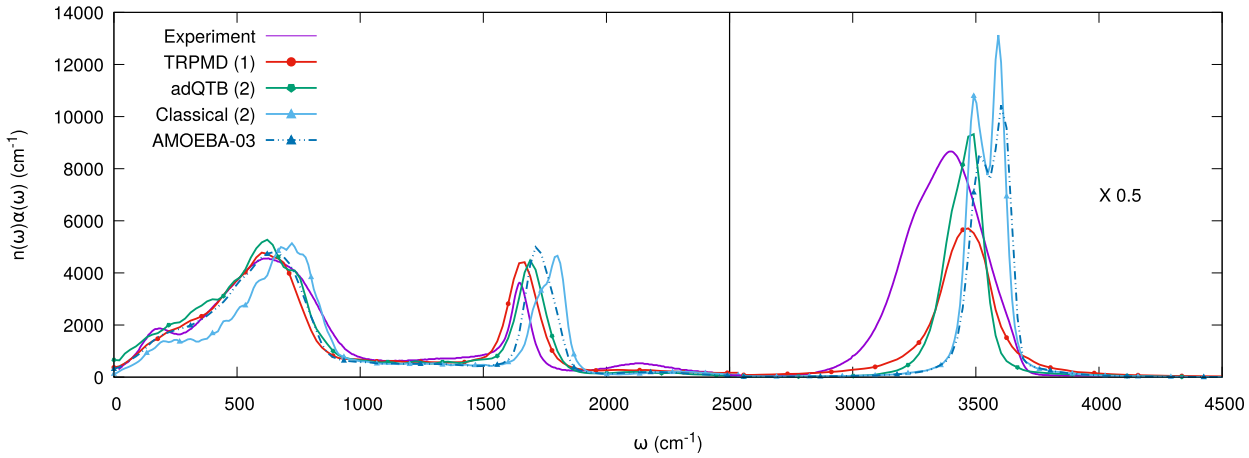


Figure 2. IR absorption spectra computed at 300 K and ρ = 0.997 g·cm^{-3} with the Q-AMOEBA water model the adQTB and PIMD methods with their respective set of parameters (green and red curves, respectively). The *classical* curve (light blue) is obtained from classical MD simulations with the Q-AMOEBA (adQTB) model. They are compared with AMOEBA03 (classical MD) and experimental data.[84] The right part of the plot, corresponding to the stretching mode is multiplied by 0.5. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text).

features are very similar for AMOEBA03, TRPMD and adQTB (although slightly more intense for the adQTB) and are red-shifted compared to the results obtained in classical simulations with Q-AMOEBA (light blue), showing the impact of NQEs even at those low frequencies. The general shape of these low-frequency features are in good agreement with the experimental spectrum and even display some substructure at around 200 cm^{-1} that is due to slow induced-dipole dynamics[53,89] and is absent for nonpolarizable models, even when including NQEs.[53]

Thermodynamic Properties. Thermodynamic properties of the two Q-AMOEBA set of parameters were computed from temperatures ranging from 249.15 to 369.15 K under 1 atm. As shown in Figure 3, the characteristic bell shape of the density versus temperature curve is well captured by the Q-AMOEBA force fields. The deviations from the experimental result are minor, with average densities at ambient temperature of 0.998 and 0.999 g·cm^{-3} for the PIMD and adQTB methods respectively, compared to the experimental value of 0.997 g·cm^{-3} (less than 0.2% difference). In both high and low ranges of temperature, the density is slightly underestimated by a maximum of 0.85% at 369.15K. The figure also shows the curve obtained from a classical MD simulation using the Q-AMOEBA (adQTB) set of parameters (light blue curve). It differs significantly from the AMOEBA03 classical reference, which shows that the changes made to the force field parameters have a sizable impact on the density. Even more importantly, the classical MD curve displays strong discrepancies with the corresponding adQTB and PIMD results: The temperature of maximum density T_{MD} is shifted to higher values, while the density is significantly reduced at low temperature and increased at high temperatures with respect to the results obtained including NQEs. To explain this impact of NQEs, we come back to the peculiarities of liquid water that explain the unusual bell shape of the density curve. The latter is related to the tetrahedral arrangement of the water molecules, which tends to produce a loosely packed local structure. For $T < T_{MD}$, increasing the temperature allows hydrogen bonds to strain and break more easily, which in turns enables larger deviations from the tetrahedral order with a better packing efficiency and therefore an increase of the density. However, for $T > T_{MD}$, a competitive process tends to dominate: The breaking of H-bonds tends to increase nearest-neighbors distances which leads to the more usual thermal expansion observed at high temperature.[91] Within the Q-AMOEBA model, we observe that the inclusion of NQEs tends to increase the density for $T < T_{MD}$ and to decrease it above T_{MD}, which indicates that NQEs lead to an overall weakening of the hydrogen bonds over the whole temperature range. This is consistent with experimental observations of isotope effects in liquid water, most of which suggest that hydrogen bonds are weaker in H_2O than in D_2O,[45] the latter isotope being heavier hence *more classical*. In particular, the experimental T_{MD} shifts from 277.15 K for H_2O to 284.32 K for D_2O and even 286.55 K for T_2O, which agrees qualitatively with the observed shift between the quantum and classical MD curves within the Q-AMOEBA model.

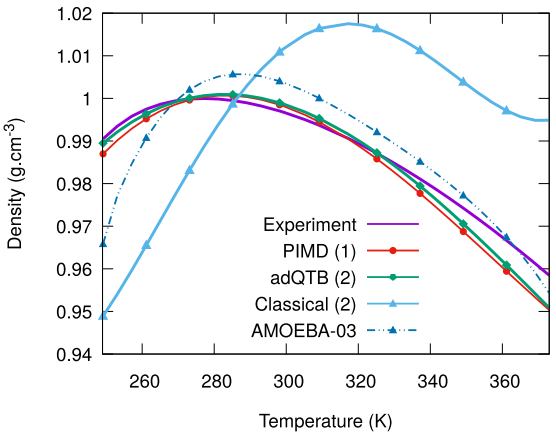


Figure 3. Density of liquid water at P = 1 atm as a function of the temperature for different AMOEBA models. The dashed line represents the density while using the Q-AMOEBA PIMD method, whereas the dot-dashed lines are obtained with the Q-AMOEBA adQTB method. Experimental data come from ref 90. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text).

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

Table 5. Average Energy Contributions of Q-AMOEBA Obtained with the Different Methods at 298.15 K[a]

	bond stretching	angle bending	Urey–Bradley	van der Waals	atomic multipoles	polarization	nonbonded
classical (2)	0.75	0.43	−0.03	5.48	−11.43	−5.19	−11.14
adQTB (2)	5.55	1.22	−0.11	4.37	−9.95	−4.43	−10.01
PIMD (1) (32 beads)	5.43	1.19	−0.12	4.32	−9.87	−4.35	−9.9
PIMD (1) (64 beads)	5.57	1.20	−0.12	4.33	−9.88	−4.38	−9.93

[a]The classical row corresponds to the adQTB set of parameter. All energies are given in kcal·mol^{-1} per water molecule and obtained from 1 ns simulations at constant volume corresponding to the density ρ = 0.997 g·cm^{-3}. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text).

This reasoning is further confirmed by the study of the different nonbonded energy contributions of the Q-AMOEBA force field in Table 5. PIMD values are superior by approximately 1.0 kcal·mol^{-1} per water molecule at 298.15 K to their classical counterpart, a trend which is consistent over the whole temperature range. Notably, the impact of NQEs on the hydrogen bond strength strongly depends on the underlying water model.[92] For example, in the q-TIP4P/f model, densities computed in PIMD or in classical MD are very similar,[65] whereas in the TTM2.1-F force field both curves are overestimated[93] and the PIMD density is lower than its classical counterpart at all temperatures.

It has been recognized that NQEs also play an important role in determining the enthalpy of vaporization ΔH_{vap} of water and that they should be taken into account explicitly for accurate prediction of this observable.[94,95] Assuming water vapor to be an ideal gas, one may find

$$\Delta H_{vap} = E_g - E_{liq} + P(V_g - V_l) \quad (7)$$

where E_g is the average total energy in the gas phase and E_{liq} is the average total energy per molecule in the liquid phase. As shown in Figure 4, classical enthalpy of vaporization is overestimated and decreases with the inclusion of NQEs. The Q-AMOEBA classical result (dashed light blue line) yields ΔH_{vap} = 11.5 kcal·mol^{-1} at 298.15 K which is ∼0.5 kcal·mol^{-1} higher than the reference. However, the PIMD and adQTB results underestimate the ΔH_{vap} by approximately ∼0.8 kcal·mol^{-1}, similar to other sophisticated models,[96] while capturing the correct slope of the curve with increasing temperature. Since the correct slope is recovered, some hypothesis tracking back the small resulting Q-AMOEBA discrepancies (compared to experiment) to the reference *ab initio* computations can be introduced. First, as we discussed the impact of NQEs on the hydrogen bond strength, it is highly probable that our CCSD(T) reference level does not fully reflect the experimental high complexity of the potential energy surfaces of water dimers and clusters. Indeed, at this level of accuracy, small changes in gas-phase reference computations could still impact the predicted condensed-phase properties. If the explicit inclusion of triple excitations clearly increases the strength of interactions,[97] then it has been shown that CCSDTQ approaches yield to relatively small but non-negligible improvement in the optimized CCSD(T) geometry and interaction energy for the water dimer.[98] Furthermore, one can guess that the simple Q-AMOEBA functional form probably reaches its limit in capturing highly complex electron density delocalization/overlap effects that would require the inclusion of higher-order excitations. Models like SIBFA[36] or AMOEBA+[34,35] that include additional overlap/delocalization effetcs such as electrostatic penetration, charge transfer, and so on, could be potentially better able to capture such nontrivial interactions.

