



HAL
open science

Fully quantum dynamics of protonated water clusters

Félix Mouhat

► **To cite this version:**

Félix Mouhat. Fully quantum dynamics of protonated water clusters. Condensed Matter [cond-mat]. Sorbonne Université, 2018. English. NNT : 2018SORUS056 . tel-02137552

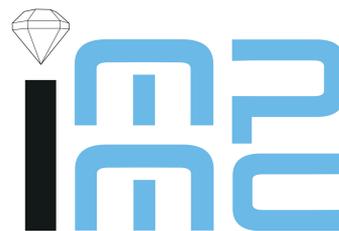
HAL Id: tel-02137552

<https://theses.hal.science/tel-02137552>

Submitted on 23 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Thèse de doctorat

Pour l'obtention du grade de

Docteur de Sorbonne Université

École doctorale de Physique et Chimie des Matériaux - ED 397

Fully quantum dynamics of protonated water clusters

Félix MOUHAT

Directeurs de thèse : Pr. Antonino Marco SAITTA, Dr. Michele CASULA

Soutenue publiquement le 07/09/2018 devant un jury composé de :

Dr. Magali BENOIT (CEMES, Toulouse) Rapportrice
Dr. Nathalie VAST (École Polytechnique, Palaiseau) Rapportrice
Pr. Pascal VIOT (Sorbonne Université, Paris) Président du jury
Pr. Dominik MARX (Ruhr Universität, Bochum) Examineur
Pr. Leonardo GUIDONI (Università degli Studi dell' Aquila) Examineur
Pr. Antonino Marco SAITTA (Sorbonne Université, Paris) Directeur de thèse
Dr. Michele CASULA (Sorbonne Université, Paris) Directeur de thèse
Pr. Rodolphe VUILLEUMIER (ENS, Paris) Membre invité

Remerciements

Tout d'abord, je souhaite remercier l'École Normale Supérieure qui, après m'avoir formé, m'a fourni une bourse de thèse me permettant de mener à bien les travaux présentés dans ce manuscrit.

Je remercie également Magali Benoit et Nathalie Vast d'avoir accepté de rapporter ce manuscrit et pour leurs remarques très pertinentes afin de l'améliorer. Mes remerciements s'adressent aussi à Pascal Viot, que j'ai eu le plaisir de connaître quand j'ai enseigné dans son équipe et qui m'a fait le plaisir de présider mon jury de thèse. Enfin, j'ai pu apprécier la présence d'experts en la personne de Leonardo Guidoni et Dominik Marx pour compléter mon jury de thèse et ai fortement apprécié nos échanges scientifiques lors de ma soutenance.

Cette formidable aventure scientifique longue de presque 4 ans n'aurait pas été possible sans le soutien permanent de mes encadrants. Commençons par Marco, dont le rôle dépasse celui du directeur de thèse puisque c'est désormais un ami. Je le remercie pour son incroyable capacité à trouver les mots justes, y compris dans les moments délicats, et l'incite à garder son formidable enthousiasme qui le caractérise. Je nous souhaite de nombreuses discussions et bières à venir ! Je tiens également à exprimer toute ma reconnaissance pour mon co-directeur Michele, pour son investissement exceptionnel et quasiment quotidien. Par ses connaissances techniques et scientifiques, il a indéniablement été le moteur principal de ces travaux de thèse. Quant à Rodolphe, nos chemins se sont croisés dès mon arrivée à l'ENS puisqu'il a été mon tuteur pédagogique. Il a été la personne qui m'a donné envie de poursuivre dans la simulation numérique et je ne le remercierai jamais assez de m'avoir guidé de manière pertinente dans mes choix, tout en veillant sur moi.

Durant ma thèse, j'ai rencontré un certain nombre de personnes dans le sympathique laboratoire de l'IMPMC et il serait impossible pour moi de toutes les lister sans en oublier. Je voudrais remercier en particulier Nicolas, alias *l'homme de la pause*, qui fut ma première vraie rencontre dans le laboratoire, avec qui j'ai passé de formidables moments dans et en dehors de l'IMPMC. Je remercie également Natan, un camarade de l'ENS qui a eu le bon goût de me rejoindre à l'IMPMC pour les vortex (= pauses de durée indéterminée, une classification existe même pour les curieux) et les pizzas/kebabs/sushis partagés ensemble. Je retiens également un excellent souvenir de Mario avec qui j'ai passé d'agréables moments en conférence dans le Colorado et à Marseille. J'ai également sincèrement apprécié les discussions que j'ai pu avoir avec mes co-bureaux : Guilherme, Abderrezak et plus récemment Antoine. Je remercie également les autres membres de mon équipe, avec qui il était toujours sympathique d'échanger avant, pendant ou après nos "group meeting" (que j'affectionnais particulièrement...) : Fabio, Giuseppe, Sarah, Adrien, Matthieu, Andrea, Gabriele et Théo (la relève !).

J'en profite pour également remercier les copains que j'ai pu me faire durant mes études à l'ENS et avec qui je continue à avoir de riches discussions scientifiques ou plus informelles. Commençons par Thibault, qui fut ma toute première rencontre à l'ENS et avec qui j'ai partagé de mémorables moments, donnant parfois lieu à l'écriture de pactes douteux. Nos

quelques milliers de pokes échangés durant notre scolarité sont une preuve indéniable de notre constante volonté de troller l'autre à tout moment de la journée ou de la nuit ! Je remercie également Lucas, notre troisième compère de nos classiques soirées Porto-Muscat avec Thibault, pour sa capacité à collecter des informations croustillantes sur toute notre promo, faisant de lui une concierge de compétition. Je remercie également Pauline pour les agréables repas et soirées jeux que nous avons partagés, et lui souhaite une bonne aventure à l'IMPIC qu'elle a rejoint récemment. En dehors de la promotion de Chimie, je souhaite remercier Jérôme avec qui j'ai affronté ma deuxième année de prépa, qui est toujours une source d'inspiration pour moi et qui m'a présenté Guillaume, mon seul camarade mathématicien avec qui nous avons testé la quasi totalité des restaurants de la rue Mouffetard. Je remercie également Julien et David, que je considère un peu comme mes aînés scientifiques puisqu'ils ont également fait de la simulation numérique, et que je retrouve parfois pour boire des bières ou faire des parties endiablées de Worms.

J'ai également pu rencontrer de nombreuses autres personnes en dehors du laboratoire lors de mes diverses activités à Paris, à savoir l'équitation et le bridge. Commençons donc par le côté sportif avec le cheval à l'École militaire où j'ai eu le plaisir de rencontrer : Marie-Valentine, Sylvain, Paul-Emile, Sonia, Nicolas, Eugénie, Soleine... J'espère que nous continuerons à partager nos traditionnels repas même si on n'est plus dans la même reprise, chacun ayant pris des chemins différents. Pour ce qui est du bridge, activité plutôt intellectuelle (non non la thèse ne me suffisait pas), là aussi les rencontres et bonnes surprises furent nombreuses : merci à Quentin, Michel et Marlène de me montrer gentiment la voie vers le bridge de haut niveau, à Thomas, Axel et Catherine, mes (ex)-partenaires de compétition préférés, à Baptiste, Constance, Camille, Ludovic, Arnaud, Julien, Olivier, Louis et autres jeunes que j'ai plaisir à voir le dimanche soir en général. Une mention spéciale à Claudine et Bernard pour leur gentillesse et leur soutien lors de mon premier festival à Biarritz. Merci aussi à Roupoil pour son humour qui m'a donné envie de le rencontrer et d'échanger régulièrement à propos de nos derniers craquages à l'enchère ou en flanc !

Jusque là, j'ai beaucoup mentionné ma vie parisienne mais je n'oublie évidemment pas mes amis de longue date. Par exemple, je sais que j'aurai toujours plaisir à retrouver Guillaume et Etienne, deux amis d'enfance que j'ai rencontrés à Montpellier où j'étais suivi pour ma surdité. J'en profite également pour remercier Adrienne et Christiane mes orthophonistes, ainsi que Jean-Pierre mon régleur, sans qui je n'aurais certainement pas pu mener mes études jusqu'à la rédaction de ce présent manuscrit. Toujours dans les amis d'enfance, je voudrais mentionner Maude avec qui j'ai toujours plaisir à échanger de tout et de rien autour d'un bon mojito, avec ou sans Colin. Enfin, il m'est impossible de ne pas citer mes amis de lycée, compagnons absolument incontournables pour mener à bien des séminaires intellectuels divers et variés (impliquant, il est vrai, une certaine quantité d'alcool et de nourriture diététique à souhait). Nous continuons d'ailleurs à partir en vacances chaque année, et j'espère que nous poursuivrons, au moins dans un futur proche, cette tradition. Merci à : Laurent, Pitou, Jiss, François, Florent, Justine, Nilou, Pauline et Apolline. À ce noyau dur s'ajoutent d'autres amis du lycée avec qui j'ai toujours plaisir à échanger : Pierre-Edouard, toujours passionné de Roland-Garros qui me fait des "live" quand je suis trop pris au laboratoire, Pierre parti à Panama, Laura, Hocine, Clément et bien d'autres que j'aurai toujours autant de plaisir à

recroiser vers Montbéliard.

Pour finir ces longs remerciements, il est naturel que je parle de ma famille, ô combien précieuse à mes yeux. Tout d'abord, merci à mes parents Thierry et Isabelle pour leur formidable soutien dès mon plus jeune âge. Ils ont su à chaque fois me pousser avec bienveillance dans mes retranchements pour que je donne le meilleur de moi-même et je ne leur serai jamais assez reconnaissant pour cela. Merci aussi à mon frère Basile, docteur avant moi, avec qui j'ai adoré vivre pendant 6 semaines durant cette thèse, alternant les moments de longues discussions, de sport et de franche rigolade. Merci aussi à Mathilde, ma petite soeur, qui sera aussi docteur un jour et qui a véritablement pris son envol récemment. Merci à mes grands-parents Pierre et Marie-Thé, et à ma grand-mère Micheline pour leurs ponctuelles attentions. Merci à ma belle famille Jean-Noël, Martine, Lou et Apolline qui m'ont accueilli chaleureusement chez eux pour me reposer pendant l'été.

"Bon Hubert, je crois qu'on a fait le tour" comme on pourrait entendre dans OSS 117, film à l'humour délicat et raffiné dont mes amis savent que je suis friand me paraît approprié pour conclure cette section dédiée aux remerciements qui, je l'espère, vous aura plu.

Contents

List of Figures	v
List of Tables	xi
List of Abbreviations	xiii
Introduction	1
I Review of electronic structure and Molecular Dynamics methods	5
1 Electronic structure calculations	7
1.1 Deterministic methods	8
1.1.1 Density Functional Theory	9
1.1.2 Quantum chemistry methods	15
1.2 Stochastic methods	21
1.2.1 The Quantum Monte Carlo approach	22
1.2.2 The Quantum Monte Carlo wave function	23
1.2.3 Variational Monte Carlo	26
1.2.4 Diffusion Monte Carlo and Lattice Regularized Diffusion Monte Carlo	29
1.2.5 Forces evaluation in Quantum Monte Carlo	34
2 Ion dynamics	37
2.1 Molecular Dynamics simulations	38
2.1.1 Time and ensemble averages	38
2.1.2 Equations of motion and Liouvillian formalism	40
2.1.3 The velocity-Verlet algorithm	41
2.1.4 Analytical force fields for water	42
2.2 Finite temperature simulations	43
2.2.1 The canonical ensemble	44
2.2.2 The Langevin thermostat	44
2.2.3 Classical Langevin Dynamics integration schemes	46
2.2.4 Validation of the Langevin thermostat	51
2.3 Dealing with quantum nuclei	52
2.3.1 Zero Point Energy and Nuclear Quantum Effects	53
2.3.2 The Path Integral approach	54
2.3.3 Path integral Langevin Dynamics	58

II	Methodological developments and testing	61
3	Development of fully quantum dynamics	63
3.1	Derivation of novel Langevin Dynamics integrators	64
3.1.1	The Classical Momentum-Position Correlator (CMPC) algorithm	64
3.1.2	Path Integral Momentum-Position Correlator (PIMPC) algorithm	66
3.1.3	Path Integral Ornstein-Uhlenbeck Dynamics (PIOUD)	69
3.1.4	Algorithm stability with deterministic forces	71
3.2	Extension to the stochastic case: correlating the noise by Quantum Monte Carlo	79
3.2.1	Evolution of the Quantum Monte Carlo wave function	79
3.2.2	Bead-grouping approximation	80
3.2.3	Quantum Monte Carlo ionic forces and noise correction	81
3.2.4	Algorithm stability with QMC forces	82
4	The Zundel ion: a benchmark system	87
4.1	Zero temperature results	88
4.1.1	Potential energy landscape	88
4.1.2	Equilibrium geometries	90
4.1.3	Accuracy of the Quantum Monte Carlo approach	95
4.2	Benchmark calculations	98
4.2.1	Validation of the classical dynamics	99
4.2.2	Validation of the quantum dynamics	100
4.3	Proton transfer in the Zundel ion	104
III	On the way to simulate fully quantum water	109
5	Interlude: the water dimer	111
5.1	Non-bonding water dimer	112
5.2	Bonding water dimer	115
5.2.1	Binding energy	115
5.2.2	Geometric properties	119
6	The protonated water hexamer	123
6.1	Zero temperature results	125
6.1.1	Potential energy landscape	125
6.1.2	Equilibrium geometries	129
6.1.3	Proton transfer static barriers	133
6.2	Classical hexamer dynamics	135
6.2.1	Radial distribution functions	136
6.2.2	Proton displacement	139
6.3	Quantum hexamer dynamics	144
6.3.1	Radial distribution functions	144
6.3.2	Proton displacement	147
6.3.3	Thermal effects versus Nuclear Quantum Effects	154

Contents

6.4	Proton transfer in the protonated hexamer	157
6.4.1	Classical counting	157
6.4.2	Quantum counting: the instanton theory	160
General conclusions and Perspectives		163
Appendix		167
A Derivation of the MP2 energy		169
B Calculation of excitation amplitudes within CC approach		171
C Wave function optimization methods		173
C.1	Stochastic Reconfiguration	173
C.2	Stochastic Reconfiguration with Hessian accelerator	174
D Noise correction and quantum integration scheme		177
D.1	Noise correlators in the CMPC algorithm	177
D.2	Quantum Monte Carlo noise correction in the CMPC algorithm	177
D.3	Quantum integration scheme	178
E Protonated hexamer geometries		181
E.1	Minimum geometry	181
E.2	Other geometries	181
F Protonated hexamer fitting parameters		185
F.1	Classical fit: the 2-species model	185
F.2	Quantum fit: the 3-species model	185
Bibliography		187
Résumé		205
Abstract		205

List of Figures

2.1	Histograms of the probability distribution to find the Zundel cation at the instantaneous temperature $T(t_i)$, imposing a constant target temperature $T_{\text{target}} = 300$ K for various thermostats. The analytic distribution $P(T T_{\text{target}})$ (see Eq. (2.46)) is represented in green.	52
2.2	Intuitive representation of the classical ring polymer as a function of the simulation temperature T	56
3.1	Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the number of quantum replicas P evolving at $T = 50$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. The time step and the friction are respectively set to $\delta t = 0.5$ fs and $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).	74
3.2	Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the number of quantum replicas P evolving at $T = 300$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. The time step and the friction are respectively set to $\delta t = 0.5$ fs and $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).	75
3.3	Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the time step δt at $T = 50$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. $P = 128$ quantum replicas are used and the friction is set to $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).	77

3.4	Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the time step δt at $T = 300$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. $P = 32$ quantum replicas are used and the friction is set to $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).	78
3.5	PIOUD evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ (top panel), the temperature T (middle panel) and the potential autocorrelation time τ_V (bottom panel) as a function of the input friction γ_0 . Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. Deterministic forces are represented in black whereas the noisy QMC forces are in red. The time step and the number of quantum replicas are respectively set to $\delta t = 1$ fs and $P = 32$	83
3.6	PIOUD evolution of the quantum kinetic energy estimators $\langle T_{\text{vir/pri}} \rangle$ as a function of the time step δt (top panel) and the number of quantum replicas P (bottom panel). Colors and symbols are the same as in Figure 3.5. The input friction was set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u. fs. The default values of the time step and the number of quantum replicas are respectively $\delta t = 1$ fs and $P = 32$	85
4.1	QMC optimized geometries for global C_2 minimum (left) and for C_s -Inv local minimum (right) of the Zundel ion ^[72] . Atom labels for the analysis of the Zundel properties are also indicated.	88
4.2	Potential energy curve (kcal/mol) of the Zundel ion projected on the central $\overline{\text{OO}}$ distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory, except CCSD(T). Each curve has its minimum as reference point.	89
4.3	Separations (\AA) between the two central oxygen atoms and the excess proton as a function of the reaction coordinate $d_{\text{O}_1\text{O}_2}$ for different computational methods. Vertical dashed lines indicate the equilibrium $d_{\text{O}_1\text{O}_2}$ for each method.	92
4.4	Excess proton (left) and average (right) gradients calculated for the Zundel ion at the CCSD (top) or CCSD(T) (bottom) level of theory for various oxygen-oxygen distances $d_{\text{O}_1\text{O}_2}$. Results are obtained with the TZ basis set.	96
4.5	Excess proton (left) and average (right) gradients calculated for the Zundel ion at the CCSD/CCSD(T) levels of theory with various basis sets. For the two tested geometries, the oxygen-oxygen distance is $d_{\text{O}_1\text{O}_2} = 2.65$ \AA	97
4.6	Intuitive representation of the QMC-driven (Path Integral) Langevin Dynamics developed in this thesis and applied to the Zundel ion. The electronic configurations adopted by the QMC walkers are depicted by the blue points, while the oxygen and the hydrogen atoms are represented by red and white spheres, respectively.	98

List of Figures

4.7	Normalized oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions obtained by CMPC-LD simulations at low temperature $T = 50$ K (left) and $T = 300$ K (right). The black curves represent the distributions obtained with analytic CCSD(T) forces, whereas the red curves correspond to VMC-LD dynamics. The friction is set to $\gamma_{\text{BO}} = 1.46 \times 10^{-3}$ a.u.	101
4.8	Normalized oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions obtained by PIOUD-LD simulations at low temperature $T = 50$ K (left) and $T = 300$ K (right). The black curves represent the distributions obtained with analytic CCSD(T) forces, whereas the red and blue curves correspond to VMC-PILD dynamics with $N_{\text{groups}} = 1$ and $N_{\text{groups}} = 16$, respectively. Quantum simulations are performed with $P = 128$ beads at $T = 50$ K, and $P = 32$ beads at $T = 300$ K. The friction is set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u.	103
4.9	Evolution of the oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions as a function of the temperature for the classical (left) or the quantum (right) Zundel ion.	106
4.10	Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC- or CCSD(T)- driven CMPC-LD (left) and PIOUD (right) simulations at $T = 50$ K (top), $T = 300$ K (middle) and $T = 900$ K (bottom). The black circles correspond to the equilibrium geometries of the Zundel ion at zero temperature obtained by CCSD(T) calculations ^[72] . Quantum simulations are performed with $P = 128$ beads at $T = 50$ K, $P = 32$ beads at $T = 300$ K and $P = 4$ beads at $T = 900$ K. The friction is set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u.	107
5.1	Different configurations of the water dimer $(\text{H}_2\text{O})_2$. (a) non-bonding water dimer. (b) Bonding water dimer.	112
5.2	Binding energy (kcal/mol) of the non-bonding water dimer $(\text{H}_2\text{O})_2$ as a function of the \overline{OO} distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory.	114
5.3	Binding energy (kcal/mol) of the bonding water dimer $(\text{H}_2\text{O})_2$ as a function of the \overline{OO} distance. Comparison between different computational methods. Left panel: DFT results. Right panel: QMC results. Structural relaxation is performed at each level of theory, except for the CCSD(T) and the LRDMC approaches. The LRDMC lattice parameter is set to $a = 0.125$	116
5.4	Binding energy (kcal/mol) of the bonding water dimer $(\text{H}_2\text{O})_2$ as a function of the \overline{OO} distance. Comparison between different Jastrow contraction schemes of the QMC WF. Structural relaxation is performed for each employed basis set.	117
6.1	Different configurations of the protonated water hexamer $\text{H}_{13}\text{O}_6^+$. (a) Zundel-like configuration with a Zundel center (H_5O_2^+ , green dashed line) and its first solvation shell (4 H_2O). (b) Eigen configuration with an Eigen cation (H_9O_4^+ , blue dashed line) on the left accompanied by two solvating water molecules (2 H_2O) on the right.	124
6.2	Atom labels employed for the analysis of the protonated water hexamer properties. This cation is composed of a Zundel-like core with 4 surrounding water molecules in the solvation shell.	126

6.3	Potential energy curve (kcal/mol) of the protonated water hexamer projected on the central \overline{OO} distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory, except for CCSD/CCSD(T). Each curve has its minimum as reference point.	127
6.4	Potential energy curve (kcal/mol) of the protonated water hexamer projected on the central \overline{OO} distance. Comparison between different computational methods for various relaxed geometries and basis sets. Each curve has its minimum as reference point.	128
6.5	Separations (\AA) between the two central oxygen atoms and the excess proton as a function of the reaction coordinate $d_{\text{O}_1\text{O}_2}$ for different computational methods. Vertical dashed lines indicate the equilibrium $d_{\text{O}_1\text{O}_2}$ for each method.	130
6.6	Oxygen-oxygen RDFs obtained by QMC-driven CMPC-LD simulations at various temperatures ranging from 200 K to 400 K. The vertical lines indicate the $\langle d_{\text{O}_1\text{O}_2} \rangle$ for each simulation.	138
6.7	Oxygen-proton RDFs obtained by QMC-driven CMPC-LD simulations at various temperatures ranging from 200 K to 400 K.	138
6.8	Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven CMPC-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high temperature $T = 400$ K (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.	140
6.9	Distributions of the proton displacement d with respect to the midpoint of the oxygen-oxygen distance (left panel) and the proton sharing coordinate δ (right panel), as a function of the temperature, for classical particles.	142
6.10	Comparison of the $P(d)$ distributions obtained from QMC-driven CMPC-LD simulations (blue) and fitted using Eq. (6.3) for $T = 200, 250, 300, 350, 400$ K.	143
6.11	Oxygen-oxygen RDFs obtained by QMC-driven PIOUD-LD simulations at various temperatures ranging from 200 K to 400 K. The vertical lines indicate the $\langle d_{\text{O}_1\text{O}_2} \rangle$ for each simulation.	145
6.12	Oxygen-proton RDFs obtained by QMC-driven PIOUD-LD simulations at various temperatures ranging from 200 K to 400 K.	146
6.13	Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven PIOUD-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high-temperature (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.	148
6.14	Centroid bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven PIOUD-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high temperature $T = 400$ K (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.	149
6.15	Left panels: quantum distributions of the proton displacement d with respect to the midpoint of the oxygen-oxygen distance with (bottom) or without (top) running averages. Right panels: quantum (top) and centroid (bottom) distributions of the proton sharing coordinate δ	151

List of Figures

6.16	Comparison of the $Q(d)$ distributions obtained from QMC-driven PIOUS-LD simulations (blue) and fitted using Eq. (6.5) for $T = 200, 250, 300, 350, 400$ K.	153
6.17	Difference between bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven LD simulations for classical (left panels), centroid (central panels) and quantum (right panels) particles. Upper panels: $\rho_{2D}(T = 300 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$. Middle panels: $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 300 \text{ K})$. Bottom panels: $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$. The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer. . .	156
6.18	Time evolution of the proton displacement and the electrostatic potential difference in the classical protonated water hexamer at room temperature ($T = 300$ K).	159
6.19	Time evolution of the proportion of the ring-polymer located on the left of $\frac{d_{O_1O_2}}{2}$ and the electrostatic potential difference in the quantum protonated water hexamer at room temperature ($T = 300$ K).	161
6.20	Evolution of the classical k_{PT} and the quantum k_{PT}^Q PT rate of the protonated hexamer as a function of the temperature.	162
6.21	Comparison of the protonated water dimer and hexamer Potential Energy Surface (left) and geometry (right) as a function of the (central) oxygen-oxygen distance $d_{O_1O_2}$. Vertical dashed lines indicate the corresponding equilibrium $d_{O_1O_2}$.	166

List of Tables

4.1	Geometric properties (distances in Å) of the C_2 -symmetry minimum of the protonated water dimer, comparison between different computational methods. . .	93
4.2	Geometric properties (angles in °) of the C_2 -symmetry minimum of the protonated water dimer, comparison between different computational methods. . . .	94
5.1	Water dimer, water monomer and binding energies for various QMC WF. The number of parameters p of the WF employed for the calculations is also indicated.	118
5.2	Geometric properties (distances in Å) of the bonding water dimer minimum, comparison between different computational methods. The * indicates that we kept the distance $\overline{O_1O_2}$ constant during the structural optimization.	120
5.3	Geometric properties (angles in °) of the bonding water dimer minimum, comparison between different computational methods.	120
6.1	Geometric properties (distances in Å) of the core of the protonated water hexamer minimum, comparison between different computational methods.	132
6.2	Geometric properties (angles in °) of the core of the protonated water hexamer minimum, comparison between different computational methods.	132
6.3	Geometric properties (distances in Å) of the solvation shell of the protonated water hexamer minimum, comparison between different computational methods.	133
6.4	Static symmetrization barriers (in Kelvin units) of the protonated water hexamer at different $d_{O_1O_2}$ for various computational methods. Results are computed as energy differences between the true minimum energy configuration and the symmetrized one.	134
6.5	Species proportions obtained by fitting distributions of the proton displacement with the 3-gaussian model of Eq. (6.3) in the classical particles case.	144
6.6	Estimated species proportions obtained by fitting distributions of the proton displacement with the 5-gaussian model in the quantum particles case.	154
6.7	Estimated number of proton jumps and its corresponding time constant τ_{jump} as a function of the temperature, for the classical protonated hexamer	159
6.8	Estimated number of proton jumps and its corresponding time constant τ_{jump} as a function of the temperature, for the quantum protonated hexamer	162
F.1	Fitting parameters obtained by fitting distributions of the proton displacement with the 3-gaussian model of Eq. (6.3) in the classical particles case.	185
F.2	Fitting parameters obtained by fitting distributions of the proton displacement with the 5-gaussian model of Eq. (6.5) in the quantum particles case.	185

List of Abbreviations

Abbreviation	Meaning
AD	Algorithmic Differentiation
BLYP	Becke-Lee-Yang-Parr
BO	Born-Oppenheimer
CBS	Complete Basis Set
CCSD/CCSD(T)	Coupled-Cluster with Single,Double (and Triple excitations
CMPC	Classical Momentum-Position Correlator
DFT	Density Functional Theory
DMC	Diffusion Monte Carlo
FN	Fixed Node
GGA	Generalized Gradient Approximation
GLE	Generalized Langevin Equation
HF	Hartree-Fock
Hell-Fey	Hellmann-Feynman
HPC	High Performance Computing
IR	InfraRed
JAGP	Jastrow Antisymmetrized Geminal Power
JSD	Jastrow Slater Determinant
KS	Kohn Sham
LDA	Local Density Approximation
LD	Langevin Dynamics
LRDMC	Lattice Regularized Diffusion Monte Carlo
MC	Monte Carlo
MD	Molecular Dynamics
MP2	Second-order Møller Plesset perturbation theory
NQE	Nuclear Quantum Effects
OUP	Ornstein-Uhlenbeck Process
PBE	Perdew-Burke-Ernzerhof
PCF	Pair Correlation Function
PES	Potential Energy Surface
PI	Path Integral
PILD	Path Integral Langevin Dynamics
PIOUD	Path Integral Ornstein-Uhlenbeck Dynamics
QTB	Quantm Thermal Bath
RDF	Radial Distribution Function
QMC	Quantum Monte Carlo
RVB	Resonating Valence Bond
SCAN	Strongly Constrained and Appropriately Normed
SD	Slater Determinant
SDE	Stochastic Differential Equation
SPC	Single Point Charge

SR	S tochastic R econfiguration
SRH	S tochastic R econfiguration with H essian accelator
SVR	S tochastic V elocity R escaling
VdW	V an d er W aals
VMC	V ariational M onte C arlo
WF	W ave F unction

Introduction

WATER, the liquid of life, displays countless remarkable properties. Among them, we can cite: a large polarity, anomalously high melting and boiling temperatures, a liquid phase denser than its solid one, characterized by a negative slope of the solid-liquid boundary region in the (P,T) phase diagram, and a large dielectric constant. Such peculiar features are due to the presence of a strong and well-structured H-bond network which significantly stabilizes the liquid water^[1]. Exploring the very rich phase diagram of water and providing an accurate description of each ice or liquid state is indeed a very active field of research^[2-6].

One of the most noticeable property is the very high proton mobility, which is approximately five times the one of ions of similar size as H_3O^+ ^[7]. Understanding the mechanisms by which hydronium H_3O^+ and hydroxyde HO^- ions are transported through water has represented both an experimental and a theoretical challenge for more than two centuries. The first explanation ever attempted was given in 1806 by von Grotthus^[8], who described the proton diffusion phenomenon in terms of a large-scale relay in the hydrogen-bonded water network. Thanks to the work of Eigen^[9] on the one hand and Zundel^[10] on the other hand, it has been suggested that hydrated protons have two preferred structures in water. The first one, the *Eigen ion*, tends to localize the excess proton on a single molecule, solvated by three surrounding water molecules forming the $\text{H}_3\text{O}^+(\text{H}_2\text{O})_3$ or H_9O_4^+ complex. The *Zundel complex* H_5O_2^+ or $[\text{H}_2\text{O} - \text{H} - \text{H}_2\text{O}]^+$ is an entity where the excess proton is delocalized and evenly shared between two side water molecules via a H-bond. These complexes are postulated to be mutually exclusive: the presence of one of them rules out the presence of the other, and they constitute long wires within the Grotthus approach.