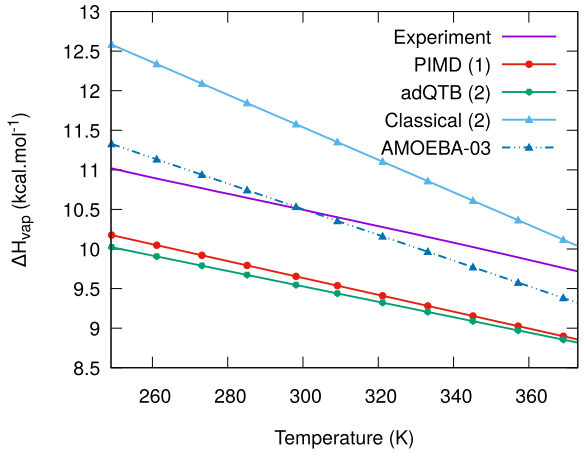


Figure 4. Enthalpy of vaporization at P = 1 atm as a function of the temperature with the different models. The dashed line represents the PIMD results and the dot-dashed lines the adQTB ones. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text). Experimental data (continuous line) are taken from ref 99.

The validity of Q-AMOEBA was further studied by computing other equilibrium properties of liquid water which were not included in the objective function of FB: the isobaric heat capacity c_p, the thermal expansion coefficient α_P, the isothermal compressibility κ_T, and the dielectric constant ε_r. The isobaric heat capacity is defined as

$$c_p = \left(\frac{\partial H}{\partial T}\right)_P \quad (8)$$

where H and T are the enthalpy and the temperature at a given pressure P. In practice, c_p is estimated as the temperature derivative of a fourth-order polynomial interpolating the values of H obtained at different temperatures. It is well-known that the heat capacity is strongly affected by NQEs. Indeed, at room temperature, the fluctuations of the high-frequency intramolecular modes are almost exclusively due to zero-point energy effects and essentially independent of the temperature. Therefore, intramolecular energy terms give almost no contribution to the heat capacity, which is not the case in the classical picture, leading to a significant overestimation of c_p when NQEs are not included explicitly.[94] Indeed, as shown in Figure 5, classical MD simulations overestimate c_p for both Q-AMOEBA and AMOEBA03. On the contrary, including NQEs allows recovering lower values much closer to experiment.

The thermal expansion coefficient α_P is calculated using analytic differentiation of a polynomial fit of the simulated density $\rho(T)$:

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

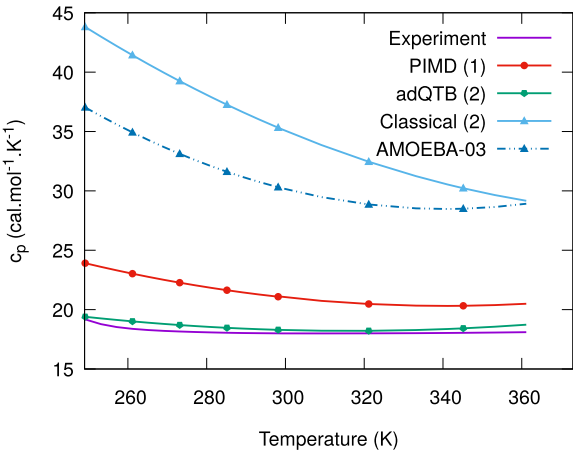


Figure 5. Isobaric heat capacity at P = 1 atm as a function of the temperature with the different models. The dashed line represents the PIMD results and the dot-dashed lines the adQTB ones. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text). Experimental data are taken from ref 99.

$$\alpha_p = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_P = -\frac{\mathrm{d}\ln\rho(T)}{\mathrm{d}}T \tag{9}$$

It shows the same trend as described above for the density (Figure 6): Classical MD yields too negative results at low temperature, a trend corrected by the inclusion of NQEs. At high temperatures, all the methods behave similarly.

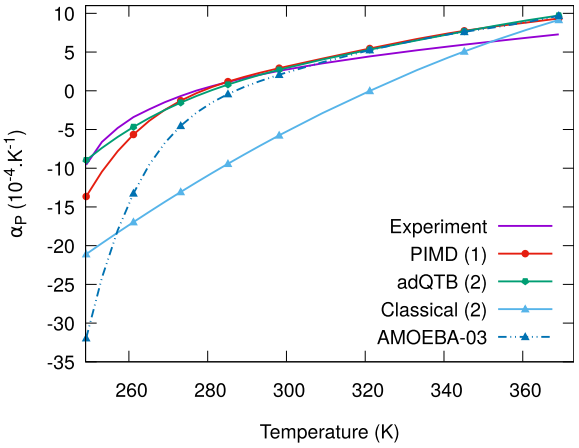


Figure 6. Thermal expansion coefficient at P = 1 atm as a function of the temperature with the different set of parameters. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text). Experimental data are taken from ref 90.

Longer *NPT* simulations of 10 ns were performed to converge the static dielectric constant and the self-diffusion coefficient and 20 ns for the isothermal compressibility (Figure 7). The isothermal compressibility that characterizes the volume change as a response to an applied pressure can also be related to the volume fluctuations in an *NPT* simulation:

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_{T,N} = \frac{1}{k_B T}\frac{\langle V^2\rangle - \langle V\rangle^2}{\langle V\rangle} \tag{10}$$

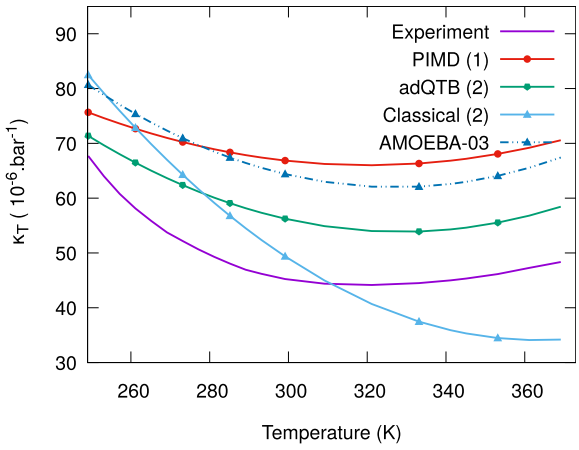


Figure 7. Isothermal compressibility at P = 1 atm as a function of the temperature with the different models. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text). Experimental data are taken from ref 100.

Simulations with classical nuclei in the Q-AMOEBA model tend to amplify the variations of this observable with T and to underestimate its value at high temperature. The explicit inclusion of NQEs reduces the variations of κ_T over the explored temperature range, and yield a shape that is very similar to the experimental curve, though with slightly larger values (particularly for the PIMD model). Overall, the results of the adQTB model are significantly improved with respect to that of the original AMOEBA03, whereas the PIMD ones are similar to it, which highlights that the original FF implicitly takes into account NQEs in its parametrization.

The static dielectric constant is calculated from the fluctuations of the total dipole moment as

$$\varepsilon_r = 1 + \frac{4\pi}{3k_B T\langle V\rangle}(\langle\mu^2\rangle - \langle\mu\rangle\cdot\langle\mu\rangle) \tag{11}$$

where $\langle\mu\rangle$ is the average total dipole moment defined in eq 5 and $\langle V\rangle$ is the average volume of the simulation box. The results (Figure 8) are satisfactory for ε_r with Q-AMOEBA and show the impact of NQEs on this property: They significantly reduce ε_r at low temperature and have a smaller influence when T increases.

The self-diffusion coefficient was evaluated at 298.15 K under 1 atm using the Einstein equation:

$$D_0 = \lim_{t\to\infty}\frac{\mathrm{d}}{\mathrm{d}t}\langle|r(t) - r(t_0)|^2\rangle \tag{12}$$

The mild Langevin thermostat (friction of γ = 1 ps^{-1}) applied to the centroid in the TRPMD method has been shown to have only a small effect on the computed diffusion whereas the adQTB method requires larger frictions (γ = 20 ps^{-1}) which affects the diffusion.[65,101] The PIMD (TRPMD) value is 2.29 ± 0.01 × 10^5 cm^2·s^{-1} which is in excellent agreement with the experimental value of 2.30 × 10^5 cm^2·s^{-1} and larger than the associated classical one (0.91 ± 0.01 × 10^5 cm^2·s^{-1}). The quantum/classical ratio of this quantity has been observed to vary widely among water models.[52,53,93] However, the adQTB value is 0.78 ± 0.01 × 10^5 cm^2·s^{-1} which is around 3 times lower than the expected value due to the large γ of the Langevin thermostat.[65] The slower diffusion can be associated with a less efficient sampling,[39] but this is largely compensated

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

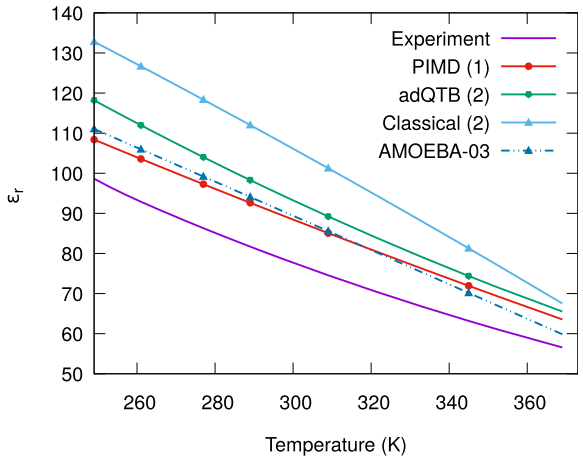


Figure 8. Static dielectric constant at P = 1 atm as a function of the temperature with the different set of parameters. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text). Experimental data are taken from ref 99.

by the considerable speed up of the adQTB method compared to path integral ones.

Finally, the molecular dipole moments were also studied in order to further confirm the previous interpretation regarding the impact of NQEs with respect to the underlying hydrogen bonds strength. Figure 9 shows that in classical MD

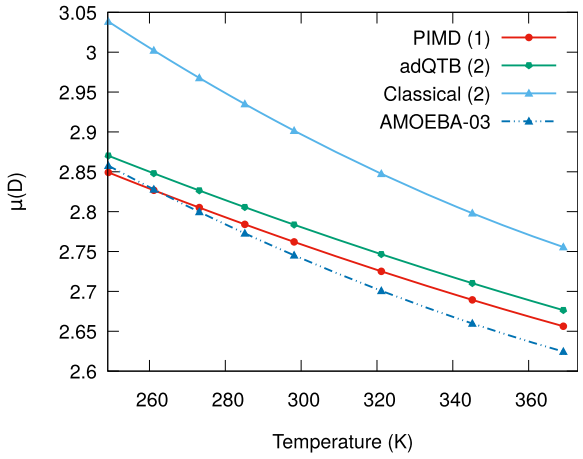


Figure 9. Molecular dipole moments at P = 1 atm as a function of the temperature with the different set of parameters. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text).

simulations with the Q-AMOEBA model the dipole moment is larger than in the original AMOEBA03 force field, indicating stronger hydrogen bonds in the new model. At room temperature, the value obtained from classical simulations with the Q-AMOEBA model is 2.90 D, whereas explicitly including NQEs via PIMD or adQTB weakens the H-bonds and reduces the dipole moment to 2.76 or 2.78 D, respectively. These values are close to those obtained in AMOEBA03, and in agreement with *ab initio* simulations and experiments.[102,103]

Heavy Water. Isotope effects are an important marker of the relevance of NQEs, since classical thermodynamics predicts different isotopes to have identical thermal equilibrium properties. Therefore, explicit treatment of NQEs is necessary to describe effects such as those arising from the substitution of H by its heavier isotope D.[104] To further validate the Q-AMOEBA approach, we present in Table 6 the values for heavy water D_2O of some thermodynamic properties previously presented, focusing on those known to be most impacted by the isotopic substitution.

Table 6. Thermodynamic Properties Computed at 298.15 K under 1 atm Pressure for Both Q-AMOEBA Models for H_2O and D_2O Compared to Experimental Values[99,105,a]

	H_2O			D_2O		
	PIMD (1)	adQTB (2)	exp.	PIMD (1)	adQTB (2)	exp.
ρ	0.998	0.999	0.997	1.109	1.106	1.104
ΔH	9.65	9.55	10.52	10.25	10.28	10.85
c_p	21.087	18.297	18.002	22.724	21.885	20.148

[a]The density ρ is given in g·cm^{-3}, the enthalpy of vaporization ΔH_{vap} in kcal·mol^{-1} and the isobaric heat capacity c_p in cal·mol^{-1}·K^{-1}. The number in parentheses refers to the set of force field parameters used in the simulations: (1) PIMD; (2) adQTB (see text).

Overall, Q-AMOEBA yields the correct trends for the isotopic substitution of H by D with an increase of the heat capacity c_p and of ΔH_{vap}. For the latter, the amplitude of the change is overestimated by a factor of approximately 2, with a change of ∼0.6 kcal·mol^{-1} for Q-AMOEBA, with respect to the experimental value of 0.3 kcal·mol^{-1}. Concerning c_p, the isotope effect is also overestimated by Q-AMOEBA (adQTB) with an increase of ∼3.6 kcal·mol^{-1} compared to the experimental 1.9 kcal·mol^{-1}. The Q-AMOEBA (PIMD) model is more accurate in its prediction of the increase of c_p under deuteration, but it slightly overestimates the heat capacity in general (for both H_2O and D_2O). This different behavior of the two Q-AMOEBA variants for this observable might be related to the slight differences in the intramolecular parameters of the two models. It is very promising that the Q-AMOEBA model is able to capture qualitatively the effects of the deuteration of water, even though the changes of the properties with isotope substitution are sometimes overestimated (it was also the case for the isotope shift of the temperature of maximum density T_{MD}, discussed in the "Thermodynamic Properties" section). The remaining discrepancies indicate that, within the tunability offered by the AMOEBA functional form, the optimization via FB leads to parameters that yield slightly too weak hydrogen bonds, which in turns leads to an overestimation of isotope effects since weak H-bonds tend to be further weakened by NQEs.[45,106] This interpretation is also consistent with the slight underestimation of ΔH_{vap} by Q-AMOEBA and with the diffusion coefficient found for D_2O: 1.45 ± 0.02 × 10^5 cm^2·s^{-1} compared to the experimental value of 1.77× 10^5 cm^2·s^{-1}.[45,107] More refined FFs (such as the AMOEBA+ functional form), when suitably reparametrized to include NQEs explicitly, might, in the future, offer even more quantitatively accurate predictions for isotope effects in water and biological systems.

Ice I_c. The study was further extended by studying the density of ice phase I_c at 78.0 K under 0 atm of Q-AMOEBA. A total of 128 beads were used to converge PIMD simulations. Q-AMEOBA gives 0.893 and 0.906 g·cm^{-3} in PIMD and adQTB, respectively, whereas the experimental value is 0.931 g·cm^{-3},[108] thus giving a maximum difference of ∼4%. Although classical MD simulations with AMOEBA03 yield a density

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

∼2% smaller than the experimental value, PIMD simulations with the same model underestimate the density by ∼6.8%. Hence, even if Q-AMOEBA probably tends to slightly overemphasize NQEs, it still improves considerably compared to PIMD simulations with AMOEBA03. This indicates that the satisfactory results obtained for ice with AMOEBA03 in classical MD simulations are due to error compensations and an implicit inclusion of NQEs in the force field parametrization, which limits its transferability. We therefore expect Q-AMOEBA to offer overall more robust predictions in this low-temperature range.

CONCLUSION

The newly introduced Q-AMOEBA model allows capturing NQEs with an improved accuracy thanks to their explicit inclusion in the parametrization process. The computational load of this parametrization was significantly reduced by resorting to the adaptive Quantum Thermal Bath method. The final Q-AMOEBA model is significantly different from the original AMOEBA03 force field and the explicit inclusion of NQEs led to an improved intramolecular potential with O−H bond length and H−O−H angle parameters closer to their gas-phase values compared to the original AMOEBA03. While Q-AMOEBA is shown to accurately reproduce gas-phase quantum chemistry computations, even for large molecular aggregates, it is also shown to be extremely robust for the prediction of condensed-phase properties. Improved results on liquid water compared to AMOEBA03 are observed, while transferability to ice is shown to be similar to previous models. The IR absorption spectrum is also better reproduced with Q-AMOEBA. Importantly, the model is also able to capture the experimental trends associated with the deuteration of water. Isotope effects are a valuable indicator of NQEs since they are absent in a classical description of the nuclei. Q-AMOEBA yields qualitatively correct predictions for the changes of thermodynamic quantities in heavy water compared to normal water, although it tends to overestimate these changes. We interpret it as a sign that the hydrogen bonds strength is slightly underestimated by Q-AMOEBA, which should be improved with more advanced functional forms such as AMOEBA+ and improved reference *ab initio* computations including higher level of couple cluster excitations. Concerning the functional form of the force field, it is shown that the inclusion of the many-body polarization effects lead to nontrivial interplay with NQEs. Since adQTB simulations can be performed at near classical cost, the improved Q-AMOEBA model can be easily extended to organic molecules, proteins, and nucleic acids, opening the possibility for the large-scale study of the importance of NQEs in biophysics.

TECHNICAL APPENDIX

Derivatives of Average Properties with Respect to Parameters in the Path Integrals Formalism

In order to optimize, the parameters of the force field using experimental data on average properties (for example, densities, enthalpy of vaporization, etc.), ForceBalance must be able to compute derivatives of these properties with respect to the FF parameters. While it is possible to simply use finite differences to numerically perform these derivatives, it is well-known that they are very sensitive to statistical noise on average values (computed over a molecular dynamics simulation) so that very long simulations would be necessary. Furthermore, one would have to perform multiple simulations with slightly different parameters in order to compute a single derivative. In ref 73, the authors propose to use the explicit form of the partition function in order to compute parametric derivatives as an average value over a single MD simulation with the current parameters. For a parameter λ and an observable $A(\mathbf{r};\lambda)$, the derivative can be expressed in the classical *NPT* ensemble as

$$\frac{\partial\langle A\rangle_\lambda}{\partial\lambda} = \langle\frac{\partial A}{\partial\lambda}\rangle_\lambda - \beta\langle A\frac{\partial E}{\partial\lambda}\rangle_\lambda - \langle A\rangle_\lambda\langle\frac{\partial E}{\partial\lambda}\rangle_\lambda \tag{13}$$

where $E(\mathbf{r}, V;\lambda)$ is the potential energy and $\langle\ldots\rangle_\lambda$ denotes an expectation value over the distribution $\rho(\mathbf{r};\lambda) \propto e^{-\beta(E(\mathbf{r},V;\lambda)+PV)}$.