Such a relay mechanism between donor and acceptor species is very difficult to probe both theoretically and experimentally. Indeed, the typical timescale for aqueous proton transport under ambient conditions is estimated to be about 1 picosecond^[11], suggesting that the involved activation barriers are tiny. Consequently, despite significant progress^[12-15], a quantitative description of this problem is not at reach yet, as shown for example by the spread of computed H^+ mobilities using Density Functional Theory-based *ab initio* Molecular Dynamics simulations (AIMD)^[16]. Probing proton transfer (PT) processes also constitutes an experimental problem, but recent advances^[17] challenge some theoretical descriptions. The complexity of describing properly PT in water stems from the variety of ingredients implied in such a process: the aqueous environment plays a key role, the involved H-bonds are very strong, *i.e.* they have the same order of magnitude as covalent bonds, and due to the light mass of the proton, Nuclear Quantum Effects (NQE) cannot be neglected. Going beyond DFT is therefore necessary to provide a reliable microscopic description of PT in aqueous system. The use of most advanced quantum chemistry methods that could address this issue is however limited to relatively small systems, due to the poor scaling of these techniques with the system size. The Quantum Monte Carlo (QMC) approach is thus a promising candidate thanks to both its remarkable accuracy and mild scaling with the number of electrons^[18]. Incorporating this method into a MD framework is nevertheless far from being trivial, as the

corresponding computed ionic forces are noisy^[19,20], due to the QMC stochastic nature^[21].

In this thesis, our aim is twofold. First, we want to propose a novel approach, capable to couple the highly accurate QMC description of the electronic properties with the fully quantum treatment of ions (nuclei) via a Path Integral Langevin Dynamics. To address the intrinsic noise affecting the QMC forces, specific algorithms must be devised. Second, we will apply this very general methodology to study H^+ defects in water clusters whose description is more attainable by high-level methods, while keeping the main physical ingredients present in liquid water. Indeed, the protonated water hexamer $\text{H}_{13}\text{O}_6^+$ is the simplest realistic system to model PT in water as its IR spectrum shows that the Zundel and Eigen motifs coexist in this cluster^[22,23].

This thesis is organized as follows. In Chapter 1, we provide the necessary theoretical background to perform very accurate electronic structure calculations, starting from the resolution of the Schrödinger equation. The use of *ab initio* calculations able to reach the chemical accuracy - defined by a few kcal/mol - is indeed necessary to provide a reliable description of PT in water and aqueous systems. The electronic correlation is treated either by deterministic techniques such as the Density Functional Theory (DFT) and quantum chemistry methods, namely the Coupled Cluster (CC) and the Møller-Plesset (MP2) approaches, or within a stochastic framework. The latter includes the QMC approach, used throughout this thesis, as it provides highly accurate results, with a mild scaling with the number of electrons of the system.

Chapter 2 is devoted to the introduction of the Molecular Dynamics (MD) method, which is frequently employed in computer simulations to characterize the properties of any physical or chemical system in gas phase or in the bulk liquid. Since we aim at studying the role of the temperature on the proton mobility in neutral or charged water clusters, using MD simulations with noisy stochastic forces, we incorporate them into a Langevin Dynamics framework. Another issue is related to the presence of significant NQE in water which deeply affect the Potential Energy Surface (PES) on which the ions move. The Feynman Path Integral (PI) approach is therefore introduced. This formalism can be recombined with the Langevin approach (PILD method), and we detail some reference algorithms to integrate the corresponding equations of motion.

In Chapter 3, we derive three novel algorithms (CMPC, PIMPC and PIOUS) to perform LD or PILD simulations of any system with noisy QMC forces. A particular attention is devoted to control the simulation temperature, as it could in principle be biased by the intrinsic QMC noise. We also introduce an original method to make the calculations with quantum nuclei, but feasible with a computational burden comparable to the classical case. The designed integrators are then tested and validated on both analytical (*i.e.* deterministic) and stochastic force fields for a benchmark system: the protonated water dimer (H_5O_2^+). The properties of this smallest water cluster are extensively studied in Chapter 4, by means of a fully quantum dynamics at finite temperature. This is, to the best of our knowledge, the very first time that such a level of accuracy is reached, on both the electronic structure and the quantum description of ion dynamics. Despite the usefulness of the Zundel cation to validate

Introduction

novel methodologies, our understanding of this system, as far as the PT process is concerned, is very limited, as the role of the solvent is not taken into account. In that perspective, we plan to study in Chapter 6 the larger protonated water hexamer, *i.e.* the solvated Zundel complex.

Before studying this cluster, Chapter 5 is devoted to the development of a QMC wave function (WF) tested on the non-bonding and bonding water dimer. We show that our QMC WF is accurate enough to recover a significant fraction of the water dimer binding energy, and more generally to capture the Van der Waals (VdW) dispersive interactions in water at a moderate computational cost.

Thanks to the tools developed in Chapter 3 and benchmarked in Chapters 4 and 5 on the Zundel ion and the water dimer, respectively, we are firmly equipped to tackle the microscopic mechanisms driving PT processes in a more realistic situation. In Chapter 6, we investigate at an unprecedented level of accuracy the interplay between thermal and nuclear quantum effects that determines the excess proton mobility inside the protonated water hexamer. In particular, we confirm the idea that NQE play a key role into the remarkably high proton diffusion in water. Moreover, we discover that the evolution of the PT rate with temperature is highly non trivial, as the excess proton mobility and the structural properties of the protonated hexamer show a non-monotonic dependence with temperature. Proton transfer is particularly efficient around ambient conditions, suggesting the presence of a sweet PT spot at room temperature, *i.e.* precisely where it could be the most relevant for biology. To evaluate the PT rate in the protonated water hexamer as a function of temperature, we introduce an original method based on the instanton theory.

We believe that all the aforementioned developments constitute a huge step forward in the understanding of the fundamental mechanisms underlying PT processes in water and in aqueous systems. Indeed, we are now able, by studying such reduced size clusters, to dissect each molecular or thermal contribution to the overall physics of PT at any temperature, particularly around ambient conditions, relevant for most chemical or biological systems.

Part I

Review of electronic structure and Molecular Dynamics methods

Electronic structure calculations

Contents

1.1	Deterministic methods	8
1.1.1	Density Functional Theory	9
1.1.2	Quantum chemistry methods	15
1.1.2.1	Second-order Møller-Plesset perturbation theory	15
1.1.2.2	Coupled-Cluster methods: CCSD and CCSD(T)	19
1.2	Stochastic methods	21
1.2.1	The Quantum Monte Carlo approach	22
1.2.2	The Quantum Monte Carlo wave function	23
1.2.2.1	Antisymmetrized Geminal Power (AGP) part	23
1.2.2.2	The Jastrow factor	25
1.2.3	Variational Monte Carlo	26
1.2.4	Diffusion Monte Carlo and Lattice Regularized Diffusion Monte Carlo	29
1.2.4.1	Diffusion Monte Carlo	29
1.2.4.2	Lattice Regularized Diffusion Monte Carlo	32
1.2.5	Forces evaluation in Quantum Monte Carlo	34

PROTON TRANSFER (PT) is an ubiquitous phenomenon occurring in strong and well-structured hydrogen bond networks. The energies implied in PT and proton diffusion are very small with barriers lower than the chemical accuracy, hereby defined as 1 kcal/mol. Consequently, it requires the use of advanced electronic structure calculations which will be reviewed in this Chapter. Indeed, to reach our goal to provide a reliable and quantitative description of proton dynamics in water clusters, it is necessary to explicitly take into account the electronic correlation. This quantity is defined as the difference between the exact energy of the system and the Hartree-Fock (HF) energy obtained in the complete basis set (CBS) limit

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}^{\infty}. \quad (1.1)$$

In the HF approach, the electrons are treated as if they moved independently from each other in a mean field generated by the other electrons: their motion is uncorrelated. This approach is far from being true and it becomes necessary to include the electronic correlation between electrons. First, each electron moves in a way so that it avoids locations in a close proximity to the instantaneous positions of all other electrons. This type of correlation is called *dynamic* correlation since it is directly related to electron dynamics.

The second component of E_{corr} arises from the construction of the HF wave function (WF) which is simply a Slater determinant (SD). That representation can be very poor to describe properly a many-body state since some electronic states can only be described using a linear

combination of nearly degenerated SD. The corresponding type of correlation is called *static* correlation since it is not related to electron dynamics.

To include electronic correlation in our study, it becomes necessary to go beyond the HF approximation. In this Chapter, we will review the different methods used to determine accurately the electronic structure of the neutral or protonated water clusters studied during this thesis. These methods will be incorporated into a Molecular Dynamics framework (MD) discussed in the Chapter 2 to describe the PT in these clusters.

The common purpose of all electronic structure methods is to find approximate solutions to the many-body Schrödinger equation:

$$\hat{H}|\Psi\rangle = E|\Psi\rangle. \quad (1.2)$$

In this thesis, we will always assume the Born-Oppenheimer (BO) approximation to hold. The Hamiltonian operator \hat{H} of a system containing N electrons and N_{at} nuclei reads as

$$\begin{aligned} \hat{H} &= -\frac{1}{2} \sum_{i=1}^N \Delta_{\mathbf{r}_i} + \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i}^N \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{a=1}^{N_{\text{at}}} \sum_i^N \frac{Z_a}{|\mathbf{q}_a - \mathbf{r}_i|} \\ &= \hat{K} + \hat{V}_{ee} + \hat{V}_{ne}, \end{aligned} \quad (1.3)$$

where \mathbf{r} (\mathbf{q}) are the electronic (ionic) positions respectively and Z_a the atomic charges. \hat{K} is the kinetic energy operator while \hat{V}_{ee} and \hat{V}_{ne} denote the electron-electron and electron-ion interaction potential. We have used atomic units in Eq. (1.3) ($\frac{1}{4\pi\epsilon_0} = e = \hbar = m_e = 1$) and this convention is kept throughout this manuscript.

We first start by the presentation of the deterministic methods (Section 1.1) including the Density Functional Theory (DFT) discussed in 1.1.1 and some modern quantum chemistry methods listed in 1.1.2. The intrinsic limitations of these methods will be detailed and discussed. Afterwards, stochastic methods (Section 1.2), especially the Quantum Monte Carlo (QMC) approach, at the core of this work, will be introduced in 1.2.1, 1.2.2, 1.2.3 and 1.2.4. Finally, we will detail how to estimate forces (1.2.5) from QMC simulations, to incorporate them into the MD framework used in the following Chapters.

1.1 Deterministic methods

In this Section, we focus on the deterministic methods for electronic structure calculations. This implies that the results, accurate or not, relevant or not, can be obtained at the machine precision without additional errors. As a consequence, there will be no error bars in the reported results, at variance with those obtained via stochastic techniques.

Let us remind that one of the main goals of a theoretician is to acquire the knowledge of the WF that contains all the useful information on the system. Indeed, some approaches directly involve the system WF such as quantum chemistry methods (Møller-Plesset perturbation theory, Coupled-Cluster approaches) as discussed in Subsection 1.1.2 or QMC techniques

(Section 1.2). Alternatively, DFT works with the electronic density of the system under study as we will see in the following.

1.1.1 Density Functional Theory

Density Functional Theory, formulated in the early 1960s, is one of the most used approaches to determine the electronic structure of physical or chemical systems. The main idea of DFT is by construction fundamentally different than WF-based methods such as HF, post-HF or QMC methods that will be described later. Instead of focusing on the ground-state WF $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, DFT works with a physical quantity that is measurable for example by X-ray diffraction^[24]: the *electronic density* $\rho(\mathbf{r})$. This quantity is defined as

$$\rho(\mathbf{r}) = N \int d\mathbf{r}_2 \dots d\mathbf{r}_N |\Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2, \quad (1.4)$$

where N is the number of electrons in the system. The key idea of DFT is that any property of a many-body system can be casted into a functional of the ground state electronic density $\rho_0(\mathbf{r})$. In their seminal paper^[25], Hohenberg and Kohn demonstrate that the external potential $\hat{V}_{ne}(\mathbf{r})$ is (to within a constant) a *unique* functional of the electronic density $\rho_0(\mathbf{r})$. Reciprocally, $\rho_0(\mathbf{r})$ is *uniquely* defined by the electron-ion external potential $\hat{V}_{ne}(\mathbf{r})$ that, in absence of external fields, only depends on the considered atomic configuration. If one considers the contributions of the Hamiltonian of Eq. (1.3) independent from the external potential, the above statement implies that it exists a *universal* density functional of the energy which can be written as:

$$\mathcal{F}[\rho(\mathbf{r})] = \min_{\Psi} \langle \Psi | \hat{K} + \hat{V}_{ee} | \Psi \rangle = K[\rho] + V_{ee}[\rho]. \quad (1.5)$$

Eq. (1.5) is valid for any external potential \hat{V}_{ne} . Using the variational principle, the DFT ground state energy $E[\rho] = \langle \Psi | \hat{H} | \Psi \rangle$ for a given specific \hat{V}_{ne} is recovered via a global minimization of $\mathcal{F}[\rho]$ with respect to the ground state electronic density $\rho \rightarrow \rho_0$.

This very appealing approach is in principle exact but suffers from at least two limitations, making its application to physical or chemical systems complex. First, there is no explicit mathematical formulation of the DFT energy $E[\rho]$ into well-known and controlled contributions. Second, despite the knowledge of the *a priori* exact electronic density of the ground state that can be measured experimentally, there is no clear way to extract other meaningful observables from ρ_0 .

Nevertheless, some of these issues can be overcome thanks to the Kohn and Sham method described in Ref. 26. In this approach, one replaces the many-body problem with N electrons to an equivalent set of N one-electron problems. This is achieved by considering a fictitious system with a total electronic density equal to the physical one $\rho(\mathbf{r})$. In this case, the electrons of the fictitious system are assumed non-interacting or independent¹.

We can thus describe each Kohn-Sham particle (KS) by a set of single-particle orbitals $\phi_i(\mathbf{r})$ with $i \in [1; N_{\text{orb}}]$. These orbitals are defined on the whole system and are named *Kohn-Sham*

¹Within the KS formulation of the DFT, the N electrons move in an *average* potential generated by the $N - 1$ other electrons, without quantum exchange interactions with the $N - 1$ electrons.

orbitals. They can be expanded either on a delocalized plane-waves basis set with a cutoff for the electronic kinetic energy (as in the Quantum Espresso suite of codes^[27]) or on a localized gaussian or Slater-type basis set (as in Molpro^[28] or other quantum chemistry codes). Given these orbitals, the electronic density of a spin-unpolarized system can be rewritten as

$$\rho(\mathbf{r}) = 2 \sum_{i \in \text{occ}} |\phi_i(\mathbf{r})|^2, \quad (1.6)$$

where the index i runs over the occupied orbitals. Exploiting the fact that KS particles have the same electronic density as that of the physical system, we can write the functional in Eq. (1.5) as a functional of the molecular orbitals $\phi_i(\mathbf{r})$:

$$\begin{aligned} \mathcal{F}[\rho(\mathbf{r})] &= \sum_i^{N/2} \langle \phi_i | -\frac{1}{2} \Delta_{\mathbf{r}} | \phi_i \rangle + \frac{1}{2} \int \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E^{XC}[\rho(\mathbf{r})] \\ &= K^{KS}[\rho(\mathbf{r})] + E^{\text{Ha}}[\rho(\mathbf{r})] + E^{XC}[\rho(\mathbf{r})]. \end{aligned} \quad (1.7)$$

Let us note that the molecular orbitals $\phi_i(\mathbf{r})$ are functionals of the electronic density $\rho(\mathbf{r})$ by applying the Hohenberg and Kohn theorem on a set of non-interacting electrons. The first term in Eq. (1.7) is the electronic kinetic energy of the independent electron system $K^{KS}[\rho(\mathbf{r})]$, the second one is the classical electron-electron repulsive term (also called the Hartree term) $E^{\text{Ha}}[\rho(\mathbf{r})]$, and the last term is, by definition, the *exchange-correlation* term: $E^{XC}[\rho(\mathbf{r})] = V_{ee}[\rho] - E^{\text{Ha}}[\rho(\mathbf{r})] + K[\rho(\mathbf{r})] - K^{KS}[\rho(\mathbf{r})]$. This term includes all the many-body effects not related to the classical electrostatic repulsion: the two first components come from the quantum particles indistinguishability, whereas the two other ones correspond to a kinetic energy correction between the physical and the non-interacting system. If, in $\hat{V}_{ee} - \hat{V}^{\text{Ha}}$, one neglects the contribution due to the electronic correlation, and computes exactly the exchange energy, DFT simply reduces to the HF approximation.

To find the corresponding KS ground state density of the system, Eq. (1.7) is minimized according to a self-consistent procedure applied to solve a set of N single-particle equations usually named the KS equations:

$$\hat{H}^{KS} \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}). \quad (1.8)$$

The KS Hamiltonian matrix is defined as $\hat{H}^{KS} = \frac{\delta \mathcal{F}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} + \hat{V}_{ne}(\mathbf{r})$ and reads as

$$\begin{aligned} \hat{H}^{KS} &= -\frac{1}{2} \Delta_{\mathbf{r}} + \underbrace{\int d\mathbf{r}' \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E^{XC}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}}_{V_{\text{eff}}(\mathbf{r})} + \hat{V}_{ne}(\mathbf{r}) \\ &= H^{1b} + V^{\text{Ha}}[\rho(\mathbf{r})] + V^{XC}[\rho(\mathbf{r})] \end{aligned} \quad (1.9)$$

where H^{1b} includes one-body contributions, namely the electronic kinetic energy and the external electron/ion potential. The set of eigenvalues problem of Eq. (1.8) is the mathematical transcription of the mapping between the true many-body problem and the fictitious particles subjected to the effective potential $V_{\text{eff}}(\mathbf{r})$ whose electronic density is in principle the same as

the exact one.

The KS equations, constituting the core of the DFT approach, are solved self-consistently until the electronic density difference between two iterations becomes lower than a user-defined threshold. The resulting eigenvalues $\{\varepsilon_i\}, i \in [1; N/2]$ are the electronic energy levels of the system described by the \hat{H}^{KS} : their physical meaning is quite complex but the most of the true electronic properties of the system are quite well reproduced by the KS solution². When the minimization procedure is finished, one can use the KS ground state electronic density and orbitals to reconstruct the true density of the physical system using Eq. (1.6).

However, to apply the KS formulation of DFT in practice, we have to face a major issue already mentioned before: we do not have an explicit form of the exchange-correlation $E^{XC}[\rho(\mathbf{r})]$ functional. Thus, an approximate form needs to be devised. Ideally, this expression should completely take into account all the many-body effects starting from the quantum electrostatic potential to the long-range Van der Waals (VdW) effects (*e.g.* polarization, dispersion).

Historically, the first very successful exchange-correlation functional $E^{XC}[\rho(\mathbf{r})]$ is the Local Density Approximation (LDA). Within this approximation, the dependence of E^{XC} has simply a local character, and is taken as the exchange-correlation energy of an homogeneous electron gas of density $\rho(\mathbf{r})$. The LDA exchange-correlation functional is

$$\begin{aligned} E_{XC}^{\text{LDA}}[\rho] &= E_X^{\text{LDA}}[\rho] + E_C^{\text{LDA}}[\rho] \\ &= \int d\mathbf{r} \rho(\mathbf{r}) \varepsilon_X^0(\rho(\mathbf{r})) + \int d\mathbf{r} \rho(\mathbf{r}) \varepsilon_C^0(\rho(\mathbf{r})). \end{aligned} \quad (1.10)$$

The exchange energy $\varepsilon_X^0(\rho)$ is given by the HF formula (in atomic units) $\varepsilon_X^0(\rho) = -\frac{3}{4}(3/\pi)^{1/3}\rho^{1/3}$. The local correlation energy $\varepsilon_C^0(\rho)$ is usually obtained by fitting a set of QMC calculations on the homogeneous electron gas model at different densities^[29,30]. This is, in practice, one of the first and most successful applications of QMC in electronic structure calculations. LDA works surprisingly well for a wide range of systems, especially for solids and crystals. Concerning water, LDA is a very poor approximation since it generally overestimates the binding energy of the water clusters by a factor 2 (from the dimer to the 20-H₂O molecules cluster)^[31–33]. This overbinding arises from a spurious exchange attraction at large distances^[31]. In this thesis, LDA will only be used in the study of the bonding and non-bonding water dimer (Chapter 5) to make a direct comparison with other much more accurate functionals or computational methods.

A first possible improvement to LDA is to include inhomogeneity in the electron density model. This can be done by including the gradient of the density $\nabla\rho(\mathbf{r})$ and its higher order derivatives under the form of a Taylor expansion of the electronic density of the system. This approach, called the gradient-expansion approximation does not actually improve significantly the LDA results because the exchange-correlation hole defined by the expansion does not satisfy the constraints of the physical exchange-correlation hole. However, these constraints can

²This is however not the case of the band gap in semiconductors, for instance.

be fulfilled if one includes *both* the density ρ and its gradient $\nabla\rho$, and specifies the magnitude of the local gradient by the so-called 'reduced gradient' defined as the dimensionless quantity $x \equiv |\nabla\rho|/\rho^{4/3}$ or sometimes as $s \equiv x/(2(3\pi^2)^{1/3})$. Hence, the Generalized Gradient Approximation (GGA) functionals have the following form:

$$\begin{aligned} E_{XC}^{\text{GGA}}[\rho] &= E_X^{\text{GGA}}[\rho] + E_C^{\text{GGA}}[\rho] \\ &= \int d\mathbf{r} f_X(\rho(\mathbf{r}), s(\mathbf{r})) + \int d\mathbf{r} f_C(\rho(\mathbf{r}), s(\mathbf{r})), \end{aligned} \quad (1.11)$$

where f_X and f_C specify the local exchange and correlation parts respectively. We note that the functional form of f_X at large reduced gradient s is of paramount importance for non-covalent interactions in molecular systems such as water, because exchange-overlap interactions depend strongly on the behaviour of f_X in the regions where the electron densities overlap^[34].

We now mention two GGAs that will be used during this thesis and we discuss their use for water and aqueous systems. Let us start with the Perdew-Burke-Ernzerhof (PBE) functional which has been widely used to study a large variety of systems, from crystals, to complex molecular systems, and of course water. Another functional very often used in literature is the BLYP functional that combines the B88^[35] and Lee-Yang-Parr^[36] approximations for exchange and correlation, respectively.

These functionals give quite reasonable results of binding energy for water clusters in gas phase, with an accurate value for the water dimer energy obtained with PBE functional compared to CC and QMC calculations, as it will be shown. However, it is well-known that these GGAs functionals tends to display a too low value of the first minimum position of the oxygen-oxygen distribution function g_{OO} of liquid water with respect to experiment, a characteristic signature of the overstructuring of the liquid. Consequently, the corresponding diffusion constant is too small (10 times smaller) and the average number of H-bonds is too large^[37]. Worse still, GGAs functionals predict that ice sinks in water, that is to say water has a lower density than ice. Indeed, BLYP functional predicts a liquid density about ~ 0.8 g/mL^[38-40], which is much lower than the experimental value 1 g/mL. Santra and co-workers^[41] demonstrated that the overstructuring of liquid water actually comes from an overbinding of single monomer molecules extracted from a MD simulation of liquid water at ambient temperature. Indeed, BLYP and PBE predict dissociation energies that are too large by 80 and 43 meV, respectively, compared to CCSD(T) calculations. This is inferior to the performance of these functionals for the equilibrium water dimer and other water clusters in the gas phase. It thus affects the strength of the H-bond in water dimers inside the liquid, giving too large dissociation energies with respect to the gas phase results.

To make a further step in the functional improvement, one can add a fraction of exact HF exchange energy defined as

$$E_X^{\text{HF}} = -\frac{1}{2} \sum_{i,j}^{N/2} \int \int d\mathbf{r} d\mathbf{r}' \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} \phi_i(\mathbf{r}') \phi_j^*(\mathbf{r}'). \quad (1.12)$$

Chapter 1. Electronic structure calculations

In this thesis, we will use the B3LYP exchange-correlation functional that involves a linear combination of GGAs and LDA, defined as

$$E_{XC}^{\text{B3LYP}} = a_0 E_X^{\text{HF}} + a_X E_X^{\text{B88}} + (1 - a_0 - a_X) E_X^{\text{LDA}} + a_C E_C^{\text{LYP}} + (1 - a_C) E_C^{\text{LDA}} \quad (1.13)$$

where E_X^{B88} and E_C^{LYP} are the exchange and correlation parts of the BLYP functional discussed above and a_0 , a_X and a_C are mixing coefficients.

As expected, the obtained binding energies of water clusters in both gas and condensed phases are considerably improved by the use of hybrid functionals such as B3LYP or PBE0 (PBE with 25% of exact HF exchange) and differ by a few meV per H-bond from the almost exact quantum chemistry calculations that will be discussed in Subsection 1.1.2. Hybrid functionals soften the structure of liquid water since they give slightly smaller number of H-bonds than GGA functionals, but similar H-bond populations^[42]. However, hybrid functionals still slightly underestimate (by 0.45 kcal/mol for the B3LYP functional) the energy difference between non H-bonded geometries of the water dimer and its global minimum^[43]. Small errors (~ 1.3 kcal/mol) in the predicted binding energies have also been observed in the protonated water dimer, trimer and tetramer in Ref. 44. The binding energies of the six first neutral water clusters are improved (by about 1.5 kcal/mol with respect to the B3LYP results) by the use of *meta*-GGA functionals (GGA functional including Laplacian of the electronic density $\nabla^2\rho$ into the exchange-correlation energy) with extended aug-cc-pV5Z Dunning basis set. Unfortunately, these calculations rapidly become unaffordable from a computational point of view, especially in a MD framework, even for moderate size clusters. It has also been recently established that B3LYP functional works remarkably well for the study of geometrical properties of protonated water clusters (from the protonated water dimer to the $\text{H}^+(\text{H}_2\text{O})_8$ cluster), in particular for the equilibrium oxygen-oxygen distance^[45]. In the meantime, PBE0 geometries are erroneous but give very accurate anharmonic frequencies for the IR spectra of these clusters. These observations suggest a major limitation of DFT to reach a quantitative description of proton transfer and diffusion in water. There is no systematic way to improve the description of fundamental properties such as equilibrium geometries, IR frequencies or dipole moments for a given family of functionals, making this method hardly reliable for the study of unknown new systems.

The fact that GGAs and hybrid functionals work reasonably well for clusters in gas phase and quite poorly in condensed phase prove these functionals are missing many-body effects such as VdW interactions or dispersion. Such interactions are essential in water because of its large polarity. The VdW interactions are however one order of magnitude weaker than the H-bonds. Moreover, they are nondirectional because of the spherical symmetry of the dispersion forces existing at large OH distances.

To include dispersion, various approaches are possible: one can simply add an atom-atom pair potential of the form $-C/R^6$ between every pair of atoms and parametrize the constant C for several choices of basis sets or functionals, from WF-based methods such as those described later. This approach is known as DFT-D or DFT-D2 and has been extensively developed by Grimme and coworkers^[46]. In spite of its empirical character, the Grimme corrections for dispersion significantly improve the description of liquid water with much more

reasonable, although slightly overestimated, values for the obtained density (1.07 – 1.13 g/mL for BLYP-D3 functional^[47]).

Another approach, used in this thesis, is to include in the exchange-correlation functional a non-local correlation term E_C^{nl} depending explicitly on electron densities at spatially separated positions. The resulting exchange-correlation functional, combined with the usual GGA representation^[48] reads as

$$E_{XC} = E_X^{\text{GGA}} + E_C^{\text{LDA}} + E_C^{\text{nl}}, \quad (1.14)$$

with the general form often assumed for the non-local correlation

$$E_C^{\text{nl}} = \int d\mathbf{r}_1 d\mathbf{r}_2 \rho(\mathbf{r}_1) \Phi(\mathbf{r}_1, \mathbf{r}_2) \rho(\mathbf{r}_2). \quad (1.15)$$

The kernel Φ , also named internal functional, is itself a functional of the density but is calibrated to ensure the correct asymptotic behavior of the dispersion interaction by falling off as $1/|\mathbf{r}_1 - \mathbf{r}_2|^6$ at large distances^[49]. In this thesis, we will use the rPW86-DF2 functional named DF2 in the following.

As expected, the VdW-corrected functionals give very close results to nearly exact quantum chemistry calculations for different configurations of the water hexamer^[50]. An additional softening compared with GGA results is also observed. This is characterized by a much lower and broader first peak of the oxygen-oxygen radial correlation function, in much better agreement with experimental data, although not yet perfect^[51]. Indeed, this time, the liquid is understructured^[52]. However, DF2 is referenced as one of the most accurate functional approach for water since it obtains the quite good score of 68% of the scoring algorithm defined by Gillan and colleagues^[53]. The scoring scheme assigns a percentage score to any chosen approximation, according to its performance for the properties (*e.g.* binding energy, sublimation energy, equilibrium distances and volumes) of the water monomer, the dimer, the hexamer, and ice structures. In that paper, the major part of the technical issues about DFT approach in water are reviewed, starting from the study of simple water clusters, to different ice structures in the perspective of using the most appropriate functional for the specific physical or chemical problem we want to tackle. For information, we report here the obtained scores for other usual functionals that will be used in this thesis. Only 14% for LDA, which is not a surprise because of the roughness of this approximation. The PBE functional gets a much more reasonable score of 56%, but far from being satisfactory to obtain quantitative results. BLYP score is not a great success neither 42% but is significantly improved with dispersion corrections: 66% for BLYP-D3, proving again the essential role played by VdW physics in water. Unfortunately, B3LYP score is not given but we can reasonably assume it is close the PBE0 one 62%.

Let us conclude this review of different functionals used to study water and aqueous systems by mentioning the Strongly Constrained and Appropriately Normed (SCAN) functional. This meta-GGA functional is, unlike most of GGA or hybrid functionals, not empirical since it satisfies 17 exact constraints on semi-local exchange-correlation functionals. SCAN functional displays a surprisingly good and quantitative agreement with quantum chemistry results for the water hexamer configurations and water monomer properties^[54], while VdW-corrected

functionals are sometimes unable to make qualitative descriptions. Very recently, Chen and coworkers performed *ab initio* MD simulations of liquid water at ambient conditions^[55]. They present SCAN as a very promising candidate to model liquid water since the obtained Radial Distribution Functions (RDFs) are very close to X-ray and neutron scattering experiments. There is still a slight overstructuring of the first peak of the $g_{\text{OO}}(r)$ while the first peak of $g_{\text{OH}}(r)$ function is too narrow. This can be explained by the lack of Nuclear Quantum Effects (NQE) of hydrogen, known to be very important to reach a quantitative description of the structure and the dynamics of water. Such effects will be discussed in the Chapter 2, where they will be accounted for within the Feynman discretized Path Integral (PI) approach.

In conclusion, simulating liquid water or gas phase clusters using DFT has been and is still a very active field of research. Indeed, the interest of DFT is motivated by the very reasonable scaling of this approach with the number of electrons of the system: $\mathcal{O}(N^3)$. It thus constitutes a very good compromise between accuracy and computational cost of *ab initio* MD simulations driven by DFT forces. Despite its exact nature by construction, DFT in practice will always suffer from the non-exact knowledge of the exchange-correlation functional. The latter cannot be universal and transferred to a large variety of physical and chemical systems because of - except the recent SCAN functional - empirical parametrization of the exchange-correlation functional.

To reach a quantitative and reliable description of water and small protonated clusters beyond the chemical accuracy, one has to make a step back and start from the Schrödinger equation (1.2) to develop theories working with the system WF, as we will see in the following Subsection.

1.1.2 Quantum chemistry methods

In this Subsection, we will introduce the basic concepts of two advanced quantum chemistry methods that will then be used to compare with the QMC results of this thesis:

- the second-order Møller-Plesset perturbation theory method also referred to as MP2;
- the Coupled-Cluster approach, with Single and Double excitations, with or without Triple excitations added in a perturbative way, named CCSD or CCSD(T), respectively. This method often constitutes the reference for direct comparison with other levels of theory because of its extreme accuracy. It is thus considered as the "golden standard" in quantum chemistry calculations.

1.1.2.1 Second-order Møller-Plesset perturbation theory

As suggested by its name, MP2 is a second-order perturbation theory in which the Schrödinger equation (1.2) is rewritten in a perturbative way as follows

$$\hat{H}|\Psi_i\rangle = (\hat{H}_0 + \lambda\hat{V})|\Psi_i\rangle = E_i|\Psi_i\rangle, \quad (1.16)$$

where we know the exact eigenfunctions and the eigenvalues of the unperturbed Hamiltonian \hat{H}_0 (this quantity will be detailed later). This approach is called the Rayleigh-Schrödinger

perturbation theory:

$$\hat{H}_0|\Psi_i^{(0)}\rangle = E_i^{(0)}|\Psi_i^{(0)}\rangle \quad \text{or} \quad \hat{H}_0|i\rangle = E_i^{(0)}|i\rangle, \quad (1.17)$$

where we have denoted $|\Psi_i^{(0)}\rangle = |i\rangle$ for compactness. λ is an ordering parameter representing the perturbation which will be then set equal to unity. We now expand the exact eigenfunctions and eigenvalues in Taylor series in λ

$$\begin{aligned} E_i &= E_i^{(0)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \dots \\ |\Psi_i\rangle &= |i\rangle + \lambda|\Psi_i^{(1)}\rangle + \lambda^2|\Psi_i^{(2)}\rangle + \dots \end{aligned} \quad (1.18)$$

We denote here $E_i^{(n)}$ the n th-order energy. Since we are interested in the *second-order* perturbative terms and more specifically the second-order electronic energy, we will truncate the Taylor expansions of Eq. (1.18) to the terms with λ^2 as a prefactor. After some mathematical manipulations and standard linear algebra that is detailed in Appendix A for the interested reader, one obtains for the second-order energy

$$E_i^{(2)} = \sum_{n \neq i} \frac{\langle i|\hat{V}|n\rangle\langle n|\hat{V}|i\rangle}{E_i^{(0)} - E_n^{(0)}} = \sum_{n \neq i} \frac{|\langle i|\hat{V}|n\rangle|^2}{E_i^{(0)} - E_n^{(0)}}. \quad (1.19)$$

In the case of MP2 theory, the unperturbed Hamiltonian is the shifted Fock operator \hat{F} , which reads as

$$\hat{H}_0 \equiv \hat{F} + \langle \Phi_0 | (\hat{H} - \hat{F}) | \Phi_0 \rangle, \quad (1.20)$$

where the Fock operator for the i -th electron is given by

$$\hat{F}(i) = \hat{h}(i) + \sum_{j=1}^{N/2} \left[2\hat{J}_j(i) - \hat{K}_j(i) \right]. \quad (1.21)$$

In the above expression, $\hat{h}(i)$ is the one-electron Hamiltonian for the i -th electron, $\hat{J}_j(i)$ is the Coulomb operator, defining the repulsive force between the i -th and j -th electrons and $\hat{K}_j(i)$ is the exchange operator between these two electrons. In Eq. (1.20), the perturbation term, also called the *correlation potential* is

$$\hat{V} \equiv \hat{H} - \hat{H}_0 = \hat{H} - \hat{F} - \langle \Phi_0 | (\hat{H} - \hat{F}) | \Phi_0 \rangle. \quad (1.22)$$

In Eqs. (1.20) and (1.22), \hat{H} is the usual electronic Hamiltonian defined in Eq. (1.3) and the SD $|\Phi_0\rangle$ filled with HF orbitals Ψ_i^{HF} , is the lowest-energy eigenfunction of the Fock operator \hat{F} :

$$\hat{F}|\Phi_0\rangle = 2 \left(\sum_{i=1}^{N/2} \varepsilon_i^{\text{HF}} \right) |\Phi_0\rangle \quad (1.23)$$

Using Eq. (1.23), it follows that

$$\hat{H}_0|\Phi_0\rangle = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle |\Phi_0\rangle \quad (1.24)$$

Chapter 1. Electronic structure calculations

so that the 0^{th} -order energy is the expectation value of \hat{H} with respect to $|\Phi_0\rangle$, that is to say, the HF energy:

$$E_{\text{MP0}} = E_{\text{HF}} = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle. \quad (1.25)$$

Since the WF $|\Phi_0\rangle$ is optimal for both \hat{H} and \hat{F} operators, it is clear from Eq. (1.22) that the first-order MP energy is equal to

$$E_{\text{MP1}} = \langle \Phi_0 | \hat{V} | \Phi_0 \rangle = 0. \quad (1.26)$$

This result is known as the Møller-Plesset theorem: the correlation potential does not contribute in first-order to the exact electronic energy. In other words, the lowest-order MP correlation energy appears in second order.

In order to obtain the MP2 formula for a closed-shell molecule, we exploit the second order energy given in Eq. (1.19) written in a basis of *doubly* excited SD determinants. Indeed, the Brillouin theorem states that singly excited SD are orthogonal to the ground state WF of the system so their contribution to the total energy is zero. After integrating out the spin, the practical formula to compute the MP2 energy is given by

$$E_{\text{MP2}} = - \sum_{i,j}^{N_{\text{orb}}} \sum_{a,b}^{N_{\text{vir}}} \frac{2 \langle \Psi_a^{\text{HF}}(1) \Psi_b^{\text{HF}}(2) | \frac{1}{r_{12}} | \Psi_i^{\text{HF}}(1) \Psi_j^{\text{HF}}(2) \rangle - \langle \Psi_a^{\text{HF}}(1) \Psi_b^{\text{HF}}(2) | \frac{1}{r_{12}} | \Psi_j^{\text{HF}}(1) \Psi_i^{\text{HF}}(2) \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \times \langle \Psi_i^{\text{HF}}(1) \Psi_j^{\text{HF}}(2) | \frac{1}{r_{12}} | \Psi_a^{\text{HF}}(1) \Psi_b^{\text{HF}}(2) \rangle \quad (1.27)$$

where N_{orb} is the number of occupied orbitals and N_{vir} is the number of virtual orbitals. The indices i and j refer to the occupied orbitals while a et b are employed for virtual electronic states.

Finally, within the MP2 approach, the total electronic energy of the system is given by the HF energy plus the second-order MP contribution

$$E = E_{\text{HF}} + E_{\text{MP2}} \leq E_{\text{HF}}. \quad (1.28)$$

In this case, the correlation energy E_{corr} , always negative by definition, is given by the MP2 energy $E_{\text{corr}} = E_{\text{MP2}}$. MP2 recovers a relatively large part of the dynamic correlation while the presence of the HF ground state ensures the inclusion of exact nonlocal exchange. The same perturbative treatment can be repeated by exploring higher order terms of the Rayleigh-Schrödinger equation (1.16) to obtain the MP3 or even the MP4 energies.

Let us discuss now the use of MP2 theory to study properties of neutral or charged water clusters, in gas or condensed phase. Concerning the elementary properties of clusters in gas phase, it is well established that MP2 results are in excellent agreement with the almost exact Coupled-Cluster methods^[56,57] that will be detailed in the forthcoming pages of this manuscript. Indeed, the binding energies and geometries of each water cluster $(\text{H}_2\text{O})_n$ ($1 \leq n \leq 20$) are very well reproduced (agreement within 0.05 kcal/mol with experimental

data). Parkkinen and his colleagues used MP2 to explore geometries and stabilization energies of the first 21 protonated water clusters, until the magical $\text{H}^+(\text{H}_2\text{O})_{21}$ known to minimize all competing energy contributions simultaneously (prism, cage...) [58]. They concluded that the global cluster energy was dominated by the geometry of the H_9O_4^+ Eigen complex. Indeed, its flat geometry favours the polyhedral cages whereas the prismatic structures are less stable.

However, towards a description of water and aqueous systems, studies of neutral or charged clusters in gas phase are used as toy models or benchmarks. Indeed, advanced electronic structure calculations should provide correct measurements of water density or proton diffusion towards an accurate description of interactions between water molecules in *condensed* phase. That is why, in the last five years, several groups focused their efforts to perform the first simulations of liquid water at ambient conditions. Del Ben and colleagues, by performing a Monte Carlo (MC) simulation of liquid water with an MP2 evaluation of the energy between a series of MC moves, proved the ability of MP2 to properly describe water in *bulk* liquid since they obtained a very satisfactory density of 1.02 g/mL. Despite a too high first peak, the generated $g_{\text{OO}}(r)$ are in very good agreement with experiments [59].

More recently, a MD simulation of liquid water at ambient conditions within MP2 and SCS-MP2 frameworks (MP2 improvement in which a separate scaling of the correlation energy contributions from antiparallel- "singlet" and parallel- "triplet" spin pairs of electrons is applied [60]) confirmed the good performances of this level of theory in describing water in condensed phase [61]. Researchers checked more critical observables such as the water self-diffusion constant $D \simeq 0.27 \text{ \AA}^2/\text{ps}^2$ being very close to the experimental value $D_{\text{exp}} = 0.23 \text{ \AA}^2/\text{ps}^2$, IR vibrational frequencies and dipole moments. The RDFs behave similarly that in Ref. 59 with a too high first peak. A better agreement with experiment for SCS-MP2 approach was found, suggesting that spin correlations could be important in water. These subtle effects will be partially discussed in Chapter 5.

In 2017, Hirata's group (already author of the aforementioned MP2 MD simulation of liquid water) compared the properties of MP2-water and DFT-water. To keep the system into the liquid phase during the MP2 simulations of water, the temperature has to be lowered to 250 K and a negative pressure of -0.6 GPa needs to be applied. Alternatively, PBE liquid water exists at 440 K and 0.3 GPa, where it has a RDF and density close to experiments. On the one hand, the polarizability has the greatest influence on the temperature needed in the simulation to make water liquid, and is significantly greater in DFT/GGA than in MP2 or SCS-MP2. Consequently, MP2-water is "cooler" than DFT-water. On the other hand, the smaller the dispersion interaction, the less dense the liquid water and the more positive a pressure is needed to maintain the correct density of the liquid. This scenario corresponds to the simulations with dispersionless DFT/GGA functionals, whereas VdW interactions are naturally taken into account by MP2, that slightly overestimates dispersion effects in liquid water. Therefore, MP2-water is "denser" than DFT-water [62].

Therefore, the MP2 approach can be considered as a serious candidate to tackle the problem of proton transfer in liquid water and other aqueous systems thanks to its robustness regarding a large set of static (RDFs, density) and dynamical (self-diffusion, IR frequencies)

properties. Nevertheless, the price to pay to reach that level of accuracy is quite expensive and is estimated as $\mathcal{O}(N^5)$ with N the number of electrons. This explains why Del Ben and coworkers performed a large DFT pre-sampling between each MC iterations and why the computed trajectory in Ref. 61, although reasonable, is short compared to MD standards for the evaluation of dynamical quantities (11 ps of trajectory instead of 40-50 ps). In the perspective of dealing with water as a *solvent*, the MP2 approach will unfortunately become prohibitive and alternative solutions must be found.

1.1.2.2 Coupled-Cluster methods: CCSD and CCSD(T)

The Coupled Cluster (CC) approach is today considered as the "golden standard" for electronic structure and quantum chemistry due to its extreme accuracy for a large variety of molecular systems. CC calculations will be used in this thesis as benchmark to validate our calculations using other electronic structure methods. Here, we provide a basic introduction to CC theory and we invite the interested reader to see Ref. 63 for further details about recent methodological advances.

The CC theory mainly relies on the exponential ansatz for the CC WF

$$|\Psi_{CC}\rangle = e^{\hat{T}}|\Phi_0\rangle, \quad (1.29)$$

where $|\Phi_0\rangle$ is the HF WF and \hat{T} is the cluster operator which is the sum of cluster operators at different excitation levels

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots + \hat{T}_N. \quad (1.30)$$

In Eq. (1.30), \hat{T}_1 is the cluster operator for the single excitations, which can be written, within the second quantization formalism, as

$$\hat{T}_1 = \sum_i^{N_{\text{orb}}} \sum_a^{N_{\text{vir}}} t_i^a \hat{a}_a^\dagger \hat{a}_i. \quad (1.31)$$

In the above equation, t_i^a represents the amplitude of the corresponding single excitation promoting an electron from the occupied orbital i to the virtual state a via the \hat{a}_i annihilation and \hat{a}_a^\dagger creation operators. When the operator $\hat{a}_a^\dagger \hat{a}_i$ acts on the HF WF, it creates the singly excited determinant $|\Phi_i^a\rangle = \hat{a}_a^\dagger \hat{a}_i |\Phi_0\rangle$.

Similarly, \hat{T}_2 is the cluster operator for double excitations and is written as

$$\hat{T}_2 = \frac{1}{4} \sum_{i,j}^{N_{\text{orb}}} \sum_{a,b}^{N_{\text{vir}}} t_{ij}^{ab} \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j, \quad (1.32)$$

where t_{ij}^{ab} are the double-excitations amplitudes. When the operator $\hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j$ acts on the HF WF $|\Phi_0\rangle$, it generates the doubly excited determinant $|\Phi_{ij}^{ab}\rangle = \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j |\Phi_0\rangle$. And so on up to the \hat{T}_N cluster operator for N -fold excitations.

For the sake of simplicity, we will in the following restrict to the case of single and double excitations, that is to say, we simply consider the operator $\hat{T} = \hat{T}_1 + \hat{T}_2$. This approach is known as Coupled Cluster Single Doubles (CCSD). By applying a Taylor expansion of the excitation operator \hat{T} , it can be demonstrated (see Appendix B for further details) that the amplitude of the triple excitations can be estimated by products of single- t_i^a and double- t_{ij}^{ab} excitations amplitudes. Moreover, it is a reasonable approximation to consider that higher-order excitations contributions to the correlation energy will be much smaller than single and double ones. Indeed, CCSD WF recovers a large fraction of the correlation energy of the system. CCSD(T) further improves upon CCSD, because also triple excitations are added, in a perturbative way. Besides, thanks to the exponential form of the WF, truncated CC methods are size-consistent, allowing for instance accurate studies of water clusters dissociation.

To determine the CC energy and the excitation amplitudes, one applies the *projection method* in which CC WF must satisfy the Schrödinger equation (1.2) in the space spanned by the HF determinant $|\Phi_0\rangle$ and excited determinants $|\Phi_i^a\rangle, |\Phi_{ij}^{ab}\rangle$ and so on

$$\begin{aligned} \langle \Phi_0 | (\hat{H} - E) | \Psi_{CC} \rangle &= 0 \\ \langle \Phi_i^a | (\hat{H} - E) | \Psi_{CC} \rangle &= 0 \\ \langle \Phi_{ij}^{ab} | (\hat{H} - E) | \Psi_{CC} \rangle &= 0. \end{aligned} \quad (1.33)$$

Using the orthonormality of the determinants, Eq. (1.33) directly gives the CC total energy

$$\begin{aligned} E &= \langle \Phi_0 | \hat{H} | \Psi_{CC} \rangle \\ &= \langle \Phi_0 | \hat{H} | \Psi_0 \rangle + \sum_i^{N_{\text{orb}}} \sum_a^{N_{\text{vir}}} c_i^a \langle \Phi_0 | \hat{H} | \Phi_i^a \rangle + \sum_{i < j}^{N_{\text{orb}}} \sum_{a < b}^{N_{\text{vir}}} c_{ij}^{ab} \langle \Phi_0 | \hat{H} | \Phi_{ij}^{ab} \rangle. \end{aligned} \quad (1.34)$$

According to Slater's rules, only first- and second-order excitations contribute to the electronic energy. Since the first term of Eq. (1.34) is simply the HF total energy $E_{\text{HF}} = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle$ and since the second term vanishes thanks to the Brillouin theorem, we obtain the CC correlation energy

$$\begin{aligned} E_{\text{corr}} &= \sum_{i < j}^{N_{\text{orb}}} \sum_{a < b}^{N_{\text{vir}}} c_{ij}^{ab} \langle \Phi_0 | \hat{H} | \Phi_{ij}^{ab} \rangle = \frac{1}{4} \sum_{i < j}^{N_{\text{orb}}} \sum_{a < b}^{N_{\text{vir}}} (t_{ij}^{ab} + 2t_i^a t_j^b) \\ &\times \left(\langle \Psi_i^{\text{HF}}(1) \Psi_j^{\text{HF}}(2) | \frac{1}{r_{12}} | \Psi_a^{\text{HF}}(1) \Psi_b^{\text{HF}}(2) \rangle - \langle \Psi_i^{\text{HF}}(1) \Psi_j^{\text{HF}}(2) | \frac{1}{r_{12}} | \Psi_b^{\text{HF}}(1) \Psi_a^{\text{HF}}(2) \rangle \right). \end{aligned} \quad (1.35)$$

Thus, this quite simple expression for the energy only involves single and double excitations amplitudes and matrix elements. Finally, let us note that this way to evaluate the CC energy is not variational, hence the obtained CC energy does not necessarily fall above the exact ground-state energy, at variance with perturbation theories such as MP2.

As already mentioned in the previous Subsections, CC techniques (mainly CCSD/CCSD(T)) are considered as the most precise single reference methods one can use today for electronic structure calculations. In the case of water, only small clusters in gas phase have been studied so far and the obtained results are in excellent agreement with experimental data for any observed quantity: equilibrium geometries, IR frequencies, bond strengths, dipole moments to cite a few^[41,64]. However, despite its elegance and its extreme reliability, the application of this method for extended system is 'nip in the bud' because of its computational cost. Indeed, the CCSD/CCSD(T) scaling laws with the number N of electrons are estimated to be in $\mathcal{O}(N^7)$, which is clearly prohibitive, without considering the additional problem of the convergence of the calculations with the basis set size. Therefore, the two main tasks of CC techniques are:

- perform benchmark calculations that can be then used for direct comparison with other levels of theory to validate new methodological approaches. This is exactly what we will do in the Chapter 4 where CCSD(T) calculations will constitute our reference to compare our QMC-MD results with;
- use CCSD/CCSD(T) reference calculations on a large set of geometries to generate very accurate Potential Energy Surface (PES) of the corresponding small molecular system under study. This approach has been used, for instance by the Bowman group, in the 2000s to build analytical potentials parametrized by fits of thousands of CCSD(T) geometries. They started with the neutral clusters^[40,65-67] before focusing their efforts on the charged dimer (Zundel ion)^[65]. The latter PES will be used in this thesis (Chapter 4), as accurate force field to benchmark our new Molecular Dynamic algorithms. Polarizable potentials are also emerging, still obtained via CCSD(T) calculations on extended Dunning basis sets^[68-70] and reproduce quite fairly water properties in both gas and condensed phases.

We have reviewed the principal deterministic methods to resolve, as precisely as possible, the electronic structure of the systems we will study in this thesis, namely the neutral and protonated water clusters. DFT is computationally cheap, but hardly reliable to predict new physics on unexplored systems, while quantum chemistry techniques are by far more accurate, but too expensive to be applied on larger systems in a close future. That is why, in the next Section (1.2), we will introduce the stochastic methods, focusing on the QMC approach, which looks to be a promising candidate to overcome the above issues.

1.2 Stochastic methods

In this Section, we will explain how it is possible to accurately determine electronic structures of a large variety of physical and chemical systems by means of stochastic methods, focusing especially on the Quantum Monte Carlo (QMC) approach. Monte Carlo refers to a set of stochastic methods relying on repeating random samplings of a given quantity to obtain a numerical estimation of it. The highly random character of this method motivated its naming, by reference to the very famous borough of Monaco where casinos are numerous.

1.2.1 The Quantum Monte Carlo approach

The problem we want to tackle here is still the same one as in Section 1.1, namely solving the Schrödinger equation (1.2). This can be done within the QMC approach, that optimizes a trial WF $\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_N)$ using the variational principle to find the best approximate solution of Eq. (1.2). The QMC final WF is supposed to be very close to the exact ground state WF of the system Ψ_0 , and must capture an important fraction of the correlation energy defined in Eq. (1.1).

One of the major advantages of stochastic techniques and QMC methods, compared with its direct competitors for advanced electronic calculations, is its ability to deal with the high dimensionality of electronic integrals at a reasonable computational cost. Furthermore, the stochastic nature of QMC algorithms makes this class of simulations particularly adapted for High Performance Computing (HPC) since random sampling can routinely be done within a parallelization scheme. Considering the recent and coming advances on the development of highly parallelized architectures such as the Tier-0 machines, the QMC approach looks very promising and we thus expect it to become more widespread in the near future.

Indeed, the very first application of QMC dates back to the work on liquid helium by MacMillan about fifty years ago^[71]. In spite of a few major successes such as the seminal work of Ceperley *et al.*^[29] on the electron gas, it seems that the use of QMC methods is yet limited to a quite restricted circle of experts. This can be partially explained by the fact that QMC approaches are computationally demanding since a large amount of generations is necessary to reach a reasonable target accuracy. Therefore, in spite of its milder theoretical scaling with respect to the system size, between $\mathcal{O}(N^3)$ and $\mathcal{O}(N^4)$, compared with other advanced electronic calculation methods, significant applications have been done only very recently. Indeed, for small or moderate size systems, this apparent gain is hindered by the very large statistical prefactor within the QMC approach, at variance with deterministic methods. For instance, Dagrada and coworkers demonstrated that the water hexamer was the crossing point between CCSD(T) and QMC calculations^[72]: QMC becomes "cheaper" than CC techniques for systems larger than $(\text{H}_2\text{O})_6$. Thus, in the perspective of studying PT reactions in liquid water or in biological systems, QMC methods are the only viable candidate for both large and accurate calculations in the condensed phase.

In the following, we will discuss the form of the QMC WF in 1.2.2 and present the two main families of QMC simulations:

- the Variational Monte Carlo (VMC) that optimizes the QMC trial WF Ψ_T through the variational principle, minimizing the total energy of the system (Subsection 1.2.3);
- the Diffusion Monte Carlo (DMC) and Lattice Regularized Diffusion Monte Carlo (LRDMC) that reinterpret the Schrödinger equation (1.2) into a diffusion equation of the trial WF in the Hilbert space of the electronic parameters (Subsection 1.2.3).

The two above techniques evaluate the WF in the $3N$ -dimensional configurational space spanned by the electronic coordinates, while other QMC techniques, that will not be detailed

here, rely on the random sampling of determinants such as Full Configuration Interaction Quantum Monte Carlo (FCIMC)^[73] or Auxiliary Field Quantum Monte Carlo (AFQMC)^[74].

1.2.2 The Quantum Monte Carlo wave function

Let us now consider the typical form of the QMC WF which will be then used in this thesis. In the following, we describe the N electrons of the system with the generalized coordinates $\{\mathbf{x}_i = (\mathbf{r}_i, \sigma_i)\}_{i=1\dots N}$ grouping electronic positions \mathbf{r}_i and spins σ_i . For convenience, we restrict our derivation to the case of spin-unpolarized systems ($N_\uparrow = N_\downarrow = N/2$) and we refer the interested reader to Refs. 75 and 76 for the spin-polarized case.

The WF used in our QMC calculations is defined by the product of two contributions as follows

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \exp[-J(\mathbf{x}_1, \dots, \mathbf{x}_N)]\Psi_D(\mathbf{x}_1, \dots, \mathbf{x}_N), \quad (1.36)$$

where $J(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is a bosonic (symmetric with respect to the exchange of two electrons) factor called the *Jastrow factor*. On the contrary, $\Psi_D(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is a fermionic (antisymmetric with respect to the exchange of two electrons) factor which ensures the entire WF being antisymmetric in agreement with fermion statistics. It is generally referred as the determinantal part of the WF since it can contain a single or a sum of determinants as we will see in the following. The QMC WF is written in such a compact form with an exponential form of the Jastrow factor to ensure a good convergence of the electronic energy with the size of the parameters set λ defining the WF. Moreover, a direct comparison can be made with the CC WF of Eq. (B.2) by making a Taylor expansion of the Jastrow factor that gives a series of determinants modulated by $J(\mathbf{x}_1, \dots, \mathbf{x}_N)$.

1.2.2.1 Antisymmetrized Geminal Power (AGP) part

First, let us focus on the determinantal part of the WF described in Eq.(1.36). The simplest assumption is that $\Psi_D(\mathbf{x}_1, \dots, \mathbf{x}_N)$ is described via a single Slater determinant (SD) which reads as

$$\Psi_D(\mathbf{x}_1, \dots, \mathbf{x}_N) = \det[\chi_i(\mathbf{x}_j)], \quad (1.37)$$

where $\chi_i(\mathbf{x}_j) = \Psi_i(\mathbf{r}_j) \otimes \Sigma_i(\sigma_j)$ are spin-orbitals. The corresponding WF is thus called a Jastrow-Slater-Determinant (JSD) WF and will mostly be used in Chapter 5. In this thesis, we are going beyond the JSD ansatz, and use the more accurate *Resonating Valence Bond* (RVB) approach whose QMC analogue is given by the *Jastrow Antisymmetrized Geminal Power* (JAGP) WF^[75,77]. This WF is at the heart of the *TurboRVB* code, used in this thesis. The JAGP ansatz is

$$\Psi_{\text{AGP}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \hat{A} [\Phi(\mathbf{x}_1, \mathbf{x}_2), \dots, \Phi(\mathbf{x}_{N-1}, \mathbf{x}_N)], \quad (1.38)$$

where \hat{A} is an antisymmetric operator and $\Phi(\mathbf{x}_i, \mathbf{x}_j)$ represents a geminal or a pairing function. The geminals are antisymmetric functions of two electron coordinates written as a product of

a spatial symmetric part and a spin singlet

$$\Phi(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{r}_i, \mathbf{r}_j) \frac{\delta(\sigma_i, \uparrow)\delta(\sigma_j, \downarrow) - \delta(\sigma_i, \downarrow)\delta(\sigma_j, \uparrow)}{\sqrt{2}}. \quad (1.39)$$

Consequently, one can consider the AGP as a generalization of the SD to electronic pairs by choosing a determinant for the antisymmetrization operator \hat{A} :

$$\Psi_{\text{AGP}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \det[\Phi(\mathbf{x}_i, \mathbf{x}_j)]. \quad (1.40)$$

In the TurboRVB code, Φ is expanded onto a localized gaussian basis set with atom-centered functions. Defining \mathbf{q} the ionic positions (see Chapters 2 and 3), the basis functions have the following form

$$G_\mu^a(|\mathbf{r} - \mathbf{q}_a|) \propto |\mathbf{r} - \mathbf{q}_a|^l e^{-\zeta_{l,n}|\mathbf{r} - \mathbf{q}_a|^2} Y_{l,m}(\Omega_{\mathbf{r} - \mathbf{q}_a}), \quad (1.41)$$

where the exponents $\zeta_{l,n}$ are variational parameters. The quantum numbers $m \in [-l; l]$ and $n \in [1; n_l]$ describe the angular momentum, while n_l represents the number of gaussians for each angular momentum shell. $Y_{l,m}(\Omega)$ are the cubic harmonics defined as real-values linear combinations of spherical harmonics.