This expression can be generalized to PI simulations using the same process with the path integral partition function. Let us denote $\langle\ldots\rangle_{N,\lambda}$ as the expectation value over the N-beads path integral distribution:

$$\rho_N(\mathbf{r}_1, \ldots, \mathbf{r}_N;\lambda) = \frac{e^{-\beta[\sum_{i=1}^N \frac{E(\mathbf{r}_i V;\lambda)}{N} + K(\mathbf{r}_1,\ldots,\mathbf{r}_N)+PV]}}{\mathcal{Z}_N(\lambda)} \tag{14}$$

with $K(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ the PI harmonic energy that couples the different beads (which is independent of λ) and $\mathcal{Z}_N(\lambda)$ the partition function that normalizes ρ_N. For convenience, let us denote $\bar{O} = \sum_{i=1}^N O(\mathbf{r}_i, V;\lambda)/N$ the average over the beads for any function O. The parametric derivative of $\langle A\rangle_{N,\lambda}$ is then:

$$\begin{aligned}
\frac{\partial\langle A\rangle_{N,\lambda}}{\partial\lambda} &= \frac{\partial}{\partial\lambda}\left[\int d\mathbf{r}_1 \ldots d\mathbf{r}_N\, \bar{A}(\lambda)\rho_N(\mathbf{r}_1, \ldots, \mathbf{r}_N;\lambda)\right] \\
&= \left\langle\frac{\partial\bar{A}}{\partial\lambda}\right\rangle_{N,\lambda} - \beta\left\langle\bar{A}(\lambda)\left(\frac{1}{N}\sum_{i=1}^N\frac{\partial E(\mathbf{r}_i;V;\lambda)}{\partial\lambda}\right)\right\rangle_{N,\lambda} \\
&- \langle\bar{A}(\lambda)\rangle_{N,\lambda}\frac{1}{\mathcal{Z}_N(\lambda)}\frac{\partial\mathcal{Z}_N}{\partial\lambda} = \left\langle\frac{\overline{\partial A}}{\partial\lambda}\right\rangle_{Z,N} \\
&- \beta\left[\left\langle\bar{A}\frac{\overline{\partial E}}{\partial\lambda}\right\rangle_{N,\lambda} - \langle\bar{A}\rangle_{N,\lambda}\left\langle\frac{\overline{\partial E}}{\partial\lambda}\right\rangle_{N,\lambda}\right]
\end{aligned}$$

which is almost identical to the classical expression except that all observables are averaged over the beads. This expression can be used in ForceBalance in order to fit the force field parameters on properties computed using quantum nuclei in the path integrals framework.

Pressure Estimator

NPT simulations were performed using the Langevin-piston method to obtain all relevant constant-pressure properties. The isotropic pressure estimator is given by

$$P_{\text{int}} = \frac{2K}{3V} - \frac{dU}{dV} \tag{15}$$

with V the volume of the simulated box and U the interatomic potential energy. In the PIMD framework, the kinetic energy K is given by the centroid virial estimator, whereas in the adQTB it is obtained as

$$K(v_1, v_2, \ldots, v_{3N}) = \eta^{-1}\sum_i\frac{1}{2}m_i v_i^2 \tag{16}$$

where η is a correction factor obtained through a deconvolution procedure in order to correct for the systematic error in the adQTB estimation of the kinetic energy due to the spectral broadening induced by the Langevin dynamics with

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

relatively large friction coefficients γ.[65] Despite this powerful procedure to correct the pressure estimation in adQTB, some differences remain with respect to PIMD calculations with the same set of parameters. More precisely, within the Q-AMOEBA (PIMD) model and for *NVT* simulations at 300 K, the average of the second term in eq 15 is −12840 and −12680 atm in adQTB and PIMD respectively, i.e., a difference of ∼1.3%. The averages of the term (kinetic part) are equal respectively to 13040 and 12640 atm (∼3.2% difference, though the PIMD value might still increase slightly for larger bead numbers). These discrepancies result in a ∼160 atm difference in the total pressure estimation. Although relatively small for condensed-phase simulations, this discrepancy causes a slight error on the density estimation in adQTB (of the order of 1%) which is the main reason for the use of two different Q-AMOEBA sets of parameters for adQTB and PIMD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c04454.

Details setup of the ForceBalance parameters optimization procedure (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Louis Lagardère − *Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France*; Email: louis.lagardere@sorbonne-universite.fr

Simon Huppert − *Sorbonne Université, Institut des NanoSciences de Paris, UMR 7588 CNRS, 75005 Paris, France*; Email: simon.huppert@sorbonne-universite.fr

Jean-Philip Piquemal − *Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France*; orcid.org/0000-0001-6615-9426; Email: jean-philip.piquemal@sorbonne-université.fr

Authors

Nastasia Mauger − *Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France*

Thomas Plé − *Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c04454

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 810367), project EMC2 (JPP). Computations have been performed at GENCI (TGCC, Bruyères le Châtel) on grant no A0070707671.

■ REFERENCES

(1) Goh, B. C.; Hadden, J. A.; Bernardi, R. C.; Singharoy, A.; McGreevy, R.; Rudack, T.; Cassidy, C. K.; Schulten, K. Computational methodologies for real-space structural refinement of large macromolecular complexes. *Annual Review of Biophysics* **2016**, *45*, 253−278.

(2) Misquitta, A. J.; Stone, A. J.; Fazeli, F. Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. *J. Chem. Theory Comput.* **2014**, *10*, 5405−5418.

(3) Walsh, T. R.; Knecht, M. R. Biointerface structural effects on the properties and applications of bioinspired peptide-based nanomaterials. *Chem. Rev.* **2017**, *117*, 12641−12704.

(4) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.

(5) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(6) Hansson, T.; Oostenbrink, C.; van Gunsteren, W. Molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2002**, *12*, 190−196.

(7) Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.* **2015**, *36*, 1990−2008.

(8) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology* **2017**, *13*, e1005659.

(9) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878−3888.

(10) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034−2053.

(11) Elber, R. Perspective: Computer simulations of long time dynamics. *J. Chem. Phys.* **2016**, *144*, 060901.

(12) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* **2004**, *120*, 10880−10889.

(13) Pierce, L. C.; Salomon-Ferrer, R.; de Oliveira, C. A.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 2997−3002.

(14) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(15) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **2013**, *3*, 198−210.

(16) Brooks, B. R.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.

(17) MacKerell, A. D.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(18) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Jr Current status of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **2015**, *1215*, 47−71.

(19) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(20) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657−1666.

(21) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(22) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(23) Sun, H. COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase ApplicationsOverview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1998**, *102*, 7338−7364.
(24) Melcr, J.; Piquemal, J.-P. Accurate Biomolecular Simulations Account for Electronic Polarization. *Front. Mol. Biosci.* **2019**, *6*, 143.
(25) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Polarizable Force Fields for Biomolecular Modeling. *Rev. Comput. Chem.* **2015**, *28*, 51−86.
(26) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371−394.
(27) Patel, S.; Mackerell, A. D., Jr; Brooks, C. L., 3rd CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004**, *25*, 1504−1514.
(28) Patel, S.; Brooks, C. L., 3rd CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004**, *25*, 1−15.
(29) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: Optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.
(30) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: first generation model and gas phase tests. *J. Comput. Chem.* **2002**, *23*, 1515−1531.
(31) Mei, Y.; Simmonett, A. C.; Pickard, F. C.; DiStasio, R. A.; Brooks, B. R.; Shao, Y. Numerical Study on the Partitioning of the Molecular Polarizability into Fluctuating Charge and Induced Atomic Dipole Contributions. *J. Phys. Chem. A* **2015**, *119*, 5865−5882.
(32) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(33) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(34) Liu, C.; Piquemal, J.-P.; Ren, P. AMOEBA+ Classical Potential for Modeling Molecular Interactions. *J. Chem. Theory Comput.* **2019**, *15*, 4122−4139.
(35) Liu, C.; Piquemal, J.-P.; Ren, P. Implementation of Geometry-Dependent Charge Flux into the Polarizable AMOEBA+ Potential. *J. Phys. Chem. Lett.* **2020**, *11*, 419−426.
(36) Naseem-Khan, S.; Lagardère, L.; Narth, C.; Cisneros, G. A.; Ren, P.; Gresh, N.; Piquemal, J.-P. Development of the Quantum-Inspired SIBFA Many-Body Polarizable Force Field: Enabling Condensed-Phase Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2022**, *18* (6), 3607−3621.
(37) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(38) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(39) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Springer, 1981; pp 331−342.
(40) Larini, L.; Lu, L.; Voth, G. A. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Chem. Phys.* **2010**, *132*, 164107.
(41) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112*, 8910−8922.
(42) Kiss, P. T.; Baranyai, A. A systematic development of a polarizable potential of water. *J. Chem. Phys.* **2013**, *138*, 204507.
(43) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(44) Vega, C.; Abascal, J. L. F.; Conde, M. M.; Aragones, J. L. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss.* **2009**, *141*, 251−276.
(45) 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. *Chem. Rev.* **2016**, *116*, 7529−7550.
(46) Agarwal, P. K.; Billeter, S. R.; Hammes-Schiffer, S. Nuclear quantum effects and enzyme dynamics in dihydrofolate reductase catalysis. *J. Phys. Chem. B* **2002**, *106*, 3283−3293.
(47) Zhang, X.; Zhou, S.; Leonik, F. M.; Wang, L.; Kuroda, D. G. Quantum Mechanical Effects in Acid-Base Chemistry. *Chem. Sci.* **2022**, *13*, 6998−7006.
(48) Pereyaslavets, L.; Kurnikov, I.; Kamath, G.; Butin, O.; Illarionov, A.; Leontyev, I.; Olevanov, M.; Levitt, M.; Kornberg, R. D.; Fain, B. On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 8878−8882.
(49) Feynman, R. P.; Hibbs, A. R.; Styer, D. F. *Quantum Mechanics and Path Integrals*; Courier Corporation, 2010.
(50) Chandler, D.; Wolynes, P. G. Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.* **1981**, *74*, 4078−4095.
(51) McBride, C.; Vega, C.; Noya, E. G.; Ramírez, R.; Sesé, L. M. Quantum contributions in the ice phases: The path to a new empirical model for water−TIP4PQ/2005. *J. Chem. Phys.* **2009**, *131*, 024506.
(52) Paesani, F.; Zhang, W.; Case, D. A.; Cheatham, T. E.; Voth, G. A. An accurate and simple quantum model for liquid water. *J. Chem. Phys.* **2006**, *125*, 184507.
(53) Habershon, S.; Markland, T. E.; Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* **2009**, *131*, 024501.
(54) Kapil, V.; Behler, J.; Ceriotti, M. High order path integrals made easy. *J. Chem. Phys.* **2016**, *145*, 234103.
(55) Pérez, A.; Tuckerman, M. E. Improving the convergence of closed and open path integral molecular dynamics via higher order Trotter factorization schemes. *J. Chem. Phys.* **2011**, *135*, 064104.
(56) Poltavsky, I.; Kapil, V.; Ceriotti, M.; Kim, K. S.; Tkatchenko, A. Accurate Description of Nuclear Quantum Effects with High-Order Perturbed Path Integrals (HOPPI). *J. Chem. Theory Comput.* **2020**, *16*, 1128−1135.
(57) Poltavsky, I.; Tkatchenko, A. Modeling quantum nuclei with perturbed path integral molecular dynamics. *Chem. Sci.* **2016**, *7*, 1368−1372.
(58) Brieuc, F.; Dammak, H.; Hayoun, M. Quantum Thermal Bath for Path Integral Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2016**, *12*, 1351−1359.
(59) Ceriotti, M.; Manolopoulos, D. E.; Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized Langevin equation. *J. Chem. Phys.* **2011**, *134*, 084104.
(60) Markland, T. E.; Manolopoulos, D. E. An efficient ring polymer contraction scheme for imaginary time path integral simulations. *J. Chem. Phys.* **2008**, *129*, 024105.
(61) Markland, T. E.; Manolopoulos, D. E. A refined ring polymer contraction scheme for systems with electrostatic interactions. *Chem. Phys. Lett.* **2008**, *464*, 256−261.
(62) Cheng, X.; Herr, J. D.; Steele, R. P. Accelerating ab initio path integral simulations via imaginary multiple-timestepping. *J. Chem. Theory Comput.* **2016**, *12*, 1627−1638.