Once this localized gaussian basis set is specified, one can expand the AGP functions in terms of atomic orbitals

$$\Phi(\mathbf{r}_i, \mathbf{r}_j) = \sum_{a,b}^{N_{\text{at}}} \sum_{\mu,\nu}^{N_{\text{basis}}} \lambda_{\mu,\nu}^{a,b} \Psi_\mu^{\text{AGP},a}(|\mathbf{r}_i - \mathbf{q}_a|) \Psi_\nu^{\text{AGP},b}(|\mathbf{r}_j - \mathbf{q}_b|), \quad (1.42)$$

where $\Psi_\mu^{\text{AGP},a} \equiv G_\mu^a$ (see Eq. (1.41)), and N_{basis} is the number of elements constituting the primitive basis of the AGP part of the WF. In the above equation, the $\mathbf{\Lambda} = \{\lambda_{\mu,\nu}^{a,b}\}$ matrix gives the strength of the valence bond between two atoms within the system restricted to a specific element of the gaussian basis set, giving an accurate picture of the chemical bonds existing in the system. Moreover, it is clear that for an extended system, many matrix elements of $\mathbf{\Lambda}$ will vanish, increasing the numerical efficiency of the QMC calculation by exploiting the *local* nature of the basis set. This point is of paramount importance for this thesis and thus will be discussed in more detail in Chapter 3.

Furthermore, it is also possible to perform a diagonalization of the AGP matrix $\mathbf{\Lambda}$ to obtain an expression for the pairing functions $\Phi(\mathbf{r}_i, \mathbf{r}_j)$ involving the *molecular orbitals* of the system:

$$\Phi(\mathbf{r}_i, \mathbf{r}_j) = \sum_k^{N_{\text{MO}}} \lambda_k^{\text{MO}} \phi_k^{\text{MO}}(\mathbf{r}_i) \phi_k^{\text{MO}}(\mathbf{r}_j). \quad (1.43)$$

By retaining only the largest eigenvalues λ_k^{MO} , one captures the relevant electronic correlations inside the system at a reasonable computational cost. It can also be proven that keeping the first $N^{\text{MO}} = N/2$ eigenvalues of the AGP matrix, the determinant of Eq. (1.40) reduces to the SD WF. The uncorrelated HF ansatz is therefore contained within the AGP wavefunction

as a special case.

Thus, the AGP part of the WF function can be seen either as a SD in its lowest level expansion, or as a linear combination of SDs. The AGP ansatz is thus particularly suited to capture the *static* electron correlation, whenever nearly degenerate low-lying orbitals are present in the system.

1.2.2.2 The Jastrow factor

The Jastrow factor, in contrast to the AGP part of the WF, is a function of electron-electron separation tailored to deal with the *dynamical* correlation of the electrons. Since this type of correlation is mainly induced by charge fluctuations of the system, the Jastrow factor is essential to include the VdW effects in the total electronic energy^[78]. The role of the Jastrow factor is also to strongly limit the double occupation of orbitals, accordingly with Pauli's exclusion principle. Moreover, it has to fulfill Kato's cusp conditions at electron/ion and electron/electron coalescence points^[79].

The functional form of the Jastrow factor is often written as a product of three terms

$$J(\mathbf{r}_1, \dots, \mathbf{r}_N) = J_1 J_2 J_3, \quad (1.44)$$

where the spin of electrons have been removed for simplicity (this case will be discussed later in the thesis, in the Chapter 5). The one-body term J_1 reads as

$$J_1 = \exp \left(- \sum_i^N \sum_j^{N_{\text{at}}} (2Z_j)^{3/4} u \left((2Z_j)^{1/4} |\mathbf{r}_i - \mathbf{q}_j| \right) \right), \quad (1.45)$$

with Z_j is the atomic charge, $u(|\mathbf{r} - \mathbf{q}|) = \frac{1 - e^{-b|\mathbf{r} - \mathbf{q}|}}{2b}$ and b is a variational parameter. This form is chosen to satisfy the aforementioned electron-ion Kato cusp conditions, used to deal appropriately with the diverging electron-nucleus Coulomb potentials at short distances. During this thesis, since we study only neutral or protonated water clusters, we keep the bare Coulomb potential only for hydrogen, while for the oxygen atoms we replace it with the Burkatzki-Filippi-Dolg (BFD) pseudopotential^[80], which is smooth at the electron-ion coalescence points. Thus, J_1 is applied only to the hydrogen atom.

Similarly, the electron-electron cusp conditions for antiparallel spin electrons are dealt with by the two-body Jastrow factor

$$J_2 = \exp \left(\sum_{i < j}^N u(|\mathbf{r}_i - \mathbf{r}_j|) \right), \quad (1.46)$$

where u is a function of the same form as in Eq. (1.45), but with a different variational parameter. Finally, many-body correlations and thus VdW physics are included in the remaining

part of the Jastrow factor

$$J_3 = \exp \left(\sum_{i < j}^N \Phi_J(\mathbf{r}_i, \mathbf{r}_j) \right), \quad (1.47)$$

with

$$\Phi_J(\mathbf{r}_i, \mathbf{r}_j) = \sum_{a,b}^{N_{\text{at}}} \sum_{\mu,\nu}^{N'_{\text{basis}}} g_{\mu,\nu}^{a,b} \Psi_{\mu}^{J,a}(\mathbf{r}_i - \mathbf{q}_a) \Psi_{\nu}^{J,b}(\mathbf{r}_j - \mathbf{q}_b), \quad (1.48)$$

where $\Psi_{\mu}^{J,a} \equiv G_{\mu}^a$ (see Eq. (1.41)), and $\mathbf{G} = \{g_{\mu,\nu}^{a,b}\}$ is the mathematical equivalent of the AGP matrix $\mathbf{\Lambda}$ for the Jastrow part of the WF. N'_{basis} is the number of primitive elements constituting the basis set of the Jastrow. The impact of N'_{basis} and the selected basis set elements on the interaction between water molecules in gas phase will be discussed in this thesis in the Chapter 5. From Eq. (1.48), it is apparent that $\Phi_J(\mathbf{r}_i, \mathbf{r}_j)$ correlates electrons sitting on different atoms a and b . In this way, interatomic-induced polarization effects are included in the WF. Therefore, J_3 is an essential ingredient to include rigorously polarizability and dispersion contribution in the QMC trial WF. Let us also remark that there is a large flexibility^[81] on the way one decides to write the QMC WF in both Jastrow and determinantal parts. The degree of sophistication of the WF must be a good compromise between computational efficiency and target accuracy, required for the type of problem we want to solve. Indeed, the WF evaluation is the most demanding task of the QMC approach. In our case, we aim at resolving tiny energy differences (beyond the chemical accuracy) so we have to be very cautious by systematically checking that our QMC WF provides relevant energies and geometries compatible with the physics of PT in water clusters.

1.2.3 Variational Monte Carlo

Variational Monte Carlo (VMC) is one of the simplest QMC approach and has been applied for the first time on fermionic systems in the seminal work of Ceperley *et al.* in the late 70s^[82]. Let us consider again the QMC antisymmetric trial WF whose expression have been detailed in the previous pages of this manuscript $\Psi_T(\mathbf{r}_1, \dots, \mathbf{r}_N)$. The electronic energy, defined as the quantum expectation value of the Hamiltonian \hat{H} given by Eq. (1.3) over the trial WF Ψ_T reads as

$$\langle \hat{H} \rangle = E_{\text{VMC}} = \frac{\int d\mathbf{R} \Psi_T^*(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R})}{\int d\mathbf{R} |\Psi_T(\mathbf{R})|^2}, \quad (1.49)$$

where $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ represent the coordinates of the QMC walker in the space spanned by electronic configurations and the spin has been omitted for simplicity. The variational principle states that the VMC energy defined in Eq. (1.49) is an upper bound to the true ground state energy of the system: $E_{\text{VMC}} \geq E_0$. Eq. (1.49) can be then rewritten under the following form

$$\begin{aligned}
 E_{\text{VMC}} &= \int d\mathbf{R} \frac{|\Psi_T(\mathbf{R})|^2}{\int d\mathbf{R}' |\Psi_T(\mathbf{R}')|^2} \frac{\hat{H}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} \\
 &= \int d\mathbf{R} \pi(\mathbf{R}) e_L(\mathbf{R}) = \langle e_L \rangle \geq E_0,
 \end{aligned}
 \tag{1.50}$$

where $\langle \cdot \rangle$ denotes an equilibrium average. In the above equation, we demonstrate that it is possible to compute the VMC energy E_{VMC} as a statistical average of the *local energy*

$$e_L = \frac{\hat{H}\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}
 \tag{1.51}$$

thanks to walkers sampling the probability density $\pi(\mathbf{R}) = \frac{|\Psi_T(\mathbf{R})|^2}{\int d\mathbf{R}' |\Psi_T(\mathbf{R}')|^2}$. This approach is known as the *importance sampling* since it favors the QMC random walk into the electron configurations space near regions where the amplitude of the QMC trial WF is high ($|\Psi_T|^2$).

Let us notice that if the trial QMC WF Ψ_T is an exact eigenstate of the quantum Hamiltonian \hat{H} , the local energy e_L is a constant since it is equal to the ground state energy of the system, independently of the configuration explored by walkers. Therefore, the closer the trial WF Ψ_T to $|\Psi_0\rangle$, the smaller the fluctuations of the estimated VMC energy E_{VMC} . This well-known phenomenon is called *zero variance property* and is extremely important to ensure the efficiency of any QMC simulation. This implies that a very careful optimization of the WF must be done before estimating the energy and other observables by a VMC calculation.

The probability density $\pi(\mathbf{R})$ is sampled using the Metropolis-Hastings algorithm first introduced by Metropolis in 1953^[83] and then generalized by Hastings 17 years later^[84]. This algorithm, also used in classical MC simulations, enables to sample any unknown probability distribution by generating a memoryless process in the $3N$ -configurational space of electronic positions. Such a random walk in which each new configuration only depends on the previous one is called a *Markov chain*. In the TurboRVB code, at each step of the Markov process, a new configuration is created by performing a single electron move. The move is then accepted or rejected by application of the Metropolis' rule^[83]. In a practical way, the MC displacement is calibrated to obtain an optimal acceptance ratio of the MC samples (about 60% of accepted moves). The efficiency of the MC sampling can also be increased if the amplitude of the single electron move depends on its distance with the nearest nucleus: if the electron is far from an ion, it moves with a larger step than the one of an electron closer to a nucleus. Moreover, when the system is composed by two distant fragments, as it will be the case in the Chapter 5 where the dissociation of the water dimer will be studied, instantaneous moves from one fragment to the other one are allowed after a given number of MC samples. This avoids to "freeze" the random walk into the configurational space of electronic positions in a confined region of the space, thus alleviating ergodicity issues.

Finally, the VMC energy E_{VMC} is straightforwardly computed as we accumulate statistics

of its local estimator, namely the local energy, as defined in Eq. (1.50):

$$E_{\text{VMC}} = \langle e_L \rangle = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{gen}}} e_L(\mathbf{R}_i), \quad (1.52)$$

where we recall N_{gen} is the number of QMC generations. Analogously, any operator \hat{O} representing a physical quantity (the oxygen-oxygen distance for instance) can be evaluated as in Eq. (1.52)

$$O_{\text{VMC}} = \langle O_L \rangle = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{gen}}} O_L(\mathbf{R}_i), \quad (1.53)$$

with $O_L = \frac{\hat{O}\Psi_T}{\Psi_T}$ being the local operator. By virtue of the central limit theorem, this estimator of the true expectation value of \hat{O} is totally free of biases. The corresponding unbiased estimator of the variance $\sigma^2[O_{\text{QMC}}]$ is given by

$$\sigma^2[O_{\text{QMC}}] = \frac{\tau_C}{N_{\text{gen}}} \sigma^2[O_L] = \frac{\tau_C}{N_{\text{gen}}} (\langle O_L^2 \rangle - \langle O_L \rangle^2), \quad (1.54)$$

where $\sigma^2[O_L]$ represents the statistical fluctuations of the local measurements of O_L . τ_C is the *autocorrelation time* of MC iterations, defined as

$$\tau_C = 1 + \frac{2}{\sigma[O_L]N_{\text{gen}}} \sum_{i < j} \text{Cov}(O_L(\mathbf{R}_i), O_L(\mathbf{R}_j)), \quad (1.55)$$

where $\text{Cov}(O_L(\mathbf{R}_i), O_L(\mathbf{R}_j)) = \langle (O_L(\mathbf{R}_i) - \langle O_L \rangle)(O_L(\mathbf{R}_j) - \langle O_L \rangle) \rangle$ is the *covariance*. τ_C is equal to the unit in the ideal case of completely uncorrelated measurements, which is in practice never the case. We thus have to perform block averages (*blocking technique*) of the Eq. (1.53) not only to obtain a reliable VMC estimation of the desired observable but also to estimate correctly its variance by significantly reducing τ_C in Eq. (1.54).

As already mentioned, the major advantage of the QMC approach over deterministic methods is thus its ability to cope with the high dimensionality of electronic integrals at a reasonable computational cost. Indeed, the central limit theorem states that the intrinsic statistical error σ_{QMC}^2 (Eq. (1.54)) on the estimated integral is independent of the dimensionality of the problem. This variance can be simply reduced by increasing the number of MC iterations to reach the desired accuracy, and a compromise must be found between accuracy and savings of computational resources.

To summarize, the guideline to perform accurate VMC calculations on physical and chemical systems, such as water clusters, is extremely simple. The very first step consists in building the WF. During this phase, we specify the mathematical form of Ψ (JSD or JAGP), the basis sets for both the Jastrow and the determinantal parts of the WF, the geometry of the system and the use of pseudopotentials for the heaviest atoms (oxygen in our case)^[80]. Then, we perform a DFT calculation, using the LDA functional, to fill the single or multiple determinants with KS orbitals, which constitute a good starting point for the optimization of the WF. Then, the QMC function is optimized using minimization procedures such as Stochastic Reconfiguration

(SR)^[85] and Stochastic Reconfiguration with Hessian accelerator (SRH)^[86] methods. From the practical point of view, this is the most delicate step in the QMC calculation. Technical details about these minimization techniques are given in Appendix C. Finally, once the WF is fully optimized and as close as possible to the ground state WF, $|\Psi_0\rangle$, of the system, we can estimate physical quantities by means of VMC statistical averages.

During the 2000s, VMC has been mostly used to study small and neutral water clusters (from the monomer to the hexamer). Looking at Ref. 87, we have the confirmation that the VMC energy of both the water monomer and dimer strongly depends on the quality of the WF. Indeed, there is a significant gain of 26 mHa (16.3 kcal/mol) in the computed energy for the water monomer when many-body corrections are included in the determinant. Sterpone and his colleagues demonstrated the ability of VMC to reproduce properly the dissociation energy curve of the bonding water dimer with a reasonable, albeit slightly underestimated (about ~ 0.5 kcal/mol), value of the dissociation energy D_e . Furthermore, thanks to the great flexibility of the JAGP WF given in Eqs. (1.40) and (1.42), they have been able to estimate the covalent (1.1(2) kcal/mol) and the correlated VdW fluctuations (1.5(2) kcal/mol) contributions to the total binding energy^[88].

More recently, Zen and coworkers made an extensive study on the role of the WF ansatz and the basis set size to properly describe the properties of the water molecule^[89]. They have reported the following general hierarchy for the QMC WF JDFT < JSD < JAGP, which clearly indicates that JAGP WF seems to be the more suitable to study water. The accuracy of JAGP-VMC calculations for the dipole are in good agreement with CCSD results, but less accurate than CCSD(T) as expected which suggests VMC is "between" these two approaches for water. Protonated dimer, namely the Zundel ion H_5O_2^+ has been studied very recently in our research team, which is the starting point of this thesis^[72]. Indeed, it has been demonstrated that obtained geometries for the Zundel complex at the VMC level are in excellent agreement with reference CCSD(T) calculations and the proton static barriers computed by VMC are slightly overestimated (about ~ 0.7 kcal/mol for $d_{\text{OO}} = 2.7$ Å) but reasonable.

To conclude, the VMC method is a simple stochastic method which has proven its efficiency to describe a large variety of physical and chemical systems, including neutral or charged water clusters, with an accuracy close to the most sophisticated quantum chemistry calculations. Therefore, thanks to its milder scaling with the system size, VMC appears to be an ideal candidate to give a correct enough PES to solve the PT problem in liquid water and this point will be largely exploited in this thesis (see Chapters 4 and 6). However, we will see in the following Subsection how to improve the VMC results via another stochastic method, namely the *Diffusion Monte Carlo* (DMC).

1.2.4 Diffusion Monte Carlo and Lattice Regularized Diffusion Monte Carlo

1.2.4.1 Diffusion Monte Carlo

Contrary to VMC, DMC is a projector technique in which the trial QMC WF is projected on the ground state WF of the system^[90]. Indeed, one can obtain the exact ground state energy

of the E_0 via the following *mixed* estimator of the quantum Hamiltonian \hat{H} :

$$E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle}. \quad (1.56)$$

However, the main issue arising from Eq. (1.56) is the need to know the exact ground state WF Ψ_0 . Let us recall the time-dependent Schrödinger equation written in the imaginary time $t \rightarrow it$

$$-\frac{\partial \Psi(\mathbf{R}, t)}{\partial t} = (\hat{H} - \Lambda) \Psi(\mathbf{R}, t), \quad (1.57)$$

where Λ is an energy offset that is for the moment left arbitrary. It is straightforward to obtain the formal solution of Eq. (1.57) given by

$$\Psi_G(\mathbf{R}, t) = e^{-(\hat{H}-\Lambda)t} \Psi_T(\mathbf{R}, t). \quad (1.58)$$

Using the spectral decomposition of the *guiding* WF Ψ_T , it is easy to show that the long-time behavior of the projected WF Ψ_G is dominated by the Hamiltonian lowest-energy eigenstate Ψ_0 having a non zero overlap with the initial Ψ_T . Consequently, in the infinite time limit, one can compute the DMC energy as the following mixed estimator

$$\begin{aligned} E_{\text{DMC}} &= \frac{\langle \Psi_G | \hat{H} | \Psi_T \rangle}{\langle \Psi_G | \Psi_T \rangle} = \int d\mathbf{R} \frac{\Psi_G^*(\mathbf{R}, t) \Psi_T(\mathbf{R})}{\int d\mathbf{R}' \Psi_G^*(\mathbf{R}', t) \Psi_T(\mathbf{R}')} \frac{\hat{H} \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} \\ &= \int d\mathbf{R} \bar{\pi}(\mathbf{R}, t) e_L(\mathbf{R}) = \langle e_L \rangle. \end{aligned} \quad (1.59)$$

The above equation is the DMC version of Eq. (1.50) where we have operated the substitution $|\Psi_T|^2 \rightarrow \Psi_G^* \Psi_T$ and the importance sampling is made on the mixed density $\bar{\pi} = \frac{\Psi_G^*(\mathbf{R}, t) \Psi_T(\mathbf{R})}{\int d\mathbf{R}' \Psi_G^*(\mathbf{R}', t) \Psi_T(\mathbf{R}')}$. One can apply Eqs.(1.53) and (1.54) to evaluate the DMC observables and their variance, since Eq. (1.59) also contains the zero variance property.

The only significant difference between DMC and VMC comes from the sampling of the mixed density $\bar{\pi}(\mathbf{R}, t)$ since it cannot be directly interpreted as a probability density because of the sign of $\Psi_G \Psi_T$ (we consider real WF for simplicity without loss of generality). Indeed, if one applies a naive sampling of $\bar{\pi}(\mathbf{R}, t)$, the projected Ψ_G would be bosonic in the long time limit. This would clearly violate Pauli's antisymmetrization principle that imposes the ground state WF of a quantum system to be fermionic. To solve that issue, we apply the very efficient *Fixed-Node approximation* (FN) which consists in restricting the importance sampling in a region of the space of electronic configurations where the Ψ_T and Ψ_G have the same sign^[29]. Such a region is called a *nodal pocket* and the tiling theorem ensures that they all are strictly equivalent and contain all the necessary information about $\bar{\pi}(\mathbf{r}, t)$ ^[91]. Within the FN approximation, it is clear that the projected WF $\Psi_G(\mathbf{R}, t)$ will have the same nodal structure that the guiding WF $\Psi_T(\mathbf{R})$ which is by principle not exact because Ψ_T is not equal to the ground state WF Ψ_0 . The systematic error induced by this hypothesis is the so-called FN error and can be fortunately neglected in our thesis. Indeed, Caffarel *et al.* demonstrated the ability of DMC to reproduce almost exactly the experimentally estimated ground state

Chapter 1. Electronic structure calculations

energy of the water molecule $E_{\text{H}_2\text{O}} = -76.4389 \text{ Ha}^{[92]}$ since they obtained the striking value of $E_{\text{H}_2\text{O}} = -76.43894(12) \text{ Ha}^{[93]}$.

Let us denote $f(\mathbf{R}, t) = \Psi_G(\mathbf{R}, t)\Psi_T(\mathbf{R}) \geq 0 \forall \{\mathbf{R}, t\}$ the unnormalized distribution we want to sample within FN approximation. Its time evolution is dictated by the time-dependent Schrödinger equation which reads as

$$-\frac{\partial f(\mathbf{R}, t)}{\partial t} = \underbrace{-\frac{\Delta_{\mathbf{R}}}{2} f(\mathbf{R}, t)}_{\text{diffusion}} + \underbrace{\nabla_{\mathbf{R}} \cdot (\mathbf{v}_D(\mathbf{R}) f(\mathbf{R}, t))}_{\text{drift}} + \underbrace{\left[e_L(\mathbf{R}) - \left(1 + \frac{\partial}{\partial t} \right) \Lambda \right]}_{\text{branching}}, \quad (1.60)$$

where we denote *drift velocity* $\mathbf{v}_D(\mathbf{R}) = \nabla \ln \Psi_T$. The above equation can thus be interpreted as a master equation driving the diffusion of QMC walkers in the configurational space of electronic positions. The first term of the r.h.s of Eq. (1.60) is a purely diffusive contribution, while the second term takes into account the drift impacting the random walk. Finally, the last term governs the walkers' population by deciding which walker will die or survive during the Markovian dynamics via its *branching* ratio.

Afterwards, we express the analytical solution of $f(\mathbf{R}, t)$ in its integral form as follows

$$f(\mathbf{R}, t) = \int d\mathbf{R}' \mathcal{G}(\mathbf{R}; \mathbf{R}', t) f(\mathbf{R}', 0), \quad (1.61)$$

where the propagator $\mathcal{G}(\mathbf{R}; \mathbf{R}', t)$ is the Green function of Eq. (1.60) which fulfills the initial condition $\mathcal{G}(\mathbf{R}; \mathbf{R}', 0) = \delta(\mathbf{R}, \mathbf{R}')$. An explicit expression of this propagator can be derived by simple application of the Trotter-Suzuki formula^[94,95] for the time evolution of the exponential operator $e^{-\tau \hat{H}}$ for small values of τ . This approximation leads to a time-step error that will be discussed more extensively in a different context in Chapter 3.

From a practical point of view, the DMC simulation starts with a VMC thermalization of the walkers where we set $f(\mathbf{R}, 0) = \Psi_{\text{VMC}}^2$. Then, we simulate the diffusion process described in Eq. (1.60) and we regularly update the walkers' dynamics by killing or generating new walkers via the branching algorithm. The energy shift Λ is modulated on-the-fly to push the random walk in the regions where the mixed probability density $f(\mathbf{R}, t)$ is large. We invite the interested reader to the complete reviews^[91,96] for further details.

The statistical estimation of physical quantities in DMC is a bit more complex than in VMC since the used estimators are no longer pure, but *mixed* due to the sampling of the mixed probability distribution $\bar{\pi}(\mathbf{R}, t)$. Hence, one can rewrite the Eq. (1.53) for the mixed estimator of a given observable O

$$O_{\text{mix}} = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{gen}}} \hat{O}_L(\mathbf{R}_i), \quad (1.62)$$

where $\{\mathbf{R}_i\}$ are sampled configurations from the mixed distribution $f(\mathbf{R})$. It can be shown that this mixed estimator O_{mix} coincides with the "pure" average $O_0 = \frac{\langle \Psi_0 | \hat{O} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}$ if and only

if $[\hat{O}, \hat{H}] = 0$. Otherwise, it is possible to extrapolate the pure, hence the physical, average via the following relation:

$$O_0 \simeq 2O_{\text{mix}} - O_{\text{VMC}} + \mathcal{O}(\delta\Psi^2). \quad (1.63)$$

In the above equation, $\delta\Psi$ simply represents the difference between the true ground state WF Ψ_0 and the QMC trial WF Ψ_T . This implies that the applicability of the formula given in Eq. (1.63) is limited by the quality of the trial WF Ψ_T which means that we have to do a very careful VMC minimization of the WF to obtain satisfactory results.

From a practical point of view, DMC provides excellent results, always in agreement with experimental data and advanced electronic structure calculations. Indeed, already in 1994, DMC calculations carried on rigid water dimer gave reasonable value for both geometries and energetics^[97]. Rigid body DMC has then been used to study tunnelling splitting (*i.e.* energy difference between two states of different symmetry) of the water dimer and trimer^[98]. The DMC dissociation energy of the water dimer is found to be between 5.03(7) and 5.47(9) kcal/mol, which is in excellent agreement with the experimental value $D_e = 5.44 \pm 0.7$ kcal/mol^[87]. DMC calculations also predict good binding energies of excess electrons in the $(\text{H}_2\text{O})_6^-$ cluster^[99]. More recently, Gillan *et al.* studied in detail the six first water clusters $(\text{H}_2\text{O})_n$ for $n = 1, \dots, 6$. They verified that DMC results are clearly superior to DFT results (obtained with hybrid functionals), especially for the description of the various isomers of water hexamer, always in good agreement with CCSD(T) results^[100]. They conclude their work by claiming that DMC can now be used as a benchmark for larger clusters and liquid water.

1.2.4.2 Lattice Regularized Diffusion Monte Carlo

In spite of its extreme accuracy, the use of DMC is still quite limited in practice due to its important computational cost. Indeed, its scale has been proven to be of the order of $\mathcal{O}(Z^{5.5-6.5})$ with the atomic number Z for small and medium size systems^[101,102]. To overcome this bottleneck, the core electrons (of oxygen atoms in our case) need to be replaced by cheap but accurate pseudopotentials. As already mentioned, we use pseudopotentials specifically designed for QMC simulations developed by Burzatki, Filippi and Dolg^[80]. The use of a pseudopotential for the core electrons has a direct impact on the many-body Hamiltonian of the system \hat{H} that now contains an *effective* contribution for the potential energy. The latter is a *semi-local* (it contains both local and non-local parts) and angular momentum dependent operator whose role is to reproduce the core-valence Coulombic repulsion and the orthogonality between core and valence states.

Such an inclusion of non-local pseudopotentials into the quantum Hamiltonian \hat{H} has a dramatic impact within the FN approximation since it seems impossible to constrain the non-local contributions of the pseudopotential within a single nodal pocket. Consequently, one solution consists in neglecting the non-local terms in the imaginary time equation (1.60) using the locality approximation^[103]. Nevertheless, the price to pay is the breaking of the variational

principle since the local Hamiltonian eigenstate can have a lower energy than the true ground state one. To solve that issue, Casula and his colleagues developed the Lattice Regularized Diffusion Monte Carlo (LRDMC) that restaures the variational principle of standard DMC algorithms^[104,105].

In this approach, the Green function driving the propagation of the mixed distribution $f(\mathbf{R}, t)$ in the imaginary time $t \rightarrow it$ is no longer solved by direct application of Trotter-Suzuki formula at short times but is discretized on a grid with lattice parameter a . Hence, it is necessary to discretize the non-local part of the Hamiltonian, which reads as

$$\hat{H}_{\text{NL}}^a = -\frac{1}{2} \sum_{i=1}^N \Delta_{\mathbf{r}_i^a}(\theta_i, \phi_i) + \hat{V}_{\text{NL}} + \mathcal{O}(a^2), \quad (1.64)$$

where $\Delta_{\mathbf{r}_i^a}(\theta_i, \phi_i)$ is the discretized Laplacian and \hat{V}_{NL} is the non-local contribution to the potential coming from the use of a pseudopotential. At each update of the electronic position during the diffusion process of the QMC walkers, the direction of the lattice is chosen with *random* angles θ and ψ generated from a normal distribution. In this way, by repeating this randomization procedure at each step, one explores continuously the space thus ensuring ergodicity with the nearest neighbors.

So as to gain accuracy, we can also discretize the local part of the pseudopotential, the electron-ion electrostatic interaction, by imposing the continuous and the discretized versions of the local energy to be the same $e_L^a(\mathbf{R}) = e_L(\mathbf{R})$. One obtains the following expression for the local potential:

$$\hat{V}_L^a(\mathbf{R}) = \hat{V}_L(\mathbf{r}) - \frac{(\Delta_{\mathbf{r}} - \Delta_{\mathbf{r}}^a)\Psi_T(\mathbf{R})}{2\Psi_T(\mathbf{R})}, \quad (1.65)$$

where \hat{V}_L is the local contribution to the pseudopotential. Finally, the full discretized quantum Hamiltonian reads as

$$\hat{H}^a = \hat{H}_{\text{NL}}^a + \hat{V}_L^a(\mathbf{R}) + \hat{V}_{ee} + \mathcal{O}(a^2), \quad (1.66)$$

where only the electron-electron Coulomb interaction is not discretized. The above equation involves a lattice Hamiltonian defined on a continuous space, which is evaluated by a finite number of matrix elements connecting two electron positions. The DMC time step error is here replaced by the lattice space errors that increases quadratically with the lattice parameter a (see Eq. (1.66)). The smaller the lattice parameter is, the more computationally expensive but accurate the LRDMC simulation is. In practice, the exact result can be extrapolated by performing several LRDMC calculations at various values of a and by taking the limit $a \rightarrow 0$. Finally, the mixed average of the energy is evaluated using the discretized Hamiltonian

$$E_{\text{LRDMC}} = \frac{\langle \Psi_G | \hat{H}^a | \Psi_T \rangle}{\langle \Psi_G | \Psi_T \rangle}. \quad (1.67)$$

As its DMC analogue, the reliability of the LRDMC approach has been assessed by Sterpone *et al.*, who demonstrated that the LRDMC extrapolated ($a \rightarrow 0$) value of the water monomer dipole $\mu_{\text{LRDMC}} = 1.870(10)$ D is in better agreement with the experimental value $\mu_{\text{exp}} = 1.855$ D than within VMC $\mu = 1.890(8)$ D. The LRDMC water dimer binding energy $D_{e,\text{exp}} = 4.9 \pm$

0.1 kcal/mol is also very close to reference quantum chemistry calculations: $D_{e,MP2} = 4.99$ kcal/mol and $D_{e,CCSD(T)} = 5.02$ kcal/mol^[88]. These results will be verified on both the bonding and non-bonding water dimer in the Chapter 5 of this thesis. In the meantime, PT static barriers estimated at various inter-oxygen distances in the Zundel complex are improved in LRDMC with respect to VMC, albeit still slightly overestimated (about ~ 0.3 kcal/mol at $d_{OO} = 2.6 - 2.7\text{\AA}$) compared to CCSD(T)^[72]. However, since we aim at performing MD simulations of small water clusters, which require an important amount of QMC estimations of the electronic energy to evaluate the ionic forces driving the system, the LRDMC approach cannot be employed, due to its excessive computational cost. Consequently, during this thesis, the electronic PES and forces will be evaluated by means of VMC, as it will be discussed in the following Subsection.

1.2.5 Forces evaluation in Quantum Monte Carlo

We derive here the method to evaluate forces by means of QMC approach. Unlike in deterministic DFT, obtaining forces within a stochastic framework is by far more difficult. Still working within the BO approximation in which the ions move on a PES defined by the electronic energy of the system, one can define the $3N_{\text{at}}$ -dimensional force acting on all N_{at} atoms of the system

$$\mathbf{f} = -\nabla_{\mathbf{q}} E_{\text{VMC}}[\Psi]. \quad (1.68)$$

In the above equation, $\nabla_{\mathbf{q}}$ is the gradient relative to cartesian coordinates $\{\mathbf{q}\}$ of the nuclei and $E_{\text{VMC}}[\Psi]$ is the variational energy already defined in Eq. (1.49). Being a functional of Ψ , $E_{\text{VMC}}[\Psi]$ has both an *implicit* and *explicit* dependences arising from:

- the Hamiltonian \hat{H} , defined in Eq. (1.3) that also explicitly depends on the nuclear coordinates $\{\mathbf{q}\}$ as we will see in Chapter 2;
- the QMC WF Ψ that explicitly depends on a set of electronic parameters $\{\boldsymbol{\lambda}\}$ varying implicitly with the ionic positions $\{\mathbf{q}\}$ through the use of a *localized* gaussian basis set (Eq. (1.41)).

Consequently, the local energy $e_L(\mathbf{R}; \mathbf{q})$ defined in Eq. (1.51) also depends on $\{\mathbf{q}\}$ through both the Hamiltonian \hat{H} and the WF Ψ . We can expand Eq. (1.68) into three contributions:

$$\mathbf{f} = \mathbf{f}^{\text{Hell-Fey}} + \mathbf{f}^{\text{Pulay}} + \mathbf{f}^{\boldsymbol{\lambda}}, \quad (1.69)$$

with

$$\begin{aligned} \mathbf{f}^{\text{Hell-Fey}} &= -\frac{\langle \Psi | \nabla_{\mathbf{q}} \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \\ \mathbf{f}^{\text{Pulay}} &= -2 \frac{\langle \Psi | O_{\mathbf{q}} \hat{H} | \Psi \rangle - \langle \Psi | O_{\mathbf{q}} | \Psi \rangle E_{\text{VMC}}}{\langle \Psi | \Psi \rangle}; \quad O_{\mathbf{q}} | \Psi \rangle = \nabla_{\mathbf{q}} | \Psi \rangle, \\ \mathbf{f}^{\boldsymbol{\lambda}} &= \nabla_{\boldsymbol{\lambda}} E_{\text{VMC}} \cdot \nabla_{\mathbf{q}} \boldsymbol{\lambda}. \end{aligned} \quad (1.70)$$

In principle, the term $\mathbf{f}^{\boldsymbol{\lambda}}$ is the most complicated because the derivatives $\nabla_{\mathbf{q}} \boldsymbol{\lambda}$ are very difficult to evaluate. Fortunately, this contribution cancels out when the set of electronic parameters

$\{\lambda\}$ is fully optimized, as the energy of the system becomes minimal. The Hellmann-Feynman (Hell-Fey) term, which is the only component of the QMC generalized force \mathbf{f} in the deterministic case, $\mathbf{f}^{\text{Hell-Fey}}$ is the only contribution which would survive if the Hell-Fey theorem were applicable. This is not the case, because the WF is not an eigenstate of \hat{H} . Nevertheless, when the optimized WF approaches an eigenstate of \hat{H} , the Pulay term gets smaller and the HF term becomes dominant.

Since there is a systematic and intrinsic noise affecting the QMC calculations due to their stochastic nature, it is clear that the Hell-Fey and the Pulay terms of the generalized QMC force have a finite variance. Consequently, the simple computation of such terms by finite differences of the electronic energies between two different ionic configurations is completely inefficient. Indeed, one has to impose small ionic displacements δq to solve tiny energy differences but unfortunately, the energy derivatives diverge as $\frac{1}{\delta q}$ as $\delta q \rightarrow 0$. This is due to the propagation of errors that make the obtained error on the energy difference has the same amplitude that the ionic force we want to compute.

Computing the Hell-Fey and Pulay terms with finite variance in a fast way is of paramount importance to make QMC-based MD simulations feasible. To overcome the aforementioned issue, some technical improvements, such as the correlated sampling (CS) and the space warp coordinate transformation (SWCT)^[106], have been done. A further step in the evaluation of efficient and accurate forces is the algorithmic differentiation (AD), recently introduced by Sorella and Capriotti^[107]. Thanks to the AD, computing all components of the ionic forces is only four times more expensive than the cost of an energy calculation.

Thanks to all the techniques described above, we are able in this thesis to evaluate the QMC forces \mathbf{f} in an affordable computational cost with increasing system size. These forces are however affected by an intrinsic noise that could bias the dynamics of the water clusters under study. In the Chapter 2, we will explain the existing methods to incorporate and control the finite variance of forces into a MD framework.

/

Ion dynamics

Contents

2.1	Molecular Dynamics simulations	38
2.1.1	Time and ensemble averages	38
2.1.2	Equations of motion and Liouvillian formalism	40
2.1.3	The velocity-Verlet algorithm	41
2.1.4	Analytical force fields for water	42
2.2	Finite temperature simulations	43
2.2.1	The canonical ensemble	44
2.2.2	The Langevin thermostat	44
2.2.3	Classical Langevin Dynamics integration schemes	46
2.2.3.1	The Bussi algorithm	47
2.2.3.2	The Attacalite-Sorella algorithm	48
2.2.4	Validation of the Langevin thermostat	51
2.3	Dealing with quantum nuclei	52
2.3.1	Zero Point Energy and Nuclear Quantum Effects	53
2.3.2	The Path Integral approach	54
2.3.3	Path integral Langevin Dynamics	58

UNDERSTANDING the mechanisms by which hydronium H_3O^+ and hydroxyde HO^- ions are transported through water has represented both an experimental and a theoretical challenge for more than two centuries^[108]. The theoreticians have the advantage to use computer simulations to directly investigate, at the microscopic length scale and on very short timescales, the proton motion inside clusters in the gas phase or the bulk liquid. Such kinds of calculations, named Molecular Dynamics (MD) simulations, are performed by keeping constant some selected thermodynamic parameters of the system, such as the temperature. Thanks to the knowledge of the interactions driving the ion dynamics of the studied system, we are able to generate representative configurations of protonated water clusters. Later, we can extract physical observables that are the thermodynamic averages of the configurations visited by the system.

In the perspective of describing quantitatively and accurately proton transfer in water, we have gathered the first essential ingredient in the Chapter 1: the interactions between protons and water molecules are very accurately described by means of *ab initio* methods. Since the proton diffusion is very sensitive to the fluctuations of the H-bond network^[109], controlling the temperature of the simulations is of paramount importance. Besides, in addition to thermal fluctuations, one has to take into account the Nuclear Quantum Effects (NQE) due to the light mass of the hydrogen atom. Such quantum effects for the nuclei are crucial, even at

room temperature^[110].

In this Chapter, we will provide the necessary framework to describe *both* thermal and nuclear quantum effects in aqueous systems. First, we recall in Section 2.1 the general methodology to perform MD simulations, with any force field. Then, we will focus on the Langevin Dynamics (LD), which enables to perform MD simulations at constant temperature (Section 2.2), particularly with stochastic forces. Afterwards, we discuss the importance of NQE in water and aqueous systems in Section 2.3. Finally, we demonstrate the possibility to combine together the Feynman Path Integral (PI) approach (2.3.2) with the Langevin Dynamics to perform Path Integral Langevin Dynamics (PILD) simulations in Subsection 2.3.3.

2.1 Molecular Dynamics simulations

Similarly to the classical Monte Carlo (MC) methods, the MD technique is a way to calculate the equilibrium properties of a given system. Still working under the Born-Oppenheimer (BO) approximation, let us write the classical Hamiltonian operator for the nuclei, which differs to the electronic Hamiltonian given in Eq. (1.3):

$$H(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{N_{\text{at}}} \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}). \quad (2.1)$$

In the above expression, $V(\mathbf{q})$ is the potential energy describing the interactions between particles or ions. It can be described empirically using analytical functions of the interatomic coordinates \mathbf{q} , with adjustable parameters parametrized to reproduce at best the experimental data for a given system. Such an approach is computationally cheap, but very system-specific. This limits its application to a given class of problems. To be more general and *transferable* to any physical or chemical system, it is by far better to evaluate the potential energy term $V(\mathbf{q})$ starting from the first principles of quantum mechanics. In the *ab initio* approaches, one solves the many-body Schrödinger equation at fixed ionic positions \mathbf{q} by means of the techniques seen in Chapter 1, to obtain the electronic energy of the system $E_{\text{elec}}(\mathbf{q})$. The latter defines the PES in which the nuclei move during the dynamics, within the BO approximation.

In Eq. (2.1), we denote $\mathbf{p} \equiv (\mathbf{p}_1, \dots, \mathbf{p}_{N_{\text{at}}})$ the particles momenta ($\mathbf{p}_i = m_i \mathbf{v}_i$ with m_i being the mass of the atom i) and $\mathbf{q} \equiv (\mathbf{q}_1, \dots, \mathbf{q}_{N_{\text{at}}})$ their coordinates. The $6N_{\text{at}}$ -dimensional space in which these coordinates (\mathbf{p}, \mathbf{q}) evolve is referred to as the *phase space*.

2.1.1 Time and ensemble averages

The goal of any MD simulation is thus to generate enough ionic configurations to obtain a full and satisfactory exploration of the phase space. In this case, the dynamics is *ergodic* and one can compute the time average of a generic observable O

$$\langle O \rangle_t = \frac{1}{t} \int_0^t dt' O(t'). \quad (2.2)$$

Eq. (2.2) corresponds to the Boltzmann formulation of statistical mechanics of time averages. In practice, there is another way, mostly used in MC simulations, to compute ensemble averages, which corresponds to the Gibbs' version of statistical mechanics. Nevertheless, only and only if the dynamics is ergodic, the *ergodic theorem* states that both approaches are equivalent:

$$\begin{aligned}\langle O \rangle &= \lim_{t \rightarrow \infty} \langle O \rangle_t = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' O(t') \\ &= \int d\mathbf{p} d\mathbf{q} \rho(\mathbf{p}, \mathbf{q}) O(\mathbf{p}, \mathbf{q}).\end{aligned}\tag{2.3}$$

In the above equation, $\rho(\mathbf{p}, \mathbf{q})$ corresponds to a probability density (similar to $\pi(\mathbf{R}, t)$ in the previous Chapter) that depends on the employed statistical ensemble for the simulation. This point will be discussed more extensively in the next section (Section 2.2). Eq. (2.3) is fundamental since it ensures the exact correspondance between time and ensemble averages, leaving us the possibility to choose the most suitable approach for the considered problem. In practice, the number of MD samples or steps N_{steps} is finite and the computed averages are approximated by

$$\langle O \rangle \simeq \frac{1}{N_{\text{steps}}} \sum_{i=1}^{N_{\text{steps}}} O(\mathbf{p}(t_i), \mathbf{q}(t_i)),\tag{2.4}$$

where $O(\mathbf{p}(t_i), \mathbf{q}(t_i))$ is the instantaneous value of the microscopic estimator of the macroscopic observable O we want to compute. Consequently, similarly to the MC case, the measures are affected by an error bar that decreases as the square root of the number of MD steps.

Therefore, the length of the MD trajectories must be adapted to the considered problem for at least two reasons. First, the number of MD steps N_{steps} must be large enough to ensure a full visit of the accessible phase space by the system under study. This is characterized by a convergence of the static properties, such as the Radial Distribution Functions (RDFs), that will be discussed in the Chapters 4 and 6. Second, due to the variety of interactions occurring in water (covalent bonds, H bonds, VdW interactions), intramolecular and intermolecular vibration modes of water clusters have very different energy and timescales. The minimum number of MD iterations must be set to capture a significant number of "rare" events, related to the softest vibration modes of the system, that can occur during the dynamics. The aforementioned reasons explain why we generated longer MD trajectories (about a factor 1.5) for the protonated water hexamer than for the benchmark tests on the Zundel ion.

In practice, the question of the length of our MD simulations is very important since it seems well established that about 50 ps trajectories are required to obtain converged thermal averages for liquid water, simulating at least 32 water molecules^[111]. Because of the computational cost for the evaluation of the electronic PES by QMC, all the MD simulations carried out during this thesis will be shorter (about 20-25 ps). Therefore, we need to ensure that in spite of the shortness of these trajectories, we do not miss any relevant physics. This issue will be discussed in Chapters 4 and 6.

2.1.2 Equations of motion and Liouvillian formalism

Going back to Eq. (2.4), it is clear that to compute any physical quantity, we only need to explicitly know the momenta (or the velocities) $\mathbf{p}(t_i)$ and the positions $\mathbf{q}(t_i)$ of the ions at each time interval t_i during the dynamics. The phase space distribution $\rho(\mathbf{p}, \mathbf{q})$ defines the probability $d\mathbf{p}d\mathbf{q}\rho(\mathbf{p}, \mathbf{q})$ to find the system in the infinitesimal phase space volume $d\mathbf{p}d\mathbf{q}$. Its time evolution is governed by the *Liouville equation* which reads as

$$\frac{\partial \rho}{\partial t} = -\{\rho, H\}, \quad (2.5)$$

where $\{., H\} = \sum_{i=1}^{3N_{\text{at}}} (\nabla_{\mathbf{p}_i} H \nabla_{\mathbf{q}_i} - \nabla_{\mathbf{q}_i} H \nabla_{\mathbf{p}_i})$ denotes the Poisson bracket. The Liouville equation (2.5) can be rewritten in terms of the Liouville operator $i\mathcal{L}$ as follows

$$\frac{\partial \rho}{\partial t} = -i\mathcal{L}\rho. \quad (2.6)$$

The phase space volume is conserved during the dynamics and its *symplectic* nature is ensured by the Liouville theorem. Therefore, the integration schemes we will develop in Chapter 3 must fulfill this fundamental property. The time evolution of the $6-N_{\text{at}}$ dimensional phase space vector $\mathbf{\Gamma} = \{\mathbf{p}_i, \mathbf{q}_i\}_{i=1, \dots, N_{\text{at}}}$ is given by

$$\frac{d\mathbf{\Gamma}(t)}{dt} = i\mathcal{L}\mathbf{\Gamma}. \quad (2.7)$$

One can integrate the above equation and obtain its formal solution which reads as

$$\mathbf{\Gamma}(t) = e^{i\mathcal{L}t}\mathbf{\Gamma}(0), \quad (2.8)$$

where $\mathbf{\Gamma}(0) = \{\mathbf{p}_i(0), \mathbf{q}_i(0)\}_{i=1, \dots, N_{\text{at}}}$ corresponds to the initial conditions of the dynamics. In the MD runs carried out in this thesis, we often take the zero temperature equilibrium geometry as starting initial positions. The initial velocities are generated randomly with the constraint that the total kinetic energy of the system must be compatible with the target temperature of the simulation in agreement with the equipartition theorem. Such an initialization is straightforward and routinely applied in our MD simulations, using the Box-Muller algorithm^[112].

Afterwards, the phase space vector $\mathbf{\Gamma}$ evolves according to the Hamilton equations of motion, given by

$$\begin{aligned} \frac{d\mathbf{p}_i}{dt} &= \dot{\mathbf{p}}_i = -\nabla_{\mathbf{q}_i} H = \mathbf{f}_i, \\ \frac{d\mathbf{q}_i}{dt} &= \dot{\mathbf{q}}_i = \nabla_{\mathbf{p}_i} H = \frac{\mathbf{p}_i}{m_i} = \mathbf{v}_i. \end{aligned} \quad (2.9)$$

where \mathbf{f}_i are the ionic forces defined in Eq. (1.68) that derive from the PES. Such forces can be either deterministic or stochastic as it will be the case in this thesis since they will be computed by means of QMC methods. For the moment, let us consider for simplicity the more usual deterministic case. In practice, the formal solution proposed in Eq. (2.8) can hardly be

exploited to compute physical quantities since there is no exact mathematical solution to the many-body (N_{at} atoms here) problem. As a consequence, we perform a number of numerical approximations to the propagator $e^{i\mathcal{L}t}$. In the following, we will detail one typical integration scheme of the Hamilton equations of motion.

2.1.3 The velocity-Verlet algorithm

We will here give a derivation of one of the most famous MD algorithms, namely velocity-Verlet, using the Liouvillian formalism. This will serve as a guide when we will later derive an original algorithm for Path Integral Langevin Dynamics (PILD). The Verlet algorithm^[113] has been widely used to perform MD simulations of any physical or chemical system, included the water clusters in the early 1980s^[114]. In the following, we present the velocity-Verlet algorithm, a simple version of the Verlet algorithm relying on the Trotter-Suzuki break-up of the Liouvillian propagator for small time intervals δt

$$e^{i\mathcal{L}_V\delta t} = e^{i\mathcal{L}_p\delta t/2}e^{i\mathcal{L}_q\delta t}e^{i\mathcal{L}_p\delta t/2} + \mathcal{O}(\delta t^2), \quad (2.10)$$

where $i\mathcal{L}_\lambda = \sum_{i=1}^{N_{\text{at}}} \dot{\lambda}_i \nabla_{\lambda_i}$, with λ_i a generic component of the momentum or the coordinate of the particle i . The above equation (2.10) is then a product of translation operators that result in the velocity-Verlet integration scheme, whose elementary steps are listed below:

1. update of the particle momenta after a half time step δt :

$$\mathbf{p}_i(t + \delta t/2) = e^{i\mathcal{L}_p\delta t/2}\mathbf{p}_i(t) = \mathbf{p}_i(t) + \frac{\delta t}{2}\mathbf{f}_i(t); \quad (2.11)$$

2. update the ionic positions at δt :

$$\mathbf{q}_i(t + \delta t) = e^{i\mathcal{L}_q\delta t}\mathbf{q}_i(t) = \mathbf{q}_i(t) + \delta t \frac{\mathbf{p}_i(t + \delta t/2)}{m_i}; \quad (2.12)$$

3. new evaluation of the ionic forces via $\mathbf{f}_i(t + \delta t) = -\nabla_{\mathbf{q}_i}V(t + \delta t)$, which represents the most time-consuming part of the calculation, especially for *ab initio* MD simulations;
4. final update of the particle momenta after the full time step δt :

$$\mathbf{p}_i(t + \delta t) = e^{i\mathcal{L}_p\delta t/2}\mathbf{p}_i(t + \delta t/2) = \mathbf{p}_i(t + \delta t/2) + \frac{\delta t}{2}\mathbf{f}_i(t + \delta t). \quad (2.13)$$

The above steps are repeated N_{step} times to generate a trajectory of total time $t_{\text{traj}} = N_{\text{step}}\delta t$. In deriving the velocity-Verlet algorithm from the Trotter-Suzuki break-up of the Liouvillian, we have used the fact that $e^{a\nabla_x}f(x) = f(x + a)$. The latter equality can be checked by deriving both terms by algebra or by a Taylor expansion.

Unlike its standard Verlet version, velocity-Verlet algorithm is able to provide *simultaneously* both positions and momenta (or velocities) of the system at each time interval t_i , which is necessary to compute observables depending explicitly on both \mathbf{q} and \mathbf{p} , as in Eq. (2.4). Let us notice that this numerical propagation of the phase space vector $\mathbf{\Gamma}(t)$, although accurate,

is not exact and suffers from a time step error which is quadratic (see Eq. (2.10)). In practice, it is desirable to work with the largest possible time step δt to optimize the phase space exploration in the fewest number of MD iterations, so as to save computational resources when the force evaluations are very demanding. Nevertheless, we are limited by the fact that the chosen time step δt has to be smaller than the characteristic period of vibration of the softest (intermolecular) modes and by the intrinsic numerical stability of our integrator. Typically, for classical or *ab initio* MD simulations, the employed time step is about $\delta t = 0.5 - 1$ fs. These considerations will have their importance in our thesis, especially in Chapter 3 when we develop novel integration schemes to perform fully quantum dynamics simulation with QMC forces.

Thanks to the knowledge of the propagator of the equations of motion, we are now able to obtain complete MD trajectories where both momenta and positions are evaluated at each elementary MD iteration, as described above. To generate dynamics sampling representative configurations $\Gamma(t)$ of the true physical system, the forces \mathbf{f}_i acting on each ion must be specified. They are computed as conjugated gradients of the Potential Energy Surface and their reliability strongly depends on the accuracy of the PES. In standard classical MD calculations, the PES is often given by an empirical potential that is parametrized to reproduce at best the experimental behavior of the system. Below, we quickly review some of these force fields.

2.1.4 Analytical force fields for water

Numerous *force field* approaches have been developed for water since more than half a century. Among them, we can cite the Lennard-Jones^[115] and the Coulomb-Buckingham^[116] potentials, that describe the dispersion and repulsion forces between particles. These potentials are the starting point of the simplest water models, where water molecules are considered to be rigid with partial charges attributed to oxygen and hydrogen atoms because of the large electronegativity of the oxygen atom (Single Point Charge or SPC model^[117]). Various flavors have then been designed, for example, by improving the Coulomb contribution to the potential by creating fictitious charges on various sites such as TIP3P/TIP4P^[118] or TIP5P^[119]. The five-sites model TIP5P predicts very good water densities ρ for temperatures ranging from 235 to 370 K. These non-polarizable force fields give quite reasonable values of the dipole moment μ , the average density ρ , and the self-diffusion constant D . The RDFs are however not very well reproduced (understructuration) and the dynamics are faster than in the experiments^[120]. Therefore, these models are not accurate enough to provide a satisfactory description of water and proton transfer in gas or condensed phase.

A straightforward improvement to these quite simplistic models is to include polarizability and flexibility into the empirical functional describing the interaction potential between water molecules. This is the case of the SPC/E and flexible SPC models that gives the best bulk water dynamics and structure among these empirical models^[120]. Such explicit representations of the water can already be computationally expensive for very large system made of thousands of atoms evolving along very long trajectories (more than 10 ns). To solve that issue, hybrids models with an implicit description of the surrounding water molecules can be applied, but will not be detailed here. We invite the interested reader to read the complete

and recent review of Skyner *et al.* for further details^[121].

In this thesis, we do not aim at studying very large aqueous systems to model chemical or biological processes. On the contrary, we make the choice to restrict our study to small neutral or protonated water clusters. Indeed, in such finite systems, there is no further bias due to possible finite-size errors made during a MD simulation, as it could be the case at the hydrodynamic limit of liquids^[122]. Besides, small water clusters have the significant advantage to reduce the number of possible scenarios for PT, making possible a clearer and more detailed analysis of each elementary process driving an effective PT. Finally, the energetics and PT barriers are very subtle to capture because of the flatness of the electronic PES for such a phenomenon. Therefore, using advanced electronic structure calculations detailed in the first Chapter is mandatory. In this case, the ionic forces \mathbf{f}_i are evaluated in a rigorous way from first principles calculations during the so-called *ab initio* MD. By construction, they are much more accurate and reliable, but also more computationally expensive, especially in the QMC case. Thus, for the time being, this class of MD simulations is limited to small systems (about 50 atoms maximum) with rather short trajectories (maximum 100 ps).

To summarize, MD simulations constitute an intuitive and efficient tool to characterize at the microscopic scale the elementary steps of the PT mechanism on a large variety of chemical or biological systems. Such calculations are complementary to experiments since they can confirm the experimental data and they are also able to predict the behavior of a system that has not been investigated experimentally under specific conditions. This is, for instance, the case of gas phase water clusters, where it can be difficult to perfectly isolate the desired clusters one wants to analyze. Indeed, they are often weakly bonded to rare gases such as neon to make easier the IR spectrum detection^[123]. Moreover, experiments are usually carried out at room temperature. Computer simulations have a greater flexibility in choosing the thermodynamic parameters. This is an advantage for the theoreticians who can explore more easily the phase diagram of the system, by choosing which thermodynamic parameters should be kept constant during the MD simulation. This arbitrary choice defines the *thermodynamic ensemble* in which we will sample the probability density $\rho(\mathbf{p}, \mathbf{q})$ given in Eq. (2.3). In the following Section, we will describe the case where the temperature of the system is supposed to be constant along the MD trajectory.

2.2 Finite temperature simulations

Constant temperature MD simulations are routinely used to predict the microscopic behavior of biological or chemical systems as a function of the temperature. This implies the use of a *thermostat* that regulates the temperature of the system during the dynamics. Many thermostating methods have been developed so far and we have to select the more appropriate strategy for the problem we want to tackle. This point is far from being trivial because each thermostat has its own advantages and drawbacks and their relative impact on equilibrium or non equilibrium properties is still discussed^[124].

Fixing the temperature of the system is not sufficient to fully determine its thermodynamic properties using statistical mechanics. Indeed, three thermodynamic parameters are required

to characterize the thermodynamic state of the system and they define the *thermodynamic ensemble* in which the system evolve. In this thesis, we aim at simulating isolated water clusters in the gas phase. In this case, the most appropriate ensemble is the canonical one, which is by far the most used ensemble for MD simulations of any chemical systems. For the specific case of water, it appears to be more reliable than the microcanonical ensemble^[125]. Its associated thermodynamic potential is the Helmholtz free energy $F(N_{\text{at}}, T) = U - TS$. We will detail in the following Subsection (see 2.2.1) the expression of its probability density $\rho(\mathbf{\Gamma})$ and the technique we will employ to sample it.

2.2.1 The canonical ensemble

The canonical ensemble assigns the following probability density

$$\rho(\mathbf{\Gamma}) = \frac{e^{-\beta H(\mathbf{\Gamma})}}{Q(N_{\text{at}}, T)}, \quad (2.14)$$

where $\beta = \frac{1}{k_B T}$ is a thermal prefactor and $Q(N_{\text{at}}, T)$ is the partition function of the system, defined by

$$Q(N_{\text{at}}, T) = \int d\mathbf{\Gamma} e^{-\beta H(\mathbf{\Gamma})}. \quad (2.15)$$

The canonical partition function $Q(N_{\text{at}}, T)$ is related to the Helmholtz free energy via

$$F(N_{\text{at}}, T) = -k_B T \ln Q(N_{\text{at}}, T). \quad (2.16)$$

Using Eq. (2.16), it is straightforward to obtain the thermodynamic properties of the system, such as its internal energy U , through the following thermodynamic derivative

$$U = -\frac{\partial}{\partial \beta} \ln Q(N_{\text{at}}, T). \quad (2.17)$$

Response functions, such as the heat capacity C_v can be easily deduced by deriving Eq. (2.17) with respect to the temperature of the system.

In the following, we will detail the employed method, used in this thesis, to control the temperature of our MD simulations, Since we employ a stochastic method to compute the forces \mathbf{f}_i acting on the ions during their dynamics, there is a further bias of the system temperature arising from the intrinsic error of the QMC forces. This point thus requires attention in the choice of the thermostating method that will be adopted.

2.2.2 The Langevin thermostat

Many thermostating schemes have been developed so far to control the temperature of MD simulations with *deterministic* forces. Among them, we can cite the Andersen^[126], the Berendsen^[127], and the Nosé-Hoover^[128,129] thermostats. Let us also mention the very efficient Stochastic Velocity Rescaling (SVR) thermostat developed by Bussi and coworkers^[130], also adapted to perform isothermal-isobaric (N_{at}, P, T) simulations^[131]. These thermostating strategies have been widely used to perform MD simulations of bulk water at constant temperature^[132,133] within a *deterministic* framework.

As already mentioned, the ionic forces driving the dynamics of the ions are computed by QMC, implying that they are *stochastic* by nature. Consequently, the aforementioned thermostats can no longer be used, since they are not built to compensate the intrinsic noise affecting the ionic forces and the temperature of the system would be biased. Incorporating noisy forces into a MD framework is not trivial, and stochastic dynamics must be employed.

The main representative of this class of MD simulations is the *Langevin approach*. From the beginning, this strategy relies on the idea that a physical or chemical system is, in practice, never isolated. Therefore, the particles constituting the system stochastically interact with neighboring atoms composing the solvent and are consequently driven by a *Brownian motion*. The Langevin approach is based on the mathematical idea that one can construct a mapping between this many-body problem and a Langevin Dynamics (LD) where the degrees of freedom related to the solvent have been replaced by a stochastic (Wiener) process. In our case, we use LD for other purposes since we plan to simulate isolated water clusters in gas phase. However, this framework turns out to be the more appropriate to deal with the intrinsic noise affecting the ionic forces computed by QMC, keeping the system at a constant temperature. The Langevin equation is given by

$$\begin{aligned}\dot{\mathbf{p}}_i(t) &= - \underbrace{\boldsymbol{\gamma}\mathbf{p}_i(t)}_{\text{dissipation}} + \mathbf{f}_i(\mathbf{q}(t)) + \underbrace{\boldsymbol{\eta}_i(t)}_{\text{fluctuation}}, \\ \dot{\mathbf{q}}_i(t) &= \frac{\mathbf{p}_i}{m_i}.\end{aligned}\tag{2.18}$$

This equation is slightly more complex than the Hamilton equations of motion (see Eq. (2.9)) since it contains two additional terms. Indeed, to include the temperature of the system during the dynamics, it is necessary to account for thermal fluctuations of the particles that may collide each other. Therefore, this *dissipative* contribution is compensated by the presence of random forces $\boldsymbol{\eta}_i(t)$ that stochastically drive the ion dynamics.