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

(63) Fanourgakis, G. S.; Markland, T. E.; Manolopoulos, D. E. A fast path integral method for polarizable force fields. *J. Chem. Phys.* **2009**, *131*, 094102.
(64) 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. *J. Chem. Theory Comput.* **2019**, *15*, 2863−2880.
(65) Mauger, N.; Plé, T.; Lagardère, L.; Bonella, S.; Mangaud, E.; Piquemal, J.-P.; Huppert, S. Nuclear Quantum Effects in Liquid Water at Near Classical Computational Cost Using the Adaptive Quantum Thermal Bath. *J. Phys. Chem. Lett.* **2021**, *12*, 8285−8291.
(66) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic Improvement of a Classical Molecular Model of Water. *J. Phys. Chem. B* **2013**, *117*, 9956−9972.
(67) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956−972.
(68) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933−5947.
(69) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8566.
(70) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827−7843.
(71) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *J. Chem. Theory Comput.* **2005**, *1*, 1128−1132.
(72) Wang, L. *ForceBalance: Systematic Force Field Optimization*; SimTK, 2014.
(73) Wang, L.-P.; Head-Gordon, T.; Ponder, J. W.; Ren, P.; Chodera, J. D.; Eastman, P. K.; Martinez, T. J.; Pande, V. S. Systematic improvement of a classical molecular model of water. *J. Phys. Chem. B* **2013**, *117*, 9956−9972.
(74) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *journal of physical chemistry letters* **2014**, *5*, 1885−1891.
(75) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(76) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* **2014**, *140*, 234116.
(77) Xantheas, S. S.; Aprà, E. The binding energies of the D2d and S4 water octamer isomers: High-level electronic structure and empirical potential results. *J. Chem. Phys.* **2004**, *120*, 823−828.
(78) Yoo, S.; Aprà, E.; Zeng, X. C.; Xantheas, S. S. High-Level Ab Initio Electronic Structure Calculations of Water Clusters (H2O)16 and (H2O)17: A New Global Minimum for (H2O)16. *J. Phys. Chem. Lett.* **2010**, *1*, 3122−3127.
(79) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-level ab initio calculations for the four low-lying families of minima of (H2O)20. I. Estimates of MP2/CBS binding energies and comparison with empirical potentials. *J. Chem. Phys.* **2004**, *121*, 2655−2663.
(80) Jorgensen, W. L.; Jenson, C. Temperature dependence of TIP3P, SPC, and TIP4P water from *NPT* Monte Carlo simulations: Seeking temperatures of maximum density. *J. Comput. Chem.* **1998**, *19*, 1179−1186.
(81) Bates, D. M.; Tschumper, G. S. CCSD(T) complete basis set limit relative energies for low-lying water hexamer structures. *J. Phys. Chem. A* **2009**, *113*, 3555−3559.
(82) Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, 074506.
(83) Soper, A. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **2000**, *258*, 121−137.
(84) Bertie, J. E.; Ahmed, M. K.; Eysel, H. H. Infrared intensities of liquids. 5. Optical and dielectric constants, integrated intensities, and dipole moment derivatives of water and water-d2 at 22.degree.C. *J. Phys. Chem.* **1989**, *93*, 2210−2218.
(85) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press, 2001.
(86) Benson, R. L.; Trenins, G.; Althorpe, S. C. Which quantum statistics-classical dynamics method is best for water? *Faraday Discuss.* **2020**, *221*, 350−366.
(87) Plé, T.; Huppert, S.; Finocchi, F.; Depondt, P.; Bonella, S. Anharmonic spectral features via trajectory-based quantum dynamics: a perturbative analysis of the interplay between dynamics and sampling. *J. Chem. Phys.* **2021**, *155*, 104108.
(88) Rossi, M.; Kapil, V.; Ceriotti, M. Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation. *J. Chem. Phys.* **2018**, *148*, 102301.
(89) Impey, R.; Madden, P.; McDonald, I. Spectroscopic and transport properties of water: model calculations and the interpretation of experimental results. *Mol. Phys.* **1982**, *46*, 513−539.
(90) Kell, G. S. Density thermal expansivity, and compressibility of liquid water from 0.deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *Journal of Chemical & Engineering Data* **1975**, *20*, 97−105.
(91) Stillinger, F. H. Water revisited. *science* **1980**, *209*, 451−457.
(92) Li, C.; Paesani, F.; Voth, G. A. Static and Dynamic Correlations in Water: Comparison of Classical Ab Initio Molecular Dynamics at Elevated Temperature with Path Integral Simulations at Ambient Temperature. *J. Chem. Theory Comput.* **2022**, *18*, 2124−2131.
(93) Paesani, F.; Iuchi, S.; Voth, G. A. Quantum effects in liquid water from an ab initio-based polarizable force field. *J. Chem. Phys.* **2007**, *127*, 074506.
(94) Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F. On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016**, *145*, 194504.
(95) Guillot, B.; Guissani, Y. How to build a better pair potential for water. *J. Chem. Phys.* **2001**, *114*, 6720−6733.
(96) Medders, G. R.; Babin, V.; Paesani, F. Development of a "first-principles" water potential with flexible monomers. III. Liquid phase properties. *J. Chem. Theory Comput.* **2014**, *10*, 2906−2910.
(97) Reinhardt, P.; Piquemal, J.-P. New intermolecular benchmark calculations on the water dimer: SAPT and supermolecular post-Hartree–Fock approaches. *Int. J. Quantum Chem.* **2009**, *109*, 3259−3267.
(98) Lane, J. R. CCSDTQ Optimized Geometry of Water Dimer. *J. Chem. Theory Comput.* **2013**, *9*, 316−323.
(99) Wagner, W.; Pruß, A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water Substance for General and Scientific Use. *J. Phys. Chem. Ref. Data* **2002**, *31*, 387−535.
(100) Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0. deg. to 150.deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *Journal of Chemical and Engineering data* **1975**, *20*, 97−105.
(101) Maginn, E. J.; Messerly, R. A.; Carlson, D. J.; Roe, D. R.; Elliot, J. R. Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0]. Living. *J. Comput. Mol. Sci.* **2018**, *1*, 6324.
(102) Badyal, Y.; Saboungi, M.-L.; Price, D.; Shastri, S.; Haeffner, D.; Soper, A. Electron distribution in water. *J. Chem. Phys.* **2000**, *112*, 9206−9208.

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

(103) Gubskaya, A. V.; Kusalik, P. G. The total molecular dipole moment for liquid water. *J. Chem. Phys.* **2002**, *117*, 5290−5302.
(104) Paesani, F. Hydrogen bond dynamics in heavy water studied with quantum dynamical simulations. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19865−19875.
(105) Kell, G. S. Effects of isotopic composition, temperature, pressure, and dissolved gases on the density of liquid water. *J. Phys. Chem. Ref. Data* **1977**, *6*, 1109−1131.
(106) Li, X.-Z.; Walker, B.; Michaelides, A. Quantum nature of the hydrogen bond. *Proc. Natl. Acad. Sci. U.S.A* **2011**, *108*, 6369−6373.
(107) Price, W. S.; Ide, H.; Arata, Y.; Söderman, O. Temperature dependence of the self-diffusion of supercooled heavy water to 244 K. *J. Phys. Chem. B* **2000**, *104*, 5874−5876.
(108) Kuhs, W. F.; Bliss, D. V.; Finney, J. L. High-Resolution Neutron Powder Diffraction Study of Ice I_c. *J. Phys. Colloques* **1987**, *48*, C1-631−C1-636.

https://doi.org/10.1021/acs.jpcb.2c04454
J. Phys. Chem. B 2022, 126, 8813−8826

Bibliography

[1] T. Dauxois, Physics Today **61**, 55 (2008).

[2] C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot, and H. Grubmüller, *Best bang for your buck: Gpu nodes for gromacs biomolecular simulations* (2015).

[3] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al., PLoS computational biology **13**, e1005659 (2017).

[4] L. C. Pierce, R. Salomon-Ferrer, C. Augusto F. de Oliveira, J. A. McCammon, and R. C. Walker, Journal of chemical theory and computation **8**, 2997 (2012).

[5] O. Adjoua, L. Lagardère, L.-H. Jolly, A. Durocher, T. Very, I. Dupays, Z. Wang, T. J. Inizan, F. Célerse, P. Ren, et al., Journal of chemical theory and computation **17**, 2034 (2021).

[6] R. Elber, The Journal of chemical physics **144**, 060901 (2016).

[7] A. K. Faradjian and R. Elber, The Journal of chemical physics **120**, 10880 (2004).

[8] L. Etheve, J. Martin, and R. Lavery, Nucleic acids research **44**, 9990 (2016).