The Langevin *damping matrix*, $\boldsymbol{\gamma} = \{\gamma_{ij}\}_{i,j=1,\dots,N_{\text{at}}}$ can be either diagonal or non diagonal, depending on the complexity of the process we are studying. For an isotropic system constituted by identical particles moving under the action of deterministic forces, the friction matrix can be described via a single constant γ . On the contrary, when the ionic forces are computed within a stochastic approach, as it is the case in this thesis, the form of the damping matrix is more complex. Consequently, the integration of the Langevin equations of motion defined in Eq. (2.18) is far from being trivial and these mathematical issues will be discussed in the next Subsection 2.2.3 and in Chapter 3, dedicated to the development of novel LD integrators. To ensure a correct sampling of the equilibrium properties, the random forces $\boldsymbol{\eta}_i(t)$ must satisfy $\langle \boldsymbol{\eta}_i(t) \rangle = 0$. Its time autocorrelation function, however, depends on the nature of the noise applied to the system. In most cases, the random forces are taken δ -correlated, according to a gaussian stationary process which verifies

$$\langle \boldsymbol{\eta}_i(t)\boldsymbol{\eta}_i(t') \rangle = g_i\delta(t-t'),\tag{2.19}$$

2.2. Finite temperature simulations

where g_i is a constant. Eq. (2.19) thus defines a *white noise* driving the LD of the system. The value of g , which depends on the Langevin friction γ and the temperature T of the system is imposed by the fact that the fluctuations generated by the action of the random forces $\boldsymbol{\eta}_i(t)$ must compensate the dissipative term encoding the collisions of the system with the heat bath. This mathematical relationship is the Fluctuation-Dissipation Theorem (FDT)^[134], given by:

$$\rho(\omega) = \langle \boldsymbol{\eta}_i(\omega) \boldsymbol{\eta}_i(-\omega) \rangle = \frac{k_B T m_i \gamma}{\pi}, \quad (2.20)$$

where $\rho(\omega)$ is the *power spectrum*¹ of the random force. Taking the Fourier transform of Eq. (2.19) and using Eq. (2.20), one obtains

$$g_i = 2m_i k_B T \gamma \quad (2.21)$$

for the constant prefactor in random force autocorrelations. Let us remark that one is free to apply a *colored noise* thermostat to the system, by modifying the mathematical laws ruling the dynamics of the random forces $\boldsymbol{\eta}_i(t)$.

As water is the solvent of a countless number of chemical or biological reactions^[135], the optimal theoretical framework to model the solvation effects of water is represented by the LD, which have thus been widely used in aqueous systems. For instance, LD calculations are used to characterize the solubility and the dynamics of organic molecules in bulk water^[136,137] or to study protein folding and dynamics^[138,139]. Nevertheless, LD simulations can also be used to study small neutral or protonated water clusters in gas phase since they provide an accurate and efficient sampling of the canonical probability density, without any considerations about the environment. Indeed, the LD approach is an ideal candidate for our QMC-driven MD simulations since the stochastic nature of the dynamics compensates the intrinsic error affecting the VMC forces. By incorporating directly the QMC noise into the random forces within a LD framework accordingly to the FDT, we can control the temperature of our MD simulations.

However, standard integration schemes such as the velocity-Verlet algorithm (see Subsection 2.1.3) cannot be used for the integration of stochastic equations of motion. Consequently, specific algorithms must be devised and will be detailed in the following Subsection.

2.2.3 Classical Langevin Dynamics integration schemes

In this Subsection, we will present two examples of LD algorithms to familiarize ourselves with this fascinating equation bringing a lot of perspectives for mathematicians, physicists, chemists and biologists. Indeed, since its derivation by Paul Langevin in 1908, this stochastic differential equation has been widely studied to derive an efficient integrator of the Langevin equation of motion^[140–145]. Designing an efficient Langevin algorithm is of paramount importance in our case since, due to the cost of QMC evaluations of the PES, we are compelled to work with a large time step δt to limit at most the number of LD iterations N_{steps} .

¹white in this case.

2.2.3.1 The Bussi algorithm

In Ref. 143, Bussi and his colleagues derived an efficient algorithm to propagate the Langevin equations of motion, solving the damping problem even in the difficult case of the high friction limit. We derive it here to familiarize with the integration of the Langevin equations of motion, as the core of this propagator will be used in Chapter 3 to devise novel quantum algorithms. In the following, we will consider, for the sake of simplicity and without loss of generality, that all the particles have the same physical mass m . Including the expression given in Eq. (2.21) into the Langevin equation (2.18), one obtains the following system for the Langevin dynamics:

$$\begin{aligned}\dot{\mathbf{p}}_i(t) &= -\gamma\mathbf{p}_i(t) + \mathbf{f}_i(\mathbf{q}(t)) + \sqrt{2mk_B T\gamma}\boldsymbol{\eta}_i(t), \\ \dot{\mathbf{q}}_i(t) &= \frac{\mathbf{p}_i(t)}{m}.\end{aligned}\tag{2.22}$$

In the above system $\mathbf{f}_i(\mathbf{q}(t))$ is a *deterministic* force while $\boldsymbol{\eta}_i(t)$ is a normalized Wiener white noise. These coupled equations can be formulated in terms of the probability density $P(\boldsymbol{\Gamma}, t)$ that evolves according to the Fokker-Planck equation^[146]

$$\frac{\partial P(\boldsymbol{\Gamma}, t)}{\partial t} = i\mathcal{L}_{\text{FP}}P(\boldsymbol{\Gamma}, t),\tag{2.23}$$

where the Fokker-Planck Liouvillian operator $i\mathcal{L}_{\text{FP}}$ is naturally written as a sum of three contributions

$$i\mathcal{L}_{\text{FP}} = i\mathcal{L}_p + i\mathcal{L}_q + i\mathcal{L}_\gamma.\tag{2.24}$$

The two first components $i\mathcal{L}_p$ and $i\mathcal{L}_q$ have already been introduced in Eq. (2.10). The last part of the Fokker-Planck Liouvillian reads as

$$i\mathcal{L}_\gamma = -\gamma \left(\nabla_{\mathbf{p}}\mathbf{p} + \frac{m}{\beta}\nabla_{\mathbf{p}}^2 \right)\tag{2.25}$$

The Fokker-Planck equation (2.23) can be similarly solved formally to Eq. (2.8). Then, one can apply a double Trotter break-up, analogously to the velocity-Verlet algorithm (see Eq. (2.10)), to obtain the full propagator

$$e^{i\mathcal{L}_{\text{FP}}\delta t} = e^{i\mathcal{L}_\gamma\frac{\delta t}{2}} e^{i\mathcal{L}_p\frac{\delta t}{2}} e^{i\mathcal{L}_q\delta t} e^{i\mathcal{L}_p\frac{\delta t}{2}} e^{i\mathcal{L}_\gamma\frac{\delta t}{2}}.\tag{2.26}$$

Of course, as we will see in the following, this choice of the Liouvillian splitting is not unique and can be adapted to each specific situation. The separation applied in Eq. (2.26) is correct, even in the high friction limit $\gamma \rightarrow +\infty$, provided the MD time step δt is small enough. After some algebra, it is straightforward to demonstrate that the elementary steps of the proposed integrators are:

1. update of the particle momenta, under the action of the thermostat, after a half time step δt (Bussi's scheme):

$$\mathbf{p}_i(t^+) = e^{i\mathcal{L}_\gamma\frac{\delta t}{2}} \mathbf{p}_i(t) = c_1\mathbf{p}_i(t) + c_2\xi_1(t);\tag{2.27}$$

2. update the ionic positions at δt :

$$\mathbf{q}_i(t + \delta t) = e^{i\mathcal{L}_p \frac{\delta t}{2}} e^{i\mathcal{L}_q \delta t} \mathbf{q}_i(t) = \mathbf{q}_i(t) + \frac{\mathbf{p}_i(t^+)}{m} \delta t + \frac{\mathbf{f}_i(\mathbf{q}_i(t))}{m} \frac{\delta t^2}{2}; \quad (2.28)$$

3. new evaluation of the ionic forces via $\mathbf{f}_i(t + \delta t) = -\nabla_{\mathbf{q}_i} V(t + \delta t)$;

4. second update of the particle momenta to conclude the deterministic part of the dynamics:

$$\mathbf{p}_i(t^- + \delta t) = e^{i\mathcal{L}_p \frac{\delta t}{2}} \mathbf{p}_i(t^+) = \mathbf{p}_i(t^+) + \frac{\delta t}{2} (\mathbf{f}_i(\mathbf{q}_i(t)) + \mathbf{f}_i(\mathbf{q}_i(t + \delta t))); \quad (2.29)$$

5. final update of the particle momenta, under the action of the thermostat, after the full time step δt :

$$\mathbf{p}_i(t + \delta t) = e^{i\mathcal{L}_\gamma \frac{\delta t}{2}} \mathbf{p}_i(t^- + \delta t) = c_1 \mathbf{p}_i(t^- + \delta t) + c_2 \xi_2(t + \delta t), \quad (2.30)$$

and so on. In the above thermostating steps 1 and 5, $\xi_1(t)$ and $\xi_2(t + \delta t)$ are two independent gaussian numbers with unitary variance, and the coefficients c_1 and c_2 are:

$$\begin{aligned} c_1 &= e^{-\gamma \frac{\delta t}{2}} \\ c_2 &= \sqrt{(1 - c_1^2) \frac{m}{\beta}}. \end{aligned} \quad (2.31)$$

To summarize, this integration scheme is very intuitive since the combination of the two inner stages in Eq. (2.26) is a simple velocity-Verlet while the outermost components of the propagator correspond to the action of the thermostat. Indeed, $\mathbf{p}_i(t^+)$ in Eq. (2.27) represents the instantaneous momentum of the particle i just after the action of the thermostat which implies that the temperature of the system should be estimated just before the half thermostat iteration giving $\mathbf{p}_i(t^+)$. This strategy is very efficient for deterministic forces. In the case of stochastic (QMC) forces, another algorithm has been proposed by Attaccalite and Sorella.

2.2.3.2 The Attaccalite-Sorella algorithm

In the following, we derive the Attaccalite-Sorella algorithm because it is the very first algorithm specifically devised for the integration of the Langevin equation of motion with noisy QMC forces. At variance with the previous algorithm, it is not based on the Liouvillian formalism. Indeed, the momenta and the positions are evolved in a single operation, thanks to the use of momentum-position correlators. This approach will also be used in Chapter 3 to design the novel classical algorithm.

As described in Chapter 1, due to the stochastic nature of the QMC method, the QMC forces $\mathbf{f}_i(\mathbf{q})$ are affected by an intrinsic error bar. This finite variance, in principle, bias the temperature of the simulation. Let us rewrite the FDT, given by Eq. (2.20), as follows

$$\gamma_{ij} = \frac{\alpha_{ij}}{2m_i k_B T}, \quad (2.32)$$

Chapter 2. Ion dynamics

with k_B the Boltzmann constant, and

$$\boldsymbol{\alpha}_{ij}\delta(t-t') = \langle \boldsymbol{\eta}_i(t)\boldsymbol{\eta}_j(t') \rangle \quad (2.33)$$

is the *force covariance matrix* which simply reduces to a diagonal form for deterministic forces^[147,148]. Indeed, in this case, the ionic forces fluctuations are completely uncorrelated ($\langle \delta\mathbf{f}_i\delta\mathbf{f}_j \rangle = 0$ for $i \neq j$) and a white noise is applied to the system. On the contrary, due to their finite variance, the QMC forces are correlated between different (particles) components ($\langle \delta\mathbf{f}_i^{\text{QMC}}\delta\mathbf{f}_j^{\text{QMC}} \rangle \neq 0$ for $i \neq j$). Consequently, the force covariance and the Langevin damping matrices have non-trivial off-diagonal elements: this situation corresponds to QMC-correlated noise. We point out this *spatially* correlated noise is *white* because of the Markovian nature of the QMC approach. It thus differs from the more frequently used *colored* noise, which is *time* correlated.

The dynamics generated by the stochastic differential equation (2.18) is Markovian, as the noise fluctuations are locally (*i.e.* are non zero only for $i = j$) correlated in time. Indeed, the number of MC samples to evaluate the electronic energy at each nuclear iteration is large enough to fully decorrelate the updated ionic forces $\mathbf{f}_i(\mathbf{q}(t + \delta t))$ from the older ones $\mathbf{f}_i(\mathbf{q}(t))$. Therefore, the Langevin approach can naturally deal with noisy QMC forces, as already shown by Sorella and coworkers in the classical MD framework^[149–151]. Hereafter, we are going to use transformed variables, more convenient to handle, achieved by the following mass scaling:

$$\begin{aligned} q_i &= q_i^0 \sqrt{m_i} \\ p_i &= \frac{p_i^0}{\sqrt{m_i}} \\ \eta_i &= \eta_i^0 \sqrt{m_i} \\ f_i &= \frac{f_i^0}{\sqrt{m_i}}, \end{aligned} \quad (2.34)$$

where $\{q_i^0, p_i^0, \eta_i^0, f_i^0\}_{i=1, \dots, 3N_{\text{at}}}$ are the original coordinates and $\{q_i, p_i, \eta_i, f_i\}_{i=1, \dots, 3N_{\text{at}}}$ denote the transformed ones. After applying the above transformation to Eq. (2.18), the corresponding second order Langevin dynamics reads as:

$$\dot{\mathbf{p}} = -\gamma\mathbf{p} + \mathbf{f}(\mathbf{q}) + \boldsymbol{\eta}(t) \quad (2.35)$$

$$\dot{\mathbf{q}} = \mathbf{p}, \quad (2.36)$$

Eq. (2.35) belongs to a general class of Stochastic Differential Equations (SDEs), whose solution is explicit whenever both the friction matrix and the force are independent of \mathbf{q} ^[152]

$$\mathbf{p}(t') = e^{-\gamma(t'-t)}\mathbf{p}(t) + \int_t^{t'} ds e^{\gamma(s-t')} (\mathbf{f}(s) + \boldsymbol{\eta}(s)). \quad (2.37)$$

Starting from this form, Attaccalite *et al.*^[149] proposed to use this integration algorithm in a QMC-based MD, because in short time interval $t' - t = \delta t$ the q -dependence of the force and the friction can be neglected and can be assumed to remain at their initial values $\mathbf{f}(t)$ and $\gamma(t)$, leading to an explicit integration of the LD at discrete equally spaced times $t_n = n\delta t$:

$$\mathbf{p}_{n+1} = e^{-\gamma_n\delta t}\mathbf{p}_n + \mathbf{\Gamma}(\mathbf{f}_n + \tilde{\boldsymbol{\eta}}) \quad (2.38)$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mathbf{p}_n\delta t, \quad (2.39)$$

where the time evolution has been discretized with time step δt , and the subscripts refer to the corresponding time slice, such that $\mathbf{p}_n = \mathbf{p}(t_n)$, $\mathbf{q}_n = \mathbf{q}(t_n)$, and $\mathbf{f}_n = \mathbf{f}(\mathbf{q}(t_n))$. In the above equations, the $\mathbf{\Gamma}$ matrix is defined as

$$\mathbf{\Gamma} = \boldsymbol{\gamma}_n^{-1}(1 - e^{-\boldsymbol{\gamma}_n \delta t}), \quad (2.40)$$

and it multiplies the forces \mathbf{f} under the approximation that δt is small enough to disregard the time dependence of $\mathbf{f}(\mathbf{q})$ in Eq. (2.37). In the same interval δt , the time dependent stochastic noise $\boldsymbol{\eta}(t)$ is integrated, leading to the stochastic variable $\tilde{\boldsymbol{\eta}}$, whose mean and variance-covariance are

$$\begin{aligned} \langle \tilde{\boldsymbol{\eta}} \rangle &= 0, \\ \langle \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\eta}} \rangle &= k_B T \boldsymbol{\gamma}_n^2 \coth\left(\boldsymbol{\gamma}_n \frac{\delta t}{2}\right), \end{aligned} \quad (2.41)$$

respectively. These relations have been derived directly from the FDT (2.32), with the hypothesis that the fluctuations (2.33) are local in time.

The time step δt and the friction matrix $\boldsymbol{\gamma}$ are algorithmic parameters to be defined in order to carry out the classical LD based on this scheme. If the forces are deterministic, $\boldsymbol{\gamma}$ can take a diagonal form, such that $\gamma_{ij} = \gamma_{\text{BO}} \delta_{ij}$, where γ_{BO} is a value appropriately chosen to optimize the efficiency of the canonical sampling, as we will see in Chapter 3. According to Eq. (2.41), this implies a diagonal $\langle \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\eta}} \rangle$, *i.e.* a non-correlated noise. In the case of noisy forces, $\tilde{\boldsymbol{\eta}}$ can be separated into a contribution $\delta \mathbf{f}$ (*already* included in the QMC forces \mathbf{f}) and $\tilde{\boldsymbol{\eta}}^{\text{ext}}$. The latter contribution has to be explicitly added in Eq. (2.38), such that its covariance added to the external noise variance satisfies:

$$\langle (\tilde{\boldsymbol{\eta}}^{\text{ext}})^T \tilde{\boldsymbol{\eta}}^{\text{ext}} \rangle + \langle \delta \mathbf{f}^T \delta \mathbf{f} \rangle = \langle \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\eta}} \rangle, \quad (2.42)$$

where $\langle \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\eta}} \rangle$ is given by Eq. (2.41) and $\langle \delta \mathbf{f}^T \delta \mathbf{f} \rangle$ is the QMC force covariance, which will have in general non-zero off-diagonal matrix elements. Thus, if the forces are noisy, $\boldsymbol{\gamma}$ is non-diagonal, as already mentioned. In that case, one has to rotate the momenta and positions in the frame which diagonalizes $\boldsymbol{\gamma}$, in order to perform numerically the Markov iteration in Eqs. (2.38) and (2.39).

The previous standard algorithm for QMC-LD dynamics has been further developed in Ref. 150 and later in Ref. 151, to improve the integration of the positions \mathbf{q} in Eq. (2.39), by taking into account the evolution of $\mathbf{p}(t)$ during the time interval δt . We will detail in Chapter 3 an alternative derivation of these algorithms in the case of *both* classical and quantum particles and their efficiency will be compared with standard LD integrators. The presented integrator is, however, already computationally efficient since, in 2015, Zen and coworkers performed the very first QMC-driven LD simulation of bulk water. They simulated 32 molecules into a cubic box with periodic boundary conditions moving in an electronic PES estimated by VMC calculations. The obtained RDFs display a very good agreement with X-ray and neutron scattering experiments, particularly concerning the position of the oxygen-oxygen peak in the radial distribution function, at variance of previous DFT attempts^[151].

2.2.4 Validation of the Langevin thermostat

To further justify the choice to work with a Langevin thermostat, we will compare the efficiency of this thermostating strategy to the Nosé-Hoover thermostat in the specific case of water clusters.

First, let us recall how the temperature of system can be evaluated along a MD simulation. Using the equipartition theorem, the instantaneous temperature can easily be related to the kinetic energy of the system as follows

$$T(t) = \frac{2E_c(t)}{Gk_B}. \quad (2.43)$$

k_B is the Boltzmann constant and G is the number of degrees of freedom of the system (it is equal to $G = 3N_{\text{at}}$ in the absence of holonomic constraints). The classical kinetic energy $E_c(t)$ is simply

$$E_c(t) = \sum_{i=1}^{N_{\text{at}}} \frac{\mathbf{p}_i(t)^2}{2m_i} = \sum_{i=1}^{N_{\text{at}}} \frac{1}{2} m_i \mathbf{v}_i(t)^2, \quad (2.44)$$

and combining Eqs. (2.43) and (2.44), we obtain the following relation for the temperature

$$T(t) = \frac{\sum_{i=1}^{N_{\text{at}}} m_i \mathbf{v}_i(t)^2}{3N_{\text{at}}k_B}. \quad (2.45)$$

As already mentioned, the above equation can be inverted to generate initial velocities in agreement with the desired temperature of the simulation.

We have tested the Nosé-Hoover thermostat for the (N_{at}, T) simulation of the classical Zundel ion with forces computed from the analytic PES, parametrized via CCSD(T) calculations^[65]. The obtained distribution of the instantaneous temperatures $T(t_i)$, imposing the constant target temperature T_{target} , is represented in Figure 2.1 (black curve). The result is compared with a LD generated using a second-order integration scheme^[142]. The red curve corresponds to the Langevin result with the same force field, while the blue curve, more noisy, is the obtained result with QMC forces on a quite small sample. The green curve is the analytic result of the probability $P(T|T_{\text{target}})$ to measure the instantaneous temperature of the system $T(t_i)$ at T when the temperature of the thermostat is T_{target} . The corresponding distribution is

$$P(T|T_{\text{target}}) \propto \sqrt{T}^{3N_{\text{at}}-2} e^{-\frac{3N_{\text{at}}T}{2T_{\text{target}}}}, \quad (2.46)$$

where $N_{\text{at}} = 7$ for the Zundel complex.

It is clear from Figure 2.1 that the Nosé-Hoover thermostat keeps the Zundel ion at the correct ambient temperature on average, but fails to describe properly thermal fluctuations. This result looks independent from the mass of the thermostat (hence the strength of the coupling with the heat bath) and comes from the non ergodicity of the Nosé-Hoover thermostat. This deficiency has been pointed out, even in the case of simple systems such as the harmonic oscillator^[153]. This problem can however be solved by the use of *chains* of thermostats^[154]. On

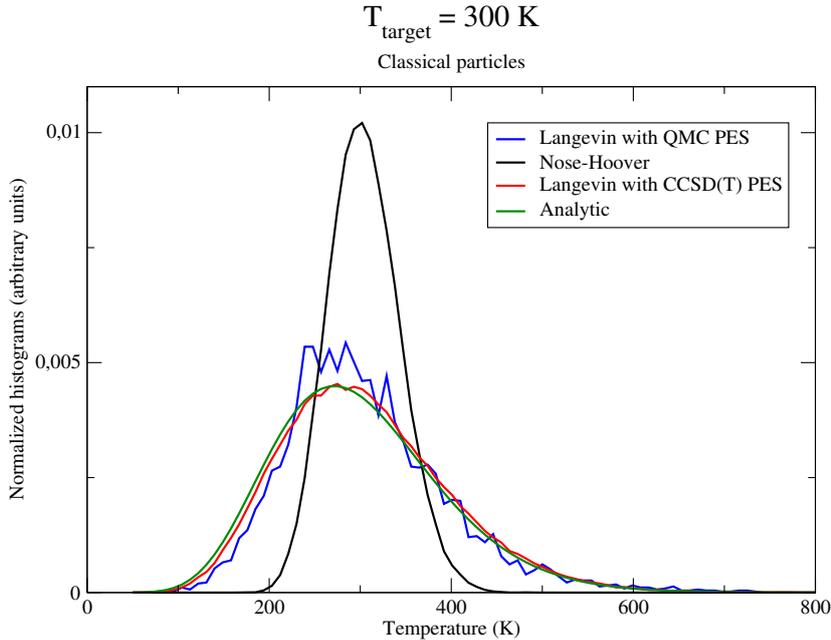


Figure 2.1 – Histograms of the probability distribution to find the Zundel cation at the instantaneous temperature $T(t_i)$, imposing a constant target temperature $T_{\text{target}} = 300 \text{ K}$ for various thermostats. The analytic distribution $P(T|T_{\text{target}})$ (see Eq. (2.46)) is represented in green.

the contrary, the Langevin simulations, either with deterministic or with stochastic forces, give satisfactory results. Therefore, this approach appears to be a promising candidate to perform MD simulations in the canonical ensemble of clusters and finite systems in the gas phase, without a major bias on the dynamics.

To put it in a nutshell, we have demonstrated the ability of the LD framework to perform accurate and efficient constant MD simulations of finite systems at finite temperature, with or without stochastic forces. Within this framework, we are thus able to make an accurate study of the impact of thermal effects in PT mechanisms inside neutral or charged water clusters, using advanced electronic structure calculations to properly probe the energetics of such phenomena. Nevertheless, to obtain quantitative results for the proton dynamics in water clusters, a key ingredient is missing. Indeed, due to its very light mass, the *quantum* nature of the proton must be taken into account, as we will see in the Section 2.3.

2.3 Dealing with quantum nuclei

In order to better understand the reason that justifies the possible significance of the quantum effects within a given system, one has to go back to the definition of the *thermal de Broglie wavelength*. Indeed, this quantity denoted $\Lambda(T)$ roughly defines the average de Broglie wavelength of the gas particles in an ideal gas at the specified temperature:

$$\Lambda(T) = \sqrt{\frac{h^2}{2\pi m k_B T}}, \quad (2.47)$$

where h is the Planck constant and m the mass of the considered particle. Using Eq. (2.47), one finds, at room temperature:

$$\begin{aligned}\Lambda(\text{H}, T = 300 \text{ K}) &= 1 \text{ \AA} \\ \Lambda(\text{O}, T = 300 \text{ K}) &= 0.25 \text{ \AA}.\end{aligned}\tag{2.48}$$

When the thermal de Broglie wavelength Λ is comparable or larger than the characteristic length of the potential confinement, the system behaves quantum mechanically. For water, hydrogen atoms are subjected to sensible variations of the potential for very small displacements (about 0.1 Å). To give a simple macroscopic manifestation of NQE, the heavy water D_2O^2 has higher freezing ($T_f = 276.98 \text{ K}$ or $3.82 \text{ }^\circ\text{C}$) and boiling ($T_b = 374.5 \text{ K}$ or $101.4 \text{ }^\circ\text{C}$) points than the pure water^[155]. Therefore, it is clear that Nuclear Quantum Effects (NQE) of proton in water cannot be neglected, even at room temperature. We will detail in the following how these effects can be described theoretically and be incorporated into a LD framework.

2.3.1 Zero Point Energy and Nuclear Quantum Effects

At variance with classical mechanics, the lowest state of quantum systems has a *finite* energy because of the Heisenberg uncertainty principle. The most common definition of the molecular Zero Point Energy (ZPE) is the energy difference between the vibrational ground state and the lowest point on the Born-Oppenheimer (BO) potential energy surface^[156]. In theoretical calculations, ZPE can be estimated by a direct calculation of the vibrational frequencies of the system since one has

$$ZPE = \frac{\hbar\omega_0}{2},\tag{2.49}$$

where $\hbar = \frac{h}{2\pi}$ is the reduced Planck constant and ω_0 is the frequency of the lowest vibration energy mode. In the case of the hydronium and hydroxyde water clusters, there is some speculation that the lowest vibrational state might lie above the PT barrier^[157,158]. This suggests that, in these systems, ZPE effects tend to provide the required energy to overcome the PT static barrier. Within the transition state theory, it is then possible to directly measure the impact of the ZPE effects on reaction rates for fast PT transfers (about 10-40 fs) over small barriers^[159]. The impact of the ZPE and thermal effects on the kinetics of PT transfer on the protonated water hexamer will be discussed in Chapter 6.

Afterwards, there is another noticeable consequence of the quantum nature of the proton: *proton tunneling* between two adjacent sites. The two available donor and acceptor sites, *i.e.* oxygen atoms in our case, are bounded by a double well potential whose shape, width and height depend on the considered system. It has been established that the probability for a particle to tunnel is inversely proportional to its mass and the width of the potential barrier. A proton is about 2000 times more massive than an electron, so it has a much lower probability of tunneling compared to an electron. Nevertheless, proton tunneling still occurs

² $m_D = 2m_H$, hence $\Lambda(\text{D}) = \Lambda(\text{H})/\sqrt{2}$.

especially at low temperatures and high pressures where the potential barrier is very thin.

Proton tunneling is usually associated to H-bonds that are shared between two proton donor and acceptor water molecules. This is equivalent to a proton resting in one of the wells of a double well potential, as described above. When proton tunneling occurs, the H bond and covalent bonds of the considered complex are switched. The same proton has the same probability of tunneling back to its original site provided that the double well potential is *symmetrical*. Proton tunneling has been first characterized in high pressure ices, in particular in the ice 'VII' phase, which is non molecular but symmetric^[160].

Quantum tunneling has also been observed in small water clusters such as the cyclic water trimer, whose chiral structure is composed by six different enantiomers. Proton tunneling enables conformational changes of this cyclic cluster^[161] and has been confirmed in larger clusters such as (H₂O)₄ or (H₂O)₅^[162]. The physical interpretation of the mechanisms implying proton entanglement and tunneling is still a fascinating, but delicate task for the scientific community. Indeed, let us remind that the Grotthuss mechanism suggests that PT reactions are *concerted*^[163] and it might be difficult to fully distinguish all the necessary conditions to obtain fast and efficient PT reactions. This is particularly true in the case of biological systems, where proton tunneling is known to be crucial to improve the understanding of enzymatic reactions^[164], or DNA base pairs interactions^[165,166].

Zero Point Energy and proton quantum tunneling effects are characteristic signs of the so-called NQE. These effects must be taken into account since they deeply affect the PES and the energetics of PT phenomenon. It is necessary to explicitly include them into our methodology since, by omitting them, one would spoil all the efforts made to improve the electronic description of the problem, as we have seen in Chapter 1. Indeed, the lack of NQE necessarily imply deficiencies in the good description of liquid water RDFs. In Ref. 167, it is demonstrated that, with NQE, the obtained RDFs of liquid water are significantly improved, even with a GGA functional for the electronic part (BLYP). In this thesis, NQE are included via the Feynman Path Integral (PI) approach which is the most used tool to describe such effects in the liquid state.