[9] M. Bagherpoor Helabad, S. Volkenandt, and P. Imhof, Frontiers in molecular biosciences **7**, 4 (2020).

[10] M. De Vivo, M. Masetti, G. Bottegoni, and A. Cavalli, Journal of medicinal chemistry **59**, 4035 (2016).

[11] L. Wang, S. D. Fried, S. G. Boxer, and T. E. Markland, Proceedings of the National Academy of Sciences **111**, 18454 (2014).

[12] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, Physical review letters **110**, 065702 (2013).

[13] K. M. Hickson, J.-C. Loison, D. Nuñez-Reyes, and R. Méreau, The journal of physical chemistry letters **7**, 3641 (2016).

[14] M. Benoit, D. Marx, and M. Parrinello, Nature **392**, 258 (1998).

[15] Y. Bronstein, P. Depondt, F. Finocchi, and A. M. Saitta, Physical Review B **89**, 214101 (2014).

[16] S. D. Ivanov, I. M. Grant, and D. Marx, The Journal of Chemical Physics **143**, 124304 (2015).

[17] C. Vega, M. M. Conde, C. McBride, J. L. F. Abascal, E. G. Noya, R. Ramírez, and L. M. Sesé, The Journal of chemical physics **132**, 046101 (2010).

[18] F. t. Bates and G. Wignall, Physical review letters **57**, 1429 (1986).

[19] B. S. El'Kin, Solid State Ionics **37**, 139 (1990).

[20] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, Chemical reviews **116**, 7529 (2016).

[21] F. Paesani, W. Zhang, D. A. Case, T. E. Cheatham III, and G. A. Voth, The Journal of chemical physics **125**, 184507 (2006).

[22] W. H. Miller, The Journal of chemical physics **125**, 132305 (2006).

[23] R. P. Feynman, in *Feynman's Thesis—A New Approach To Quantum Theory* (World Scientific, 2005), pp. 71–109.

[24] R. Feynman, *Space–time approach to non-relativistic quantum mechanics" rev. of mod* (1948).

[25] H. Dammak, Y. Chalopin, M. Laroche, M. Hayoun, and J.-J. Greffet, Phys. Rev. Lett. **103**, 190601 (2009).

[26] M. Ceriotti, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **103**, 030603 (2009).

[27] E. Mangaud, S. Huppert, T. Plé, P. Depondt, S. Bonella, and F. Finocchi, Journal of Chemical Theory and Computation **15**, 2863 (2019).

[28] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, et al., Chemical science **9**, 956 (2018).

[29] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, Proceedings of the National Academy of Sciences **115**, 8878 (2018).

[30] S. Habershon, T. E. Markland, and D. E. Manolopoulos, The journal of chemical physics **131**, 024501 (2009).

[31] P. Ren and J. W. Ponder, The Journal of Physical Chemistry B **107**, 5933 (2003).

[32] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, et al., The Journal of Physical Chemistry B **114**, 2549 (2010).

[33] E. Schrödinger, Phys. Rev. **28**, 1049 (1926).

[34] M. Born and R. Oppenheimer, Annalen der Physik **389**, 457 (1927).

[35] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[36] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[37] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[38] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[39] A. D. Becke, The Journal of Chemical Physics **98**, 1372 (1993).

[40] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, The Journal of Physical Chemistry **98**, 11623 (1994).

[41] A. J. Cohen, P. Mori-Sánchez, and W. Yang, Chemical Reviews **112**, 289 (2012).

[42] T. D. Crawford and H. F. Schaefer, *An Introduction to Coupled Cluster Theory for Computational Chemists* (Wiley-Blackwell, 2007).

[43] R. Car and M. Parrinello, Physical review letters **55**, 2471 (1985).

[44] B. J. Alder and T. E. Wainwright, The Journal of Chemical Physics **31**, 459 (1959).

[45] M. Griebel, K. S, and G. Zumbusch, *Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications.* (Springer Publishing Company, Incorporated, 2007).

[46] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation* (Oxford Graduate Texts, 2010).

[47] M. Ceriotti, D. E. Manolopoulos, T. E. Markland, and M. Rossi, *Path Integral Methods in Atomistic Modelling - An introduction* - (CECAM, 2021).

[48] L. Verlet, Phys. Rev. **159**, 98 (1967).

[49] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, The Journal of Chemical Physics **76**, 637 (1982).

[50] C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulations* (McGraw-Hill Book Company, 1985).

[51] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, The Journal of Chemical Physics **81**, 3684 (1984).

[52] S. Nosé, Molecular Physics **52**, 255 (1984).

[53] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, vol. 1 of *Computational Science Series* (Academic Press, San Diego, 2002).

[54] D. S. Lemons and A. Gythiel, American Journal of Physics **65**, 1079 (1997).

[55] M. Tuckerman, B. J. Berne, and G. J. Martyna, The Journal of Chemical Physics **97**, 1990 (1992).

[56] S. Melchionna, The Journal of Chemical Physics **127**, 044108 (2007).

[57] R. D. SKEEL and J. A. IZAGUIRRE, Molecular Physics **100**, 3885 (2002).

[58] J. Finkelstein, G. Fiorin, and B. Seibold, Molecular Physics **118**, e1649493 (2020).

[59] B. Leimkuhler and C. Matthews, Applied Mathematics Research eXpress **2013**, 34 (2012).

[60] González, M.A., JDN **12**, 169 (2011).

[61] F. Jensen, *Introduction to Computational Chemistry* (Wiley, 2017), 3rd ed.

[62] M. Gonzalez, *Force fields and molecular dynamics simulations*, vol. 12 (Ecole thématique de la Société Française de la Neutronique, 2011).

[63] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Journal of Computational Chemistry **30**, 1545 (2009).

[64] D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, J Comput Chem **26**, 1668 (2005).

[65] W. L. Jorgensen and J. Tirado-Rives, Journal of the American Chemical Society **110**, 1657 (1988).

[66] K. M. Visscher and D. P. Geerke, The Journal of Physical Chemistry B **124**, 1628 (2020).

[67] M. Mohebifar and C. N. Rowley, The Journal of Chemical Physics **153**, 134105 (2020).

[68] I. M. Klotz, F. M. Walker, and R. B. Pivan, Journal of the American Chemical Society **68**, 1486 (1946).

[69] N. L. Allinger, Y. H. Yuh, and J. H. Lii, Journal of the American Chemical Society **111**, 8551 (1989).

[70] T. Köddermann, D. Paschek, and R. Ludwig, ChemPhysChem **8**, 2464 (2007).

[71] R. G. Parr and R. G. Pearson, Journal of the American Chemical Society **105**, 7512 (1983).

[72] D. Ongari, P. G. Boyd, O. Kadioglu, A. K. Mace, S. Keskin, and B. Smit, Journal of Chemical Theory and Computation **15**, 382 (2019).

[73] J. Chen and T. J. Martı´nez, Chemical Physics Letters **463**, 288 (2008).

[74] G. Lee Warren, J. E. Davis, and S. Patel, The Journal of Chemical Physics **128**, 144110 (2008).

[75] D. Mathieu, The Journal of Chemical Physics **127**, 224103 (2007).

[76] R. A. Nistor, J. G. Polihronov, M. H. Müser, and N. J. Mosey, The Journal of Chemical Physics **125**, 094108 (2006).

[77] S. Patel and C. L. Brooks, 3rd, J Comput Chem **25**, 1 (2004).

[78] S. Patel, A. D. Mackerell, Jr, and C. L. Brooks, 3rd, J Comput Chem **25**, 1504 (2004).

[79] G. Lamoureux and B. Roux, The Journal of Chemical Physics **119**, 3025 (2003).

[80] L. Lagardère, F. Aviat, and J.-P. Piquemal, Journal of Physical Chemistry Letters **10**, 2593-2599 (2019).

[81] J. Applequist, J. R. Carl, and K.-K. Fung, Journal of the American Chemical Society **94**, 2952 (1972).

[82] Y. Mei, A. C. Simmonett, F. C. Pickard, R. A. DiStasio, B. R. Brooks, and Y. Shao, The Journal of Physical Chemistry A **119**, 5865 (2015).

[83] D. Bedrov, J.-P. Piquemal, O. Borodin, A. D. MacKerell, B. Roux, and C. Schröder, Chemical Reviews **119**, 7940 (2019).

[84] P. E. M. Lopes, B. Roux, and A. D. MacKerell, Theoretical Chemistry Accounts **124**, 11 (2009).

[85] P. P. Ewald, Annalen der Physik **369**, 253 (1921).

[86] H. A. Stern, G. A. Kaminski, J. L. Banks, R. Zhou, B. J. Berne, and R. A. Friesner, The Journal of Physical Chemistry B **103**, 4730 (1999).

[87] H. S. Smalø, P.-O. Åstrand, and L. Jensen, The Journal of Chemical Physics **131**, 044101 (2009).

[88] J. Applequist, J. R. Carl, and K.-K. Fung, Journal of the American Chemical Society **94**, 2952 (1972).

[89] B. T. Thole, Chemical Physics **59**, 341 (1981).

[90] C. M. University, *Model box periodic boundary conditions - p.b.c.*, `http://isaacs.sourceforge.net/phys/pbc.html` (2022).

[91] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids: Second Edition* (Oxford University Press, Oxford, 2017), 2nd ed., ISBN 9780198803195.

[92] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, The Journal of Chemical Physics **103**, 8577 (1995).

[93] T. Darden, D. York, and L. Pedersen, The Journal of Chemical Physics **98**, 10089 (1993).

[94] F. H. Stillinger and A. Rahman, The Journal of Chemical Physics **57**, 1281 (1972).

[95] F. H. Stillinger and A. Rahman, The Journal of Chemical Physics **60**, 1545 (1974).

[96] A. Laio and M. Parrinello, Proceedings of the National Academy of Sciences **99**, 12562 (2002).

[97] A. F. Voter, The Journal of Chemical Physics **106**, 4665 (1997).