2.3.2 The Path Integral approach

We present here the Path Integral (PI) method that will be used in this thesis to take into account NQE. Based on the Feynman PI formulation of quantum mechanics, it has experienced a significant interest in the last twenty years. Indeed, this approach enables a rigorous evaluation of the quantum partition function $\mathcal{Z}(N_{\text{at}}, T)$ defined as:

$$\mathcal{Z}(N_{\text{at}}, T) = \text{Tr} \left[e^{-\beta H} \right]. \quad (2.50)$$

$\text{Tr}[\cdot]$ represents the trace of the operator inside the squares and H is the Hamiltonian given in Eq. (2.1). One can rewrite Eq. (2.50) as

$$\mathcal{Z}(N_{\text{at}}, T) = \int d^{3N_{\text{at}}} q \langle q | e^{-\beta H} | q \rangle, \quad (2.51)$$

where $\langle q|e^{-\beta H}|q\rangle$ are diagonal matrix elements in the coordinates space. In the case of quantum particles, the kinetic energy K and potential energy V operators do not commute, implying that we have to apply a Trotter-Suzuki break-up to evaluate these matrix elements. Indeed, using the same factorization as in Eq. (2.10) applied P times, one obtains

$$e^{-\beta H} = e^{-\beta(K+V)} = \lim_{P \rightarrow +\infty} [e^{-\beta V/2P} e^{-\beta K/P} e^{-\beta V/2P}]^P. \quad (2.52)$$

To simplify the derivation, we introduce an operator $\xi = e^{-\beta V/2P} e^{-\beta K/P} e^{-\beta V/2P}$. Combining Eqs. (2.51) and (2.52) and inserting $P - 1$ times the identity operator $I = \int dx |x\rangle\langle x|$ between the P factors of ξ , the quantum partition function reads as

$$\mathcal{Z}(N_{\text{at}}, T) = \lim_{P \rightarrow +\infty} \int d^3 q_2 \dots d^3 q_P \langle q|\xi|q_P\rangle \langle q_P|\xi|q_{P-1}\rangle \dots \langle q_2|\xi|q\rangle. \quad (2.53)$$

The interaction potential V only depends on the atomic coordinates and is consequently diagonal in the coordinate basis. The matrix elements in Eq. (2.53) can be evaluated as

$$\langle q_{j+1}|\xi|q_j\rangle = e^{-\beta V(q_{j+1})/2P} \langle q_{j+1}|e^{-\beta K/P}|q_j\rangle e^{-\beta V(q_j)/2P}, \quad (2.54)$$

for $j = 0, \dots, P-1$. The matrix element implying the kinetic energy operator K is less trivial to evaluate since one has to work in the momentum basis, as follows

$$\begin{aligned} \langle q_{j+1}|e^{-\beta K/P}|q_j\rangle &= \int dp \langle q_{j+1}|p\rangle \langle p|q_j\rangle e^{-\beta p^2/2mP} \\ \langle q_{j+1}|e^{-\beta K/P}|q_j\rangle &= \frac{1}{2\pi\hbar} \int dp e^{ip(q_{j+1}-q_j)/\hbar} e^{-\beta p^2/2mP}, \end{aligned} \quad (2.55)$$

where we have used once again the identity operator I and the knowledge of $\langle q|p\rangle = 1/\sqrt{2\pi\hbar} e^{ipq/\hbar}$. The above equation is simply a gaussian integral and combining it with Eq. (2.54), one obtains

$$\langle q_{j+1}|\xi|q_j\rangle = \left(\frac{mP}{2\pi\beta\hbar^2} \right) e^{-\frac{mP}{2\beta\hbar^2}(q_{j+1}-q_j)^2} e^{-\frac{\beta}{2P}(V(q_{j+1})+V(q_j))}. \quad (2.56)$$

By inserting the above equation into Eq. (2.53), we finally have the expression of the quantum partition function

$$\mathcal{Z}(N_{\text{at}}, T) = \lim_{P \rightarrow +\infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{\binom{P}{2}} \int d^{3f} q \exp \left\{ -\frac{1}{\hbar} \sum_{j=1}^P \left[\sum_{i=1}^{N_{\text{at}}} \frac{mP}{2\beta\hbar} (q_i^{(j+1)} - q_i^{(j)})^2 + \frac{\beta\hbar}{P} V(q_i) \right] \right\}, \quad (2.57)$$

with $f = PN_{\text{at}}$ and the cyclic boundary condition $q_i^{(P+1)} = q_i^{(1)}$ for each particle i . One can interpret the above formula as an exact evaluation of the quantum partition function $\mathcal{Z}(N_{\text{at}}, T)$ via a discretized cyclic path which depicts the quantum delocalization of the nuclei. We can notice that in the high temperature limit ($\beta \rightarrow 0$), one recovers the expression of the classical canonical partition function $Q(N_{\text{at}}, T)$ given in Eq. (2.15). Thus, the high temperature limit of the PI formalism is strictly equivalent to the classical limit.

In practice, the number of slices P is finite and the expression in Eq. (2.57) is approximated

2.3. Dealing with quantum nuclei

using the *quantum-to-classical isomorphism*^[168]. This property states that the description of a true quantum system can be replaced by a fictitious classical path in the quantum imaginary time $\tau = \frac{\beta}{P}$ which reproduces the quantum behavior of the system. This idea is at the basis of the so-called Path Integral Molecular Dynamics (PIMD) and Path Integral Monte Carlo (PIMC) techniques. In this case, the quantum partition function can be evaluated as

$$\mathcal{Z}(N_{\text{at}}, T) = \left(\frac{1}{2\pi\hbar} \right)^P \int d^f \Gamma e^{-\tau H_P(\Gamma)}, \quad (2.58)$$

where $H_P(\Gamma)$ is the quantum-to-classical isomorphism Hamiltonian, given by

$$H_P(\Gamma) = \sum_{i=1}^{3N_{\text{at}}} \sum_{j=1}^P \left(\frac{1}{2m_i} [p_i^{(j)}]^2 + \frac{1}{2} \tilde{\omega}_P^2 \left(q_i^{(j)} - q_i^{(j-1)} \right)^2 \right) + \sum_{j=1}^P V(q_1^{(j)}, \dots, q_{3N_{\text{at}}}^{(j)}). \quad (2.59)$$

In other words, the quantum nuclei are replaced by fictitious classical ring polymers whose beads are connected to each other by harmonic springs with frequency $\tilde{\omega}_P = \frac{P}{\beta\hbar}$. The lower the temperature, the more spatially extended the necklace, as intuitively drawn³ in Figure 2.2.

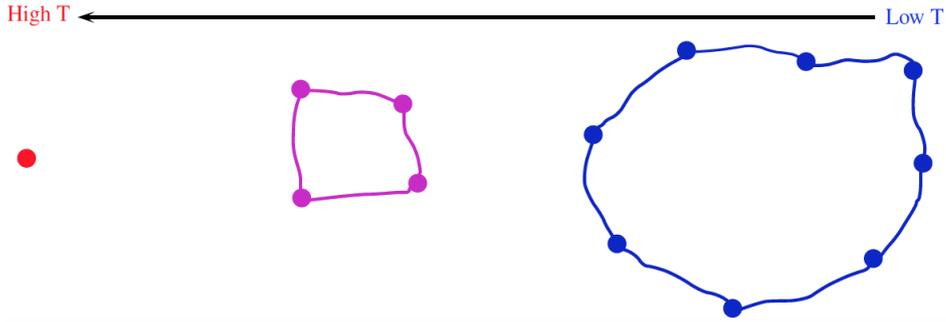


Figure 2.2 – Intuitive representation of the classical ring polymer as a function of the simulation temperature T .

The PIMD method is a way to calculate quantum mechanical properties using the recipes from classical statistical mechanics with a modified Hamiltonian containing an additional quantum kinetic term (see Eq. (2.59)). Starting from the definition of the quantum average of a generic observable O given by,

$$\langle O \rangle = \frac{1}{\mathcal{Z}} \text{Tr} \left[e^{-\beta H} O \right], \quad (2.60)$$

it is straightforward to evaluate the potential energy of the system:

$$\langle V \rangle \simeq \frac{1}{(2\pi\hbar)^f \mathcal{Z}} \int d^f \Gamma e^{-\tau H_P(\Gamma)} V_P(\mathbf{q}), \quad (2.61)$$

where $V_P(\mathbf{q}) = \frac{1}{P} \sum_{j=1}^P V(q_1^{(j)}, \dots, q_{3N_{\text{at}}}^{(j)})$ is the bead-dependent potential averaged over all the

³In principle, the number of beads P remains the same for each temperature T but is modified here for the sake of representability.

replicas. The quantum kinetic energy can also be evaluated according to

$$\langle T \rangle \simeq \frac{1}{(2\pi\hbar)^f \mathcal{Z}} \int d^f \mathbf{\Gamma} e^{-\tau H_P(\mathbf{\Gamma})} T_P(\mathbf{q}), \quad (2.62)$$

where $T_P(\mathbf{q})$ can be either chosen to be the centroid primitive estimator⁴,

$$T_{P,\text{pri}}(\mathbf{q}) = \frac{3N_{\text{at}}P}{2\beta} - \frac{1}{P} \sum_{j=1}^P \frac{1}{2} \left(\frac{q_i^{(j)} - q_i^{(j-1)}}{\hbar\tau} \right)^2, \quad (2.63)$$

or the more often employed centroid virial estimator,

$$T_{P,\text{vir}}(\mathbf{q}) = \frac{N_{\text{at}}}{2\beta} + \frac{1}{2P} \sum_{i=1}^{3N_{\text{at}}} \sum_{j=1}^P \left(q_i^{(j)} - \bar{q}_i \right) \partial_{q_i^{(j)}} \tilde{V}, \quad (2.64)$$

with $\bar{q}_i = \frac{1}{P} \sum_{j=1}^P q_i^{(j)}$, $\partial_{q_i^{(j)}} \equiv \frac{\partial}{\partial q_i^{(j)}}$ and $\tilde{V} = \sum_{j=1}^P V(q_1^{(j)}, \dots, q_{3N_{\text{at}}}^{(j)}) = PV_P$.

It is clear from Eq. (2.63) that the primitive estimator is more unstable and less convenient to use than the virial because of the possible presence of negative contributions in the quantum kinetic energy. This apparent drawback actually provides a stringent test for the path sampling since one should obtain $\langle T \rangle_{\text{vir}} = \langle T \rangle_{\text{pri}}$ for fully converged calculations in the limit $\delta t \rightarrow 0$ for any value of P . In practice this is almost never met exactly and it is particularly difficult to fulfill it for large P . Indeed, two constraints limit the choice of P . On one side, P must be large enough to recover all quantum properties of the physical system; on the other side, P should not be too large, otherwise the quantum imaginary time $\tau = \frac{\beta}{P}$ would become too small and, for example, too large fluctuations in the primitive energy evaluation would appear.

PIMD simulations have been widely used to study the impact of NQE in water clusters and liquid water. To the best of our knowledge, the first PI simulation of liquid water and the three first water clusters has been performed in 1985 with an empirical force field^[170]. Later, Marx and coworkers used PIMD simulations to demonstrate that the hydrated proton forms a fluxional defect in the H-bond network, rather than any individual idealized hydration state such as Zundel or Eigen along the water chain. They also established that the quantum tunneling of the proton is negligible and the transition state theory does not apply. Proton diffusion looks however defined by the thermally-induced H-bond breaking and formation of the second solvation shell^[13]. Many other PIMD calculations of charged water clusters or liquid water have been carried out so far,^[109,110,171–173] and they all confirm the importance of NQE, even at room temperature to describe PT in such systems. Nevertheless, the precise role of NQE in PT mechanisms is not fully elucidated yet since it can be significantly different from one system to another, and such effects are still under discussions. We refer the interested reader to the recent reviews of Ceriotti^[174] and of Markland^[175] for the latest advances in the theoretical treatment of NQE in water and aqueous systems.

⁴We refer the reader to the Ref. 169 for a complete derivation of the primitive energy.

In this thesis, we have the key advantage that our study of NQE will *a priori* not suffer from a bias coming from the electronic description of the problem thanks to the use of advanced QMC methods. QMC-MD driven simulations will be, as we have previously discussed, incorporated into a LD framework to control the QMC intrinsic noise on the computed ionic forces. Thus, we have to combine together PI and LD methodologies to perform PILD simulations of small water clusters. In the following, we will describe a reference algorithm to perform such calculations.

2.3.3 Path integral Langevin Dynamics

To perform our PIMD simulations using the quantum-to-classical mapping between the quantum nuclei and the fictitious classical ring polymers, we adopt here the choice of Parrinello and Rahman to set the physical masses for each necklace bead. We are thus using the Ring Polymer Molecular Dynamics (RPMD) approximation to real quantum correlation functions^[176], although real-time properties are lost when a thermostat is applied to the system, as we will do in the following. Various choices are available for the thermostat, ranging from the more conventional Nosé-Hoover chain^[154] to stochastic thermostating in its simple Path Integral Langevin equation (PILE) form. In the following, we will detail the PILE algorithm^[177] as it will constitute the basis of our novel algorithm to perform fully quantum dynamics simulation within QMC-PILD approach (see Chapter 3).

Within the PI framework, the Langevin equations of motion for each bead should in principle contain an additional harmonic force, coming from the interbead coupling. To make the problem easier to integrate, it is very convenient to use the *normal modes transformation* that rotates the bead momenta $p_i^{(j)}$ for $j = 1, \dots, P$ into $\tilde{p}_i^{(k)}$. In the normal mode representation, the LD of of each mode of the free ring polymer is that of an uncoupled harmonic oscillator which reads as

$$\begin{aligned}\dot{\tilde{p}}_i^{(k)} &= -m_i \omega_k^2 \tilde{q}_i^{(k)} - \gamma^{(k)} \tilde{p}_i^{(k)} + \sqrt{\frac{2m_i \gamma^{(k)} P}{\beta}} \xi_i^{(k)}(t) \\ \dot{\tilde{q}}_i^{(k)} &= \frac{\tilde{p}_i^{(k)}}{m_i},\end{aligned}\tag{2.65}$$

where $\xi_i^{(k)}(t)$ represents a gaussian white noise, as in Eq. (2.27) and $\omega_k = 2\tilde{\omega}_P \sin\left(\frac{(k-1)\pi}{P}\right)$ is the frequency of the k -th harmonic mode. To integrate the equation (2.65), Ceriotti and his colleagues proposed the following split operator propagator

$$e^{i\mathcal{L}\delta t} = e^{i\mathcal{L}_\gamma \frac{\delta t}{2}} e^{i\mathcal{L}_p \frac{\delta t}{2}} e^{i\mathcal{L}_0 \delta t} e^{i\mathcal{L}_p \frac{\delta t}{2}} e^{i\mathcal{L}_\gamma \frac{\delta t}{2}},\tag{2.66}$$

where $i\mathcal{L}_p$ and $i\mathcal{L}_\gamma$ have been defined in Eqs. (2.10) and (2.25), respectively. The Liouvillian operator $i\mathcal{L}_0$ is associated to the kinetic part of the classical isomorphism Hamiltonian given in Eq. (2.59). In other words, the step implying $i\mathcal{L}_0$, at the core of the PILE algorithm, corresponds to the exact propagation of the quantum harmonic modes *without* thermostat. A single PILD iteration is generated via the following steps:

1. apply a back and forth normal modes transformation and thermalize each free bead according to the Bussi scheme^[143]:

$$p_i^{(j)} \xrightarrow{\text{normal modes}} \tilde{p}_i^{(k)}$$

$$\tilde{p}_i^{(k)}(t_1) = e^{i\mathcal{L}\gamma\frac{\delta t}{2}} \tilde{p}_i^{(k)}(t) = c_1^{(k)} \tilde{p}_i^{(k)}(t) + \sqrt{\frac{m_i P}{\beta}} c_2^{(k)} \xi_i^{(k)}$$

$$p_i^{(j)} \xleftarrow{\text{real coordinates}} \tilde{p}_i^{(k)}, \quad (2.67)$$

with

$$\begin{aligned} c_1^{(k)} &= e^{-\frac{\delta t}{2}\gamma^{(k)}} \\ c_2^{(k)} &= \sqrt{1 - [c_1^{(k)}]^2}; \end{aligned} \quad (2.68)$$

2. update the momenta according to the velocity-Verlet algorithm:

$$p_i^{(j)}(t_2) = e^{i\mathcal{L}_p\frac{\delta t}{2}} p_i^{(j)}(t_1) = p_i^{(j)}(t_1) + \frac{\delta t}{2} f_i^{(j)}; \quad (2.69)$$

3. apply a back and forth normal modes transformation to propagate exactly the quantum harmonic modes, without the thermostat.

$$p_i^{(j)} \xrightarrow{\text{normal modes}} \tilde{p}_i^{(k)}$$

$$\begin{pmatrix} \tilde{p}_i^{(k)}(t_3) \\ \tilde{q}_i^{(k)}(t_3) \end{pmatrix} = e^{i\mathcal{L}_0\delta t} \begin{pmatrix} \tilde{p}_i^{(k)}(t_2) \\ \tilde{q}_i^{(k)}(t_2) \end{pmatrix} = \begin{pmatrix} \cos(\omega_k\delta t) & -m_i\omega_k \sin(\omega_k\delta t) \\ [1/m_i\omega_k] \sin(\omega_k\delta t) & \cos(\omega_k\delta t) \end{pmatrix} \begin{pmatrix} \tilde{p}_i^{(k)}(t_2) \\ \tilde{q}_i^{(k)}(t_2) \end{pmatrix}$$

$$p_i^{(j)} \xleftarrow{\text{real coordinates}} \tilde{p}_i^{(k)}; \quad (2.70)$$

4. update the ionic forces $\mathbf{f}_i(t + \delta t)$;

5. repeat the second and the first steps to terminate the symmetric propagation in Eq. (2.66).

Looking at Eq. (2.65), it appears that, within PILE algorithm, one has to specify the damping $\gamma^{(k)}$ of each fictitious harmonic mode. In the normal modes representation, the optimal choice for $\gamma^{(k)}$ turns out to be^[177,178]

$$\gamma^{(k)} = \begin{cases} 1/\tau_0 & k = 0 \\ 2\omega_k & k > 0, \end{cases} \quad (2.71)$$

where the τ_0 is a separate thermostat time constant for the centroid⁵. The damping $\gamma^{(k)}$ can be large, particularly when the ring polymer is stiff, as it is the case of the low temperature

⁵Indeed, by using $\gamma^{(0)} = 2\omega_0 = 0$, this mode would not be thermalized since it evolves according to a Hamiltonian dynamics.

limit. Consequently, the aforementioned choice of $\gamma^{(k)}$ causes technical issues that will be discussed in Chapter 3. Let us also notice that the choice of the propagator given in Eq. (2.66), also referred as OBABO algorithm⁶ in the LD community, is not unique. It has been recently established that a permutation of the elementary Liouvillian operators, in the BAOAB order, give additional stability to PILD simulations of water^[179], especially in the high friction limit.

Path Integral simulations can also be coupled with the Generalized Langevin Equation (GLE)^[177,180,181], where the noise is colored to simulate quantum effects and accelerate the convergence of the PIMD with the number of beads, at the price of breaking the validity of the FDT. This is not the case of the Quantum Thermal Bath (QTB) approach, where the power spectral density of the colored noise follows the *quantum* FDT. In practice, QTB give very reasonable results for structural properties of physical or chemical systems that behave harmonically^[182,183]. Nevertheless, the QTB approach has at least two major drawbacks. First, it is not able to capture strongly *anharmonic* effects^[184] so we do not expect this method to be accurate for the PT problem tackled in this thesis. Second, this technique is prone to ZPE leakage, like any other method based on classical trajectories. This drawback, arising at large Langevin damping γ values, is difficult to control and makes complex a rigorous analysis of the dynamical properties of the system^[185].

In our case, the correlated noise will *naturally* emerge from the intrinsic QMC noise on the estimated ionic forces $\mathbf{f}_i(t)$ which are physical. Therefore, we can legitimately hope to capture the whole physics of PT in water clusters within a QMC-PILD approach. The price to pay is however dramatic: the Langevin damping matrix γ is non diagonal within the QMC framework. Consequently, it does not longer commute with the matrix describing the harmonic coupling between beads, which seriously complicates the integration of the Langevin equations of motion. To address that problem, novel algorithms have thus to be devised, and such methodological developments will be at the heart of this thesis (see Chapter 3).

⁶In this terminology, $O \equiv i\mathcal{L}_\gamma$, $B \equiv i\mathcal{L}_p$ and $A \equiv i\mathcal{L}_0$.

Part II

Methodological developments and testing

Development of fully quantum dynamics

Contents

3.1	Derivation of novel Langevin Dynamics integrators	64
3.1.1	The Classical Momentum-Position Correlator (CMPC) algorithm	64
3.1.2	Path Integral Momentum-Position Correlator (PIMPC) algorithm	66
3.1.3	Path Integral Ornstein-Uhlenbeck Dynamics (PIOUD)	69
3.1.4	Algorithm stability with deterministic forces	71
3.1.4.1	Stability with respect to the number of beads P	72
3.1.4.2	Stability with respect to the time step δt	76
3.2	Extension to the stochastic case: correlating the noise by Quantum Monte Carlo	79
3.2.1	Evolution of the Quantum Monte Carlo wave function	79
3.2.2	Bead-grouping approximation	80
3.2.3	Quantum Monte Carlo ionic forces and noise correction	81
3.2.4	Algorithm stability with QMC forces	82

QUANTUM MONTE CARLO (QMC) techniques constitute a promising approach to study aqueous systems, since the scalability of the method ($\mathcal{O}(N^{3-4})$) makes the simulations of very large systems computationally affordable with a much greater accuracy than Density Functional Theory (DFT) in most cases. Recently, it has been demonstrated that QMC techniques provide results as accurate as the basis set converged CCSD(T) for small neutral or charged water clusters^[100,186]. Consequently, QMC can now be used as a benchmark method and certainly benefits from its intrinsically parallel formulation in modern supercomputers. The price to pay here is the systematic statistical uncertainty arising from the stochastic nature of the QMC approach. Some strategies have thus been developed to incorporate the intrinsic QMC noise into a classical statistical mechanics framework of nuclei at finite temperature T . For instance, the Coupled Electron Ion Monte Carlo (CEIMC) method relies on the Born-Oppenheimer (BO) approximation for treating finite temperature ions coupled with ground state electrons. The Boltzmann distribution function of the ionic degrees of freedom is sampled at fixed temperature via a Metropolis MC simulation based on the electronic energies computed during independent ground state QMC calculations^[187,188]. Another strategy, that will be employed here, is to resort to a Langevin MD approach to correlate the noise driving the Langevin Dynamics (LD) of the system by the QMC forces covariance. In that case, the wave function (WF) parameters are optimized by Variational Monte Carlo (VMC) calculations along the Molecular Dynamics (MD) path such that the electronic solution has always energies and forces as close as possible to the true BO surface^[149,150].

3.1. Derivation of novel Langevin Dynamics integrators

In 2014, Zen and coworkers applied the latter method to perform the very first QMC-based MD simulation of bulk liquid water at ambient conditions using 32 water molecules in a cubic cell with periodic boundary conditions^[151]. By obtaining a better position of the first peak of the oxygen-oxygen radial distribution function compared with most advanced functionals, they demonstrated the ability of QMC to tackle this kind of problems. However, the authors had to make some approximations: they expanded the WF over a small basis set and, more importantly, a classical description of the nuclei was adopted. Indeed, it is well known that Nuclear Quantum Effects (NQE) play a crucial role in the description of water or ice by deeply affecting the Radial Distribution Functions (RDFs) and distorting hydrogen bonds because of quantum disorder^[167,189–194]. In this Chapter, we propose to extend Zen’s pioneering work by including a nuclear quantum description within the QMC-driven dynamics. This is achieved within a Path Integral Langevin Dynamics (PILD) approach which, to the best of our knowledge, has never been used in the case of non-deterministic forces.

The Chapter is organized as follows. In Section 3.1, we derive three original algorithms to integrate the Langevin equations of motion for both classical and quantum particles. The designed algorithms, namely CMPC (Classical Momentum-Position Correlator), PIMPC (Path Integral Momentum Position Correlator) and PIOUD (Path Integral Ornstein-Uhlenbeck Dynamics), are very general since they can be used to propagate both deterministic and stochastic forces. A study of the numerical stability of the new path integral integrators is done in Subsection 3.1.4 for deterministic forces, where a direct comparison with existing PILD algorithms is possible. We then provide the necessary tools to perform fully quantum LD of protonated water clusters with stochastic QMC forces in Section 3.2. A bead-grouping approximation is introduced in Subsection 3.2.2, to save a significant amount of computational resources. The QMC-noise correction scheme is detailed in Subsection 3.2.3, which yields an unbiased sampling of the canonical quantum partition function, even in the presence of stochastic errors. The numerical stability analysis of the new integrators with noisy QMC forces is finally reported on Subsection 3.2.4.

3.1 Derivation of novel Langevin Dynamics integrators

In this Section, we provide two novel algorithms to integrate the Langevin equations of motion in the presence of either deterministic or stochastic forces for both classical and quantum particles. Two of these integrators (CPMC and PIOUD) have been published in Ref. 195.

3.1.1 The Classical Momentum-Position Correlator (CMPC) algorithm

The basic integration scheme developed by Attacalite *et al.*^[149], introduced in the previous Chapter, to perform QMC-driven MD simulations at finite temperature has been further developed in Refs. 150 and 151. In the latter scheme, dubbed as Classical Momentum-Position Correlator (CMPC) algorithm, the Langevin dynamics is driven by a correlated noise affecting both momenta and positions. To set up the formalism that we will be exploiting in the quantum path integral case, we provide here an alternative derivation, based on the joint

momentum-position coordinates

$$\mathbf{X} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}, \quad (3.1)$$

where \mathbf{X} is a $6N$ -dimensional vector. Analogously, we extend the definition of the ionic and random force vectors to be $\mathbf{F} = \begin{pmatrix} \mathbf{f} \\ \mathbf{0} \end{pmatrix}$ and $\mathbf{E} = \begin{pmatrix} \boldsymbol{\eta} \\ \mathbf{0} \end{pmatrix}$, respectively. In this extended basis, Eqs. (2.35) and (2.36) can be rewritten into a generalized SDE, which reads as

$$\dot{\mathbf{X}} = -\hat{\gamma}\mathbf{X} + \mathbf{F} + \mathbf{E}, \quad (3.2)$$

where in this notation the matrix $\hat{\gamma}$ represents a generalized friction that couples both momenta and positions by:

$$\hat{\gamma} = \begin{pmatrix} \gamma & \mathbf{0} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix}, \quad (3.3)$$

with the "physical" friction γ being the same $3N \times 3N$ matrix introduced in Eq. (2.35), and \mathbf{I} is the identity matrix. The formal solution of Eq. (3.2) is provided by

$$\mathbf{X}(t') = e^{-\hat{\gamma}(t'-t)}\mathbf{X}(t) + \int_t^{t'} ds e^{\hat{\gamma}(s-t')} (\mathbf{F}(\mathbf{X}(s)) + \mathbf{E}(s)). \quad (3.4)$$

Using joint coordinates would be of little use, if we were not be able to evaluate the exponential $e^{-\hat{\gamma}\delta t}$ in a closed analytic form. This is possible because the block matrix $\hat{\gamma}$ can be more conveniently rewritten in terms of Pauli matrices $\boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z$, as follows:

$$\hat{\gamma} = \frac{\gamma}{2} \otimes \mathbf{I} - \frac{\mathbf{I}}{2} \otimes \boldsymbol{\sigma}_x + i \frac{\mathbf{I}}{2} \otimes \boldsymbol{\sigma}_y + \frac{\gamma}{2} \otimes \boldsymbol{\sigma}_z. \quad (3.5)$$

Then, the exponentiation can be straightforwardly obtained by using standard Pauli matrices algebra, and the solution can be given in a closed form:

$$\mathbf{p}_{n+1} = e^{-\gamma\delta t} \mathbf{p}_n + \boldsymbol{\Gamma} (\mathbf{f}_n + \tilde{\boldsymbol{\eta}}) \quad (3.6)$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \boldsymbol{\Gamma} \mathbf{p}_n + \boldsymbol{\Theta} (\mathbf{f}_n + \tilde{\boldsymbol{\eta}}), \quad (3.7)$$

where the time evolution has been discretized with time step δt , and the subscripts refer to the corresponding time slice, such that $\mathbf{p}_n = \mathbf{p}(t_n)$, $\mathbf{q}_n = \mathbf{q}(t_n)$, and $\mathbf{f}_n = \mathbf{f}(\mathbf{q}(t_n))$. In the above equations, the other symbols are defined as

$$\begin{aligned} \boldsymbol{\Gamma} &= \gamma^{-1}(1 - e^{-\gamma\delta t}), \\ \boldsymbol{\Theta} &= \gamma^{-2}(-1 + \gamma\delta t + e^{-\gamma\delta t}), \\ \tilde{\boldsymbol{\eta}} &= \boldsymbol{\Gamma}^{-1} \int_{t_n}^{t_{n+1}} dt e^{\gamma(t-t_{n+1})} \boldsymbol{\eta}(t), \\ \tilde{\tilde{\boldsymbol{\eta}}} &= (\boldsymbol{\Theta}\gamma)^{-1} \int_{t_n}^{t_{n+1}} dt (1 - e^{\gamma(t-t_{n+1})}) \boldsymbol{\eta}(t). \end{aligned} \quad (3.8)$$

We immediately notice that the CMPC algorithm differs from those previously developed in literature^[140-145], since momenta and positions are propagated *simultaneously* in a single

3.1. Derivation of novel Langevin Dynamics integrators

iteration thanks to the use of momentum-position correlation matrices. In particular, according to Eqs. (3.6) and (3.7), not only the momenta but also the positions are affected by the integrated Langevin noise, $\tilde{\boldsymbol{\eta}}$ and $\tilde{\tilde{\boldsymbol{\eta}}}$ whose properties are detailed in Appendix D.1.