[98] A. Warshel and M. Levitt, Journal of Molecular Biology **103**, 227 (1976).

[99] T. E. Markland and M. Ceriotti, Nature Reviews Chemistry **2**, 0109 (2018).

[100] L. Wang, S. D. Fried, S. G. Boxer, and T. E. Markland, Proceedings of the National Academy of Sciences **111**, 18454 (2014).

[101] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, Phys. Rev. Lett. **110**, 065702 (2013).

[102] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, Proceedings of the National Academy of Sciences **115**, 8878 (2018).

[103] R. P. Feynman, A. R. Hibbs, and D. F. Styer, *Quantum mechanics and path integrals* (Courier Corporation, 2010).

[104] J. Cao and G. A. Voth, The Journal of chemical physics **101**, 6168 (1994).

[105] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III, Annual review of physical chemistry **64**, 387 (2013).

[106] H. Kimizuka, H. Mori, and S. Ogata, Physical Review B **83**, 094110 (2011).

[107] T. F. Miller III and D. E. Manolopoulos, The Journal of chemical physics **122**, 184503 (2005).

[108] B. J. Berne and D. Thirumalai, Annual Review of Physical Chemistry **37**, 401 (1986).

[109] A. G. Csaszar, C. Fabri, T. Szidarovszky, E. Matyus, T. Furtenbacher, and G. Czako, Physical Chemistry Chemical Physics **14**, 1085 (2012).

[110] R. Kubo, Reports on progress in physics **29**, 255 (1966).

[111] H. B. Callen and T. A. Welton, Physical Review **83**, 34 (1951).

[112] Y. Bronstein, P. Depondt, L. E. Bove, R. Gaal, A. M. Saitta, and F. Finocchi, Physical Review B **93**, 024104 (2016).

[113] H. Dammak, E. Antoshchenkova, M. Hayoun, and F. Finocchi, Journal of Physics: Condensed Matter **24**, 435402 (2012).

[114] M. Ceriotti, G. Miceli, A. Pietropaolo, D. Colognesi, A. Nale, M. Catti, M. Bernasconi, and M. Parrinello, Physical Review B **82**, 174306 (2010).

[115] K. Drużbicki, M. Krzystyniak, D. Hollas, V. Kapil, P. Slavíček, G. Romanelli, and F. Fernandez-Alonso, in *Journal of Physics: Conference Series* (IOP Publishing, 2018), vol. 1055, p. 012003.

[116] T. Qi and E. J. Reed, The Journal of Physical Chemistry A **116**, 10451 (2012).

[117] M. Parrinello and A. Rahman, The Journal of chemical physics **80**, 860 (1984).

[118] R. W. Hall and B. J. Berne, The Journal of chemical physics **81**, 3641 (1984).

[119] T. E. Markland and D. E. Manolopoulos, Chemical Physics Letters **464**, 256 (2008).

[120] T. E. Markland and D. E. Manolopoulos, The Journal of chemical physics **129**, 024105 (2008).

[121] M. Herman, E. Bruskin, and B. Berne, The Journal of Chemical Physics **76**, 5150 (1982).

[122] W. Janke and T. Sauer, The Journal of chemical physics **107**, 5821 (1997).

[123] T. M. Yamamoto, The Journal of chemical physics **123**, 104101 (2005).

[124] B. Berne and G. Harp, Advances in chemical physics **17**, 63 (2009).

[125] M. Toda, R. Kubo, N. Saitō, N. Hashitsume, et al., *Statistical physics II: nonequilibrium statistical mechanics*, vol. 2 (Springer Science & Business Media, 1991).

[126] R. Zwanzig, *Nonequilibrium statistical mechanics* (Oxford university press, 2001).

[127] I. R. Craig and D. E. Manolopoulos, The Journal of chemical physics **121**, 3368 (2004).

[128] J. Cao and G. A. Voth, The Journal of chemical physics **99**, 10070 (1993).

[129] G. Martyna, The Journal of chemical physics **104**, 2018 (1996).

[130] J. Cao and G. Martyna, The Journal of chemical physics **104**, 2028 (1996).

[131] D. Marx, M. E. Tuckerman, and G. J. Martyna, Computer physics communications **118**, 166 (1999).

[132] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, Molecular Physics **87**, 1117 (1996).

[133] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, The Journal of chemical physics **130**, 194510 (2009).

[134] S. D. Ivanov, A. Witt, M. Shiga, and D. Marx, The Journal of Chemical Physics **132**, 031101 (2010).

[135] G. Trenins, M. J. Willatt, and S. C. Althorpe, The Journal of Chemical Physics **151**, 054109 (2019).

[136] I. R. Craig and D. E. Manolopoulos, The Journal of chemical physics **122**, 084106 (2005).

[137] I. R. Craig and D. E. Manolopoulos, The Journal of chemical physics **123**, 034102 (2005).

[138] I. R. Craig and D. E. Manolopoulos, Chemical physics **322**, 236 (2006).

[139] T. F. Miller III and D. E. Manolopoulos, The Journal of chemical physics **123**, 154504 (2005).

[140] S. Habershon, G. S. Fanourgakis, and D. E. Manolopoulos, The Journal of chemical physics **129**, 074501 (2008).

[141] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, The Journal of chemical physics **140**, 234116 (2014).

[142] M. Langevin, in *Annales de chimie et de physique, Series* (1905), vol. 5, pp. 245–288.

[143] M. MASCART, Comptes rendus hebdomadaires des séances de l'Académie des sciences, série physique pp. 530–533 (1908).

[144] H. A. Kramers, Physica **7**, 284 (1940).

[145] R. F. Grote and J. T. Hynes, The Journal of Chemical Physics **73**, 2715 (1980).

[146] S. Huppert, T. Plé, S. Bonella, P. Depondt, and F. Finocchi, Applied Sciences **12**, 4756 (2022).

[147] R. Kubo, M. Toda, and N. Hashitsume, *Statistical physics II: nonequilibrium statistical mechanics*, vol. 31 (Springer Science & Business Media, 2012).

[148] O. Bedoya-Martinez, J.-L. Barrat, and D. Rodney, Physical Review B **89**, 014303 (2014).

[149] F. Brieuc, Y. Bronstein, H. Dammak, P. Depondt, F. Finocchi, and M. Hayoun, Journal of Chemical Theory and Computation **12**, 5688 (2016).

[150] J.-L. Barrat and D. Rodney, Journal of Statistical Physics **144**, 679 (2011).

[151] I. R. Sims, Nature chemistry **5**, 734 (2013).

[152] K. M. Hickson, J.-C. Loison, D. Nuñez-Reyes, and R. Méreau, The journal of physical chemistry letters **7**, 3641 (2016).

[153] P. W. Bessonette and M. A. White, The Journal of chemical physics **110**, 3919 (1999).

[154] Y. Bronstein, P. Depondt, and F. Finocchi, European Journal of Mineralogy **29**, 385 (2017).

[155] A. Sano-Furukawa, H. Kagi, T. Nagai, S. Nakano, S. Fukura, D. Ushijima, R. Iizuka, E. Ohtani, and T. Yagi, American Mineralogist **94**, 1255 (2009).

[156] F. H. Stillinger, science **209**, 451 (1980).

[157] X.-Z. Li, B. Walker, and A. Michaelides, Proceedings of the National Academy of Sciences **108**, 6369 (2011).

[158] D. M. Wilkins, D. E. Manolopoulos, S. Pipolo, D. Laage, and J. T. Hynes, The Journal of Physical Chemistry Letters **8**, 2602 (2017).

[159] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, Proceedings of the National Academy of Sciences **110**, 15591 (2013).

[160] W. S. Price, H. Ide, Y. Arata, and O. Söderman, The Journal of Physical Chemistry B **104**, 5874 (2000).

[161] E. H. Hardy, A. Zygar, M. D. Zeidler, M. Holz, and F. D. Sacher, The Journal of Chemical Physics **114**, 3174 (2001).

[162] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, in *Intermolecular forces* (Springer, 1981), pp. 331–342.

[163] R. Kumar and J. L. Skinner, The Journal of Physical Chemistry B **112**, 8311 (2008).

[164] J. L. Abascal and C. Vega, The Journal of chemical physics **123**, 234505 (2005).

[165] L. X. Dang and B. M. Pettitt, Journal of physical chemistry **91**, 3349 (1987).

[166] J. Lobaugh and G. A. Voth, The Journal of chemical physics **106**, 2400 (1997).

[167] L. X. Dang and T.-M. Chang, The Journal of chemical physics **106**, 8149 (1997).

[168] L. X. Dang, The Journal of Physical Chemistry B **102**, 620 (1998).

[169] G. S. Fanourgakis and S. S. Xantheas, The Journal of Physical Chemistry A **110**, 4100 (2006).

[170] G. Lamoureux and B. Roux, The Journal of chemical physics **119**, 3025 (2003).

[171] L. Hernández de la Peña and P. Kusalik, The Journal of chemical physics **121**, 5992 (2004).

[172] L. Hernández de la Peña and P. G. Kusalik, The Journal of chemical physics **125**, 054512 (2006).

[173] J. A. Poulsen, G. Nyman, and P. J. Rossky, Proceedings of the National Academy of Sciences **102**, 6709 (2005).

[174] J. L. Abascal and C. Vega, The Journal of chemical physics **123**, 234505 (2005).

[175] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, and J. W. Ponder, Journal of chemical theory and computation **14**, 5273 (2018).

[176] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, Journal of computational chemistry **26**, 1781 (2005).

[177] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, Journal of chemical theory and computation **4**, 435 (2008).

[178] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins: Structure, Function, and Bioinformatics **65**, 712 (2006).

[179] P. A. Kollman, Accounts of Chemical Research **29**, 461 (1996).

[180] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, Proteins: Structure, Function, and Bioinformatics **78**, 1950 (2010).

[181] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Journal of computational chemistry **25**, 1157 (2004).

[182] R. Pastor and A. MacKerell Jr, The journal of physical chemistry letters **2**, 1526 (2011).

[183] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, The Journal of chemical physics **79**, 926 (1983).

[184] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, The Journal of Physical Chemistry B **117**, 9956 (2013).

[185] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, The Journal of chemical physics **120**, 9665 (2004).

[186] M. Rossi, V. Kapil, and M. Ceriotti, The Journal of Chemical Physics **148**, 102301 (2018).

[187] R. L. Benson, G. Trenins, and S. C. Althorpe, Faraday Discussions **221**, 350 (2020).

[188] T. Plé, S. Huppert, F. Finocchi, P. Depondt, and S. Bonella, The Journal of Chemical Physics **155**, 104108 (2021).

[189] E. Mangaud, S. Huppert, T. Plé, P. Depondt, S. Bonella, and F. Finocchi, Journal of chemical theory and computation **15**, 2863 (2019).

[190] J.-L. Barrat and D. Rodney, Journal of Statistical Physics **144**, 679 (2011).

[191] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, The Journal of chemical physics **103**, 4613 (1995).

[192] M. Ceriotti, J. More, and D. E. Manolopoulos, Computer Physics Communications **185**, 1019 (2014).

[193] M. Tuckerman, *Statistical mechanics: theory and molecular simulation* (Oxford university press, 2010).

[194] G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon, The Journal of Chemical Physics **113**, 9140 (2000).

[195] J. A. Morrone and R. Car, Physical review letters **101**, 017801 (2008).

[196] F. Paesani and G. A. Voth, The Journal of Physical Chemistry B **113**, 5702 (2009).

[197] G. Reiter, J. Li, J. Mayers, T. Abdul-Redah, and P. Platzman, Brazilian Journal of Physics **34**, 142 (2004).

[198] G. Romanelli, M. Ceriotti, D. E. Manolopoulos, C. Pantalei, R. Senesi, and C. Andreani, The Journal of Physical Chemistry Letters **4**, 3251 (2013).

[199] T. J. Hele, M. J. Willatt, A. Muolo, and S. C. Althorpe, The Journal of Chemical Physics **142**, 134103 (2015).

[200] J. Beutier, R. Vuilleumier, S. Bonella, and G. Ciccotti, Molecular Physics **113**, 2894 (2015).

[201] T. Plé, S. Huppert, F. Finocchi, P. Depondt, and S. Bonella, The Journal of chemical physics **151**, 114114 (2019).

[202] W. H. Miller, The Journal of Physical Chemistry A **105**, 2942 (2001).

[203] S. K. Reddy, S. C. Straight, P. Bajaj, C. Huy Pham, M. Riera, D. R. Moberg, M. A. Morales, C. Knight, A. W. Götz, and F. Paesani, The Journal of chemical physics **145**, 194504 (2016).

[204] B. Guillot and Y. Guissani, The Journal of chemical physics **108**, 10162 (1998).

[205] M. Hijazi, D. M. Wilkins, and M. Ceriotti, The Journal of Chemical Physics **148**, 184109 (2018).

[206] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr, et al., The journal of physical chemistry B **114**, 2549 (2010).

[207] S. Naseem-Khan, L. Lagardère, C. Narth, G. A. Cisneros, P. Ren, N. Gresh, and J.-P. Piquemal, Journal of Chemical Theory and Computation (2022).

[208] R. E. Duke, O. N. Starovoytov, J.-P. Piquemal, and G. A. Cisneros, Journal of chemical theory and computation **10**, 1361 (2014).

[209] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, Journal of computational chemistry **26**, 1668 (2005).

[210] R. Salomon-Ferrer, D. Case, and R. Walker, *An overview of the amber biomolecular simulation package. wires comput mol sci 3, 198–210* (2013).

[211] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Journal of computational chemistry **30**, 1545 (2009).

[212] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., The Journal of Physical Chemistry B **102**, 3586 (1998).

[213] P. E. Lopes, O. Guvench, and A. D. MacKerell, in *Molecular modeling of proteins* (Springer, 2015), pp. 47–71.

[214] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., Journal of computational chemistry **31**, 671 (2010).

[215] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Journal of the American Chemical Society **118**, 11225 (1996).

[216] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, The Journal of Physical Chemistry B **105**, 6474 (2001).

[217] J. Melcr and J.-P. Piquemal, Frontiers in molecular biosciences **6**, 143 (2019).

[218] S. Patel, A. D. Mackerell Jr, and C. L. Brooks III, Journal of computational chemistry **25**, 1504 (2004).

[219] Z.-X. Wang, W. Zhang, C. Wu, H. Lei, P. Cieplak, and Y. Duan, Journal of computational chemistry **27**, 781 (2006).

[220] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, and T. A. Halgren, Journal of computational chemistry **23**, 1515 (2002).

[221] L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, The Journal of Physical Chemistry B **117**, 9956 (2013).

[222] N. L. Allinger, S. Rodriguez, and K. Chen, Journal of Molecular Structure: THEOCHEM **260**, 161 (1992).

[223] W. Benedict, N. Gailar, and E. K. Plyler, The Journal of Chemical Physics **24**, 1139 (1956).

[224] T. A. Halgren, Journal of the American Chemical Society **114**, 7827 (1992).

[225] A. J. Stone, Journal of Chemical Theory and Computation **1**, 1128 (2005).

[226] J. E. Dennis Jr, D. M. Gay, and R. E. Walsh, ACM Transactions on Mathematical Software (TOMS) **7**, 348 (1981).

[227] B. J. Smith, D. J. Swanton, J. A. Pople, H. F. Schaefer III, and L. Radom, The Journal of chemical physics **92**, 1240 (1990).

[228] D. M. Bates and G. S. Tschumper, The Journal of Physical Chemistry A **113**, 3555 (2009).

[229] S. S. Xantheas and E. Aprà, The Journal of chemical physics **120**, 823 (2004).

[230] S. Bulusu, S. Yoo, E. Apra, S. Xantheas, and X. C. Zeng, The Journal of Physical Chemistry A **110**, 11781 (2006).

[231] G. S. Fanourgakis, E. Apra, and S. S. Xantheas, The Journal of chemical physics **121**, 2655 (2004).

[232] T. H. Dunning Jr, The Journal of chemical physics **90**, 1007 (1989).

[233] R. P. Steele, R. A. DiStasio Jr, Y. Shao, J. Kong, and M. Head-Gordon, The Journal of chemical physics **125**, 074108 (2006).

[234] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, et al., Physical Chemistry Chemical Physics **8**, 3172 (2006).

[235] J. Friedrich, H. Yu, H. R. Leverentz, P. Bai, J. I. Siepmann, and D. G. Truhlar, The journal of physical chemistry letters **5**, 666 (2014).

[236] N. Sahu, S. S. Khire, and S. R. Gadre, Molecular Physics **113**, 2970 (2015).

[237] F. Paesani, S. Iuchi, and G. A. Voth, The Journal of chemical physics **127**, 074506 (2007).

[238] J. R. Lane, Journal of chemical theory and computation **9**, 316 (2013).

[239] A. Soper, Chemical Physics **258**, 121 (2000).

[240] D. Mariedahl, F. Perakis, A. Spah, H. Pathak, K. H. Kim, G. Camisasca, D. Schlesinger, C. Benmore, L. G. M. Pettersson, A. Nilsson, et al., The Journal of Physical Chemistry B **122**, 7616 (2018).

[241] R. Impey, P. Madden, and I. McDonald, Molecular Physics **46**, 513 (1982).

[242] G. R. Medders, V. Babin, and F. Paesani, Journal of chemical theory and computation **10**, 2906 (2014).

[243] E. J. Maginn, R. A. Messerly, D. J. Carlson, D. R. Roe, and J. R. Elliot, Living Journal of Computational Molecular Science **1**, 6324 (2018).

[244] Y. Badyal, M.-L. Saboungi, D. Price, S. Shastri, D. Haeffner, and A. Soper, The Journal of Chemical Physics **112**, 9206 (2000).

[245] A. V. Gubskaya and P. G. Kusalik, The Journal of chemical physics **117**, 5290 (2002).

[246] G. S. Kell, Journal of Physical and Chemical Reference Data **6**, 1109 (1977).

[247] W. Wagner and A. Pruß, Journal of Physical and Chemical Reference Data **31**, 387 (2002).

[248] C. Liu, J.-P. Piquemal, and P. Ren, Journal of chemical theory and computation **15**, 4122 (2019).

[249] C. Liu, J.-P. Piquemal, and P. Ren, The journal of physical chemistry letters **11**, 419 (2019).

[250] P. P. Poier and F. Jensen, Journal of chemical theory and computation **15**, 3093 (2019).

[251] G. A. Cisneros, Journal of Chemical Theory and Computation **8**, 5072 (2012).

[252] J. A. Rackers, R. R. Silva, Z. Wang, and J. W. Ponder, Journal of Chemical Theory and Computation **17**, 7056 (2021).

[253] L. Lagardère, L. El-Khoury, S. Naseem-Khan, F. Aviat, N. Gresh, and J.-P. Piquemal, in *AIP Conference Proceedings* (AIP Publishing LLC, 2017), vol. 1906, p. 030018.

[254] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, Journal of chemical theory and computation **9**, 4046 (2013).

[255] C. Zhang, C. Lu, Z. Jing, C. Wu, J.-P. Piquemal, J. W. Ponder, and P. Ren, Journal of chemical theory and computation **14**, 2084 (2018).

[256] R. A. Corrigan, G. Qi, A. C. Thiel, J. R. Lynn, B. D. Walker, T. L. Casavant, L. Lagardere, J.-P. Piquemal, J. W. Ponder, P. Ren, et al., Journal of chemical theory and computation **17**, 2323 (2021).

[257] P. P. Poier, T. Jaffrelot Inizan, O. Adjoua, L. Lagardère, and J.-P. Piquemal, The Journal of Physical Chemistry Letters **13**, 4381 (2022).