It is interesting to note that the momentum-position correlator formalism correlates the integrated noise, even without dealing with non-diagonal $\boldsymbol{\alpha}$ matrices (Eq. 2.32), as we will do in the QMC case (Section 3.2). We also remark that in order to derive the integrated equations of motion (3.6) and (3.7), we disregarded the \mathbf{q} -time dependence of \mathbf{f} in Eq. (3.4) in the time interval δt . For deterministic forces, where $\boldsymbol{\alpha}$ is position independent, it is the only approximation left in the discretized classical Langevin dynamics driven by Eqs. (3.6) and (3.7).

For deterministic \mathbf{f} , $\gamma_{ij} = \gamma_{\text{BO}}\delta_{ij}$, while the definition of $\boldsymbol{\gamma}$ is more general for noisy QMC forces (see Subsection 3.2.3). Once $\boldsymbol{\gamma}$ and δt have been set, the numerical evolution is performed according to Eqs. (3.6) and (3.7) in the frame which diagonalizes $\boldsymbol{\gamma}$.

3.1.2 Path Integral Momentum-Position Correlator (PIMPC) algorithm

In the following, we will extend the CMPC algorithm, described in the previous Subsection, to the quantum case. As already mentioned, this framework will allow us to incorporate noisy QMC forces in the equations of motion, by using an appropriately tailored Langevin noise without breaking the FDT, as it is explained in Section 3.2. Moreover, in our quantum algorithms, which make use of a Trotter break-up^[94] between the harmonic and the physical modes, the quantum harmonic part is integrated *exactly* together with the Langevin thermostat for the harmonic frequencies.

We start by applying the same idea as of our classical algorithm with momentum-position correlators to the quantum-to-classical mapping Hamiltonian H_P described in Eq. (2.59). The equations of motion corresponding to H_P coupled to a Langevin thermostat are those in Eq. (3.2), provided \mathbf{X} is now interpreted as a $6NP$ -dimensional vector, the thermal noise $\boldsymbol{\eta}$ lives in the $3NP$ -dimensional space, and the generalized $\hat{\boldsymbol{\gamma}}$ must be redefined in order to include the harmonic couplings between the beads. $\hat{\boldsymbol{\gamma}}$ is now a $6NP \times 6NP$ matrix, which reads as

$$\hat{\boldsymbol{\gamma}} = \begin{pmatrix} \boldsymbol{\gamma} & \mathbf{K} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix}, \quad (3.9)$$

where \mathbf{K} is $3NP \times 3NP$ matrix defined as follows:

$$K_{ih}^{(j)(k)} = \tilde{\omega}_P^2 \delta_{ih} \left(2\delta^{(j)(k)} - \delta^{(j)(k-1)} - \delta^{(j)(k+1)} \right). \quad (3.10)$$

In the above definition, we have used lower indices to indicate the particle components, the upper ones (in parenthesis) indicate the bead components, while δ_{ij} ($\delta^{(i)(j)}$ for the bead indices) is the usual Kronecker delta. The \mathbf{K} matrix is diagonal in the particle sector, as the harmonic springs in the fictitious H_P of Eq. (2.59) couple different replicas only for the same particle components. Cyclic conditions are implicitly applied in Eq. (3.10) to the matrix boundaries in the bead sector (i.e. $(0) = (P)$), as the polymers are necklaces.

Chapter 3. Development of fully quantum dynamics

As we have done in Eq. (3.5) for the classical case, we expand $\hat{\gamma}$ of Eq. (3.9) in Pauli matrices, as follows:

$$\hat{\gamma} = \frac{\gamma}{2} \otimes \mathbf{I} + \frac{\mathbf{K} - \mathbf{I}}{2} \otimes \boldsymbol{\sigma}_x + i \frac{\mathbf{K} + \mathbf{I}}{2} \otimes \boldsymbol{\sigma}_y + \frac{\gamma}{2} \otimes \boldsymbol{\sigma}_z. \quad (3.11)$$

Moreover, we require that $[\mathbf{K}, \boldsymbol{\gamma}] = 0$, which is fulfilled when the forces \mathbf{f} are deterministic. Indeed, $\boldsymbol{\gamma}$ obeys the FDT (2.32) in the extended $3NP$ dimensional space. For deterministic forces, $\boldsymbol{\gamma}$ is bead independent, while for noisy forces the matrix $\boldsymbol{\alpha}$ could be correlated according to the force covariance $\langle \delta f_i^{(j)} \delta f_h^{(k)} \rangle$, and $\boldsymbol{\gamma}$ could acquire a bead dependent contribution to fulfill the FDT. For the time being, let us assume $\boldsymbol{\gamma}$ to be bead independent, which implies that \mathbf{K} and $\boldsymbol{\gamma}$ commute. We will see in Section 3.2 how to recover this condition even in the case of QMC forces. Then, Eq. (3.11) and $[\mathbf{K}, \boldsymbol{\gamma}] = 0$ allow one to evaluate $e^{-\hat{\gamma}\delta t}$ in a closed analytic form for each (upper and lower) block component of the SDE formal solution in Eq. (3.4). The related algebra is quite tedious and we refer the reader to the Appendix D.3. The resulting integrated equations of motions lead to the following Markov chain:

$$\mathbf{p}_{n+1} = \boldsymbol{\Lambda}_{1,1}\mathbf{p}_n + \boldsymbol{\Lambda}_{1,2}\mathbf{q}_n + \boldsymbol{\Gamma}\tilde{\boldsymbol{\eta}}, \quad (3.12)$$

$$\mathbf{q}_{n+1} = \boldsymbol{\Lambda}_{2,1}\mathbf{p}_n + \boldsymbol{\Lambda}_{2,2}\mathbf{q}_n + \boldsymbol{\Theta}\tilde{\tilde{\boldsymbol{\eta}}}, \quad (3.13)$$

with the integrated $6NP$ -dimensional noise (\mathbf{E}_{int}) and force (\mathbf{F}_{int}) which are, respectively,

$$\mathbf{E}_{\text{int}} = \begin{pmatrix} \boldsymbol{\Gamma}\tilde{\boldsymbol{\eta}} \\ \boldsymbol{\Theta}\tilde{\tilde{\boldsymbol{\eta}}} \end{pmatrix} = \int_{t_n}^{t_{n+1}} dt e^{\hat{\gamma}(t-t_{n+1})} \begin{pmatrix} \boldsymbol{\eta} \\ 0 \end{pmatrix}, \quad (3.14)$$

$$\begin{aligned} \mathbf{F}_{\text{int}} &= \begin{pmatrix} \boldsymbol{\Gamma}\mathbf{f}_n \\ \boldsymbol{\Theta}\mathbf{f}_n \end{pmatrix} = \int_{t_n}^{t_{n+1}} dt e^{\hat{\gamma}(t-t_{n+1})} \begin{pmatrix} \mathbf{f}_n \\ 0 \end{pmatrix} \\ &= \hat{\gamma}^{-1} \left(\mathbf{I} - e^{-\hat{\gamma}\delta t} \right) \begin{pmatrix} \mathbf{f}_n \\ 0 \end{pmatrix}. \end{aligned} \quad (3.15)$$

The expressions for $\boldsymbol{\Lambda}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ matrices are quite complex, so they are given in Appendix D.3. Using this integration scheme, one is in principle able to work with large time steps for the dynamics since all the high frequency vibrations are substantially damped in this type of dynamics. Similarly to the classical case, we can compute the noise correlation matrix by using its definition:

$$\langle \mathbf{E}_{\text{int}}^T \mathbf{E}_{\text{int}} \rangle = \int_{-\delta t}^0 dt e^{\boldsymbol{\gamma}t} \hat{\boldsymbol{\alpha}} e^{\boldsymbol{\gamma}^\dagger t}, \quad (3.16)$$

where

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} \boldsymbol{\alpha} & 0 \\ 0 & 0 \end{pmatrix}. \quad (3.17)$$

Once again, we refer the reader to the Appendix D.3 for the analytic expression of the noise correlators in the quantum case.

After setting up the general formalism, we need to define the $\boldsymbol{\gamma}$ matrix in order to fully determine the numerical scheme, as we have done for the CMPC algorithms (Subsection 3.1.1). In the quantum case $\boldsymbol{\gamma}$ is a $3NP \times 3NP$ matrix. Moreover, we want that $[\mathbf{K}, \boldsymbol{\gamma}] = 0$.

3.1. Derivation of novel Langevin Dynamics integrators

Let us first introduce γ^{BO} as a $3N \times 3N$ matrix which is bead independent and defined as in the classical case, namely $\gamma_{ij}^{\text{BO}} = \gamma_{\text{BO}} \delta_{ij}$ for deterministic forces, with γ_{BO} a free parameter, to be optimized. In the case of noisy forces, γ^{BO} will have off-diagonal elements, as we have anticipated for the classical algorithm. This will be explained in Subsection 3.2.3. Therefore, γ^{BO} should retain all optimal damping properties related to the physical BO forces $\mathbf{f}^{(i)}$ acting on each replica (i). We assume that the physics within each replica does not change from one bead to another, thus the optimal γ^{BO} can safely be taken as bead independent.

The additional complication of the quantum case is represented by the harmonic matrix \mathbf{K} , coupling the beads to each other, whose eigenvalues are $\lambda_k = \omega_k^2$ for $k = 1, \dots, P$, where

$$\omega_k = 2\tilde{\omega}_P \sin\left(\frac{(k-1)\pi}{P}\right) \quad \text{for } k = 1, \dots, P. \quad (3.18)$$

The energy scale set by \mathbf{K} can be much larger than the one of the BO potential energy surface. In the case the BO forces \mathbf{f} are set to zero, one can compute analytically the relevant autocorrelation time τ_H (using H as operator in Eq. (3.28)) of the related Langevin dynamics. It has been shown^[177,178] that the optimal damping which minimizes τ_H is given by $\gamma_{\text{harm}}^{(k)} = 2\omega_k$, for each harmonic eigenmode k .

We would like γ to be optimal for both BO and harmonic dynamics. A way to define such a matrix is to exploit the commutator $[\mathbf{K}, \gamma] = 0$, which implies that γ and \mathbf{K} share a common set of eigenvectors. Thus, we can work in a rotated frame which diagonalizes both γ^{BO} (bead independent) and \mathbf{K} (particle independent). In this frame γ is diagonal, with eigenvalues:

$$\gamma_i^{(k)} = \gamma_i^{\text{BO}} + \gamma_{\text{harm}}^{(k)} \quad (3.19)$$

for the set of eigenvectors of both γ^{BO} (lower index) and \mathbf{K} (upper index). This uniquely defines the full matrix γ , as it can be reconstructed from its eigenvectors and eigenvalues. For noisy forces the γ_i^{BO} spectrum will be non trivial, while for deterministic forces we have $\gamma_i^{\text{BO}} = \gamma_{\text{BO}} \forall i$. Thus, even in the quantum case which mixes harmonic with physical modes, the only parameter left to tune is γ_{BO} , as the optimal damping for harmonic forces is fully determined by Eq. (3.18).

To summarize, we have introduced here an algorithm which is able to integrate simultaneously both the physical molecular and the fictitious ring polymer vibration modes in a single time step δt . This could be appealing, but it turns into a disadvantage, as we will see in the tests reported in Subsection 3.1.4. Indeed, the compromise between physical and harmonic modes made for the friction in Eq. (3.19) does not seem optimal. That choice leads to overdamped dynamics with lower diffusion. Moreover, we recall that in Eq. (3.13) we made the approximation of constant BO forces \mathbf{f} during the time step δt . This condition is hard to meet for fast vibrational modes driven by \mathbf{K} . The resulting algorithm turns out to be less stable as a function of the number of beads P and the time step δt . We are going to improve it, by resorting to a Trotter breakup in order to separate the harmonic part from the BO one.

3.1.3 Path Integral Ornstein-Uhlenbeck Dynamics (PIOUD)

The previous PIMPC algorithm can be improved by starting from the Fokker-Planck Liouville operator given in Eq. (2.24). With no assumption on the form of the Langevin damping matrix γ^1 , it is defined as

$$i\mathcal{L}_{\text{FP}} = \sum_{i=1}^{3NP} \left(\mathcal{F}_i \nabla_{p_i} + p_i \nabla_{q_i} - \sum_{j=1}^{3NP} \gamma_{ij} (\nabla_{p_i} p_j + k_B T P \nabla_{p_i} \nabla_{p_j}) \right) \quad (3.20)$$

in mass-scaled coordinates. $i\mathcal{L}_{\text{FP}}$ is built upon the Hamiltonian propagation, driven by the first two terms, and the Langevin thermostat, represented by the last two, deriving from the Fokker-Planck equation. For the sake of readability, i and j run here over all particles and beads indexed together. In Eq. (3.20), $\mathcal{F}_i \equiv f_i^{\text{BO}} + f_i^{\text{harm}}$ is the generalized force, comprising the BO and harmonic contributions, where $f_i^{\text{BO}} \equiv -\nabla_{q_i} \tilde{V}$ and $\tilde{V} = \sum_{j=1}^P V(q_1^{(j)}, \dots, q_{3N}^{(j)})$.

As we have seen in the previous Subsection, the quantum-to-classical isomorphism Hamiltonian in Eq. (2.59) includes very different energy scales. For instance, if one works with a large number of beads P , the ring polymer vibration modes can become much stiffer than the molecular vibrations arising from the BO potential. For the PIMPC algorithm, introduced in Subsection 3.1.2 and tested in Section 3.1.4, we found that propagating simultaneously all these modes is not optimal since soft modes will be overdamped and the phase space sampling will be less efficient. To overcome this problem, we would like to split the Liouvillian in Eq. (3.20) into just two operators, one containing only the physical (BO) modes, the other depending exclusively on the fictitious (harmonic) modes. To do so, we first separate the friction matrix into two contributions, in a way analogous to Eq. (3.19):

$$\gamma = \gamma^{\text{BO}} + \gamma^{\text{harm}}. \quad (3.21)$$

We can then rewrite the total Liouvillian as the sum of two terms, $i\mathcal{L}_{\text{FP}} = i\mathcal{L}^{\text{BO}} + i\mathcal{L}^{\text{harm}}$, where

$$i\mathcal{L}^{\text{harm}} = \sum_i \left(f_i^{\text{harm}} \nabla_{p_i} + p_i \nabla_{q_i} - \sum_j \gamma_{ij}^{\text{harm}} (\nabla_{p_i} p_j + k_B T P \nabla_{p_i} \nabla_{p_j}) \right), \quad (3.22)$$

$$i\mathcal{L}^{\text{BO}} = \sum_i \left(f_i^{\text{BO}} \nabla_{p_i} - \sum_j \gamma_{ij}^{\text{BO}} (\nabla_{p_i} p_j + k_B T P \nabla_{p_i} \nabla_{p_j}) \right), \quad (3.23)$$

in such a way that we can break up the evolution operator via Trotter factorization^[94]. In order to preserve the reversibility of the process and reduce at most the time step error, we approximate the propagation using a symmetric form^[196,197]:

$$e^{i\mathcal{L}\delta t} \simeq e^{i\mathcal{L}^{\text{BO}}\delta t/2} e^{i\mathcal{L}^{\text{harm}}\delta t} e^{i\mathcal{L}^{\text{BO}}\delta t/2}. \quad (3.24)$$

The equations of motion corresponding to $i\mathcal{L}^{\text{harm}}$ are those in Eq. (3.2) with the generalized $\hat{\gamma}$ of Eq. (3.9) where $\mathbf{F} = 0$. They are *linear* in both \mathbf{p} and \mathbf{q} , thus exactly solvable in an analytic

¹This matrix used to be diagonal in the Bussi derivation introduced in Chapter 2.

3.1. Derivation of novel Langevin Dynamics integrators

closed form. They describe an OUP, and their algebra and properties have been exploited in the GLE framework, where they are used to propagate external auxiliary variables together with the physical momenta in order to generate a colored noise with a corresponding effective memory kernel. In the present algorithm we are going to use the OUP to integrate directly the SDE for both \mathbf{p} and \mathbf{q} , together with the Langevin noise, namely without further splitting the Langevin thermostat in $i\mathcal{L}^{\text{harm}}$. The resulting evolution is given by the coupled equations in Eqs. (3.12) and (3.13), but with the notable difference of \mathbf{f}_n set to zero, as the BO component of the total force has been loaded in the $i\mathcal{L}^{\text{BO}}$ Trotter factors. As we mentioned in the previous Subsection, dropping the time dependence of \mathbf{f}_n was the only approximation made in the PIMPC case. Here, there is no time dependence either in \mathbf{f}_n or in γ^{harm} , and thus $e^{i\mathcal{L}^{\text{harm}}\delta t}$ is applied exactly. γ^{harm} is defined as in Subsection 3.1.2, namely it is the optimal damping for the dynamics driven by a harmonic H in its normal modes representation. However, in this case there is no low-energy damping for small frequencies in γ_{harm} . Thus, we need to introduce another one, γ_0 , specific to $i\mathcal{L}^{\text{harm}}$, such that in the normal modes representation γ^{harm} is diagonal and reads as:

$$\gamma_{\text{harm}}^{(k)} = \begin{cases} 2\omega_k & \text{if } 2\omega_k \geq \gamma_0 \\ \gamma_0 & \text{otherwise,} \end{cases} \quad (3.25)$$

for $k = 1, \dots, P$. In other words, the thermostat controlled by the friction γ_0 is mainly applied to the centroid rototranslational modes which would not be thermalized otherwise because their frequency is zero. Contrary to the other free quantum harmonic modes, the optimal damping γ_0 is not general and has to be optimized for the system under study. A reasonable value of this parameter, also present in the PILE method, can be found by performing short preliminary simulations with any accurate force field^[68,69,198,199].

The equation of motion corresponding to $i\mathcal{L}^{\text{BO}}$ is given by^[19]

$$\mathbf{p}(t) = e^{\gamma(t_0-t)}\mathbf{p}(t - \delta t/2) + \int_{t_0}^t dt' e^{\gamma(t'-t)}(\mathbf{f}(t - \delta t/2) + \boldsymbol{\eta}(t')), \quad (3.26)$$

with the initial time t_0 . Indeed, only the momenta are evolved in $i\mathcal{L}^{\text{BO}}$. Therefore, no Momentum-Position Correlator (MPC) algorithm is needed, and the resulting equations are equal to those of the simple classical Langevin algorithm introduced in Ref. 149, restricted to \mathbf{p} , which in the present case is a $3NP$ -dimensional vector. The corresponding γ^{BO} is taken block diagonal according to the bead index, and possibly bead dependent. This is very useful for QMC noisy forces, as the statistics of the QMC force covariance matrix is genuinely bead dependent, because the WF is sampled independently for each bead (see Section 3.2). In the QMC case, γ^{BO} is fundamental to correct the intrinsic QMC noise affecting the BO forces. However, for deterministic BO forces, the corresponding γ^{BO} can be taken equal to zero, and the corresponding time-discretized solution will result in a simple velocity update $\mathbf{p}_{n+1} = \mathbf{p}_n + \frac{\delta t}{2}\mathbf{f}_n$. In fact, any additional damping in the BO modes will turn into a slower algorithmic diffusion, as the thermalization is already guaranteed by the Langevin thermostat integrated in $i\mathcal{L}^{\text{harm}}$. In this limit, $i\mathcal{L}^{\text{BO}}$ will reduce to $i\mathcal{L}_p$.

To summarize, a single PIOUD iteration is generated via the following steps:

1. update the particles momenta according to the Eq. (3.26) ($e^{i\mathcal{L}^{\text{BO}}\delta t/2}$), in either the real coordinate space (deterministic forces) or in the frame that diagonalizes the Langevin damping matrix γ (stochastic forces);
2. apply a back and forth normal mode transformation to *both* propagate and thermalize the free quantum ring polymers according to the Eqs. (3.12) and (3.13) ($e^{i\mathcal{L}^{\text{harm}}\delta t}$);
3. new evaluation of the ionic forces via $\mathbf{f}(t + \delta t) = -\nabla_{\mathbf{q}}V(t + \delta t)$;
4. final update of the particles momenta as in step 1, by applying again the operator $e^{i\mathcal{L}^{\text{BO}}\delta t/2}$.

The main difference between PILE^[177] and our PIOUD algorithm^[195] is that, within PILE, $i\mathcal{L}^{\text{harm}}$ is split into $i\mathcal{L}_0$, which propagates the motion of the harmonic oscillators in their normal modes representation, and $i\mathcal{L}_\gamma$ (the part proportional to $\gamma_{ij}^{\text{harm}}$ in Eq. (3.22), $i\mathcal{L}_0$ being the remaining part in the first line of the same equation), which corresponds to the Langevin thermostat acting on the normal modes. Such a splitting of the two aforementioned Liouvillian operators is not applied within PIOUD, as the exact integration of quantum harmonic forces coupled to a Langevin thermostat is applied via $i\mathcal{L}^{\text{harm}}$. Apart from the additional Trotter break-up, in the sequence of operators given in Eq. (2.66) the Langevin thermostat on the normal modes is the outermost part of the Liouvillian decomposition, for a better control of the target temperature. Our algorithm in Eq. (3.24) performs better than the solution proposed in Eq. (2.66), since in our case $i\mathcal{L}^{\text{harm}} = i\mathcal{L}_0 + i\mathcal{L}_\gamma$ corresponds to an exact propagator including both the coordinates change and the thermalization of the free quantum ring polymers. As a consequence, one Trotter factorization is saved which in principle enables us to work with a larger time step δt or more quantum replicas P . Indeed, the Trotter break-up is obviously exact if the two operators involved commute. The split of $i\mathcal{L}^{\text{harm}}$ proposed by Ceriotti and his colleagues has the drawback that choice as the commutator of $[i\mathcal{L}_0, i\mathcal{L}_\gamma]$ is dominated by the term $\gamma_{ij}^{\text{harm}} f_j^{\text{harm}} \nabla_{p_i}$ which diverges as P^3 for large P , as it can be easily seen by the exact expression of the harmonic forces and the choice of the optimal friction. This argument suggests that our method, involving commutators at most diverging as P^2 , should have a time step error much better behaved for large P .

To conclude, we have presented two novel algorithms to perform LD simulations of small water clusters, with either deterministic or stochastic forces. In particular, by saving a Trotter break-up in the propagator with respect to the PILE algorithm, our PIOUD integrator should exhibit further stability with respect to the time step δt and the number of beads P . This improvement will be tested and discussed in the next Section.

3.1.4 Algorithm stability with deterministic forces

We test the robustness of the two algorithms, namely PIMPC and PIOUD, detailed in the previous Section, by comparing them with PILE. We perform PILD simulations on the Zundel ion with almost exact deterministic forces which are simply computed by finite differences of the CCSD(T) PES provided by Huang and coworkers^[65]. Since our only aim here is to test the different integration schemes, the number of MD iterations $N_{\text{steps}} = 10^5$ is quite small

3.1. Derivation of novel Langevin Dynamics integrators

to visit the entire phase space but sufficient to identify some possible weakness of the propagators. We recall that PIMPC denotes the propagator corresponding to the Eqs. (3.12) and (3.13) since positions and velocities evolve jointly, whereas the PIOUS integrator corresponds to Eq. (3.24), in which we separate the physical from the fictitious vibration modes of the system.

A robust PILD algorithm must be stable with both large time step δt and large number of quantum replicas P . Indeed, the collective modes of a large ring polymer become very stiff, so much more difficult to control, especially if the integration time step is large. Close values of the virial $T_{P,\text{vir}}(\mathbf{q})$ (Eq. 2.64) and the primitive $T_{P,\text{pri}}(\mathbf{q})$ (Eq. 2.63) kinetic energy estimators point that we are sampling properly both positions and momenta of all particles.

Beside temperature and kinetic energy, other observables will also be used here to quantify the numerical efficiency of the tested algorithms. A very stable propagation of the equations of motion can for instance be due to an overcautious (too small δt) integration of the dynamics, which does not constitute a real improvement since the phase space will be poorly visited. To this purpose, Ceperley introduced the algorithmic diffusion constant D which reads as^[200]

$$D = \left\langle \frac{\left[\delta t \sum_{i=1}^{3N} \sum_{j=1}^M \left(q_i^{(j)}(t + \delta t) - q_i^{(j)}(t) \right) \right]^2}{T_{\text{iter}}} \right\rangle. \quad (3.27)$$

T_{iter} is the total amount of CPU time spent for one single MD iteration at a given time step δt . Since T_{iter} is almost equal for the three algorithms under study, the algorithmic diffusion can be interpreted as the usual diffusion of quantum particles in the position space. Furthermore, we will also focus our attention on the potential autocorrelation time known to be very sensitive to Langevin damping since the softest vibration modes are sometimes very long to sample. This useful quantity has the following expression

$$\tau_V = \frac{1}{\langle V^2 \rangle - \langle V \rangle^2} \int_0^{N_{\text{steps}}\delta t} dt \langle \delta V(0) \delta V(t) \rangle, \quad (3.28)$$

with $\delta V = V - \langle V \rangle$ the fluctuation of the potential energy. τ_V needs to be minimized to obtain an optimal production run and this is achieved by working with the largest possible time step δt . Simultaneously, an optimal centroid damping γ_0 has to be found to minimize τ_V , as we will detail later.

3.1.4.1 Stability with respect to the number of beads P

We report the obtained results for the three different integration schemes (PIMPC, PILE and PIOUS) at low temperature ($T = 50$ K) in Figure 3.1 and at room temperature ($T = 300$ K) in Figure 3.2, as a function of the number of quantum replicas P . In each case, the four key observables (virial versus primitive kinetic energy, average temperature, algorithmic diffusion constant and potential autocorrelation time) described a few lines above are compared. We perform simulations working with an increasing number of beads $P = 4, 8, 16, 32, 64, 128$ and

256, using a reasonable but not-so-small time step $\delta t = 0.5$ fs.

The quantum kinetic energy estimators are known to be useful tools to determine the number of beads required at a given temperature to capture NQE. However, other quantities such as the algorithmic diffusion constant and the potential autocorrelation time can also be used to diagnose the good convergence of the quantum kinetic energy. Indeed, these observables keep increasing (quantum kinetic energy and diffusion) or decreasing (potential autocorrelation time) when the number of beads increases until they reach a plateau value. In the case of the Zundel ion, $P = 128$ quantum replicas are enough to fully recover the quantum kinetic energy at low temperature (see Figure 3.1), whereas only $P = 32$ beads are required at room temperature, according to Figure 3.2. We note that these values are consistent with those frequently used in the literature for this system^[13,172,201].

A rapid inspection of Figures 3.1 and 3.2 shows that the PIMPC approach, in which all operations are propagated into one single block without any mode separation during the dynamics, is by far less efficient and stable than the other two. Indeed, the primitive energy estimator (see Eq. (2.63)) becomes unstable for $P = 256$ at 50 K and for $P \geq 32$ at room temperature, showing that this algorithm is not able to handle the stiffest vibration modes of the ring polymer. The asymptotic value of the algorithmic diffusion constant is also dramatically reduced and even slightly decreases at room temperature for very large number of quantum replicas. Indeed, by working in a mixed space where all the physical and fictitious vibration modes are propagated during a single iteration, one has to deal with very different energy scales, spanning various orders of magnitude. This leads to an overdamping of the softest intermolecular modes which are strongly penalized because of the presence of very high frequency modes. This also explains why the potential autocorrelation time is slightly larger for this algorithm while the simulation temperature seems to be rather well-controlled. The performances of the PIOUD propagator of Eq. (3.24) and the PILE algorithm are much closer to each other. Looking at the algorithmic diffusion constant and the potential autocorrelation time, the PILE and PIOUD algorithms have the same computational efficiency. This is expected because the same normal mode transformation is applied in both approaches.

We notice that PILE exhibits a built-in stability of the average simulation temperature with respect to the number of beads. On the contrary, our algorithms (both PIMPC and PIOUD) display a natural and small time step error on the target temperature which tends to vanish when the number of beads increases. This temperature difference can be easily explained by the fact the instantaneous velocities (and so the temperature) are measured just after the thermalization step $i\mathcal{L}_\gamma$ in the PILE integration scheme, whereas they are evaluated after the $i\mathcal{L}^{\text{BO}}$ operator in our propagations. Since $i\mathcal{L}^{\text{BO}}$ includes the propagation according to the ionic forces, that are not harmonic and therefore not exactly integrated, it is reasonable to expect a larger error in this case. We remark however that an artificially small error in the target temperature is not at all important, because the temperature depends only on the velocities that have a trivial (i.e. Boltzmann) distribution. It is much more important to have the correct distribution for the coordinates q , regardless what is the distribution of the velocities. Nevertheless, when one increases the number of beads, the error made on the target temperature (which is already less than 1%) decreases. Indeed, for large P , the propagation

3.1. Derivation of novel Langevin Dynamics integrators

driven by $i\mathcal{L}^{\text{harm}}$ becomes dominant over the BO forces. This implies a more effective control of the temperature, because $i\mathcal{L}^{\text{harm}}$ corresponds to an exact integration. When looking at the top left panels of Figures 3.1 and 3.2, we notice that the PIOUD algorithm displays an almost perfect control of the kinetic energy operators at each temperature. On the contrary, the PILE algorithm exhibits a significant instability of the primitive energy for a large number of beads ($P \geq 64$) at room temperature. The contrast between the perfect average temperature and the quantum kinetic energy instabilities demonstrates that a good control of the velocities does not necessarily imply an accurate sampling of the positions. The robustness of our approach mainly relies on a simultaneous control of both positions and velocities during the dynamics via momentum-position correlation matrices.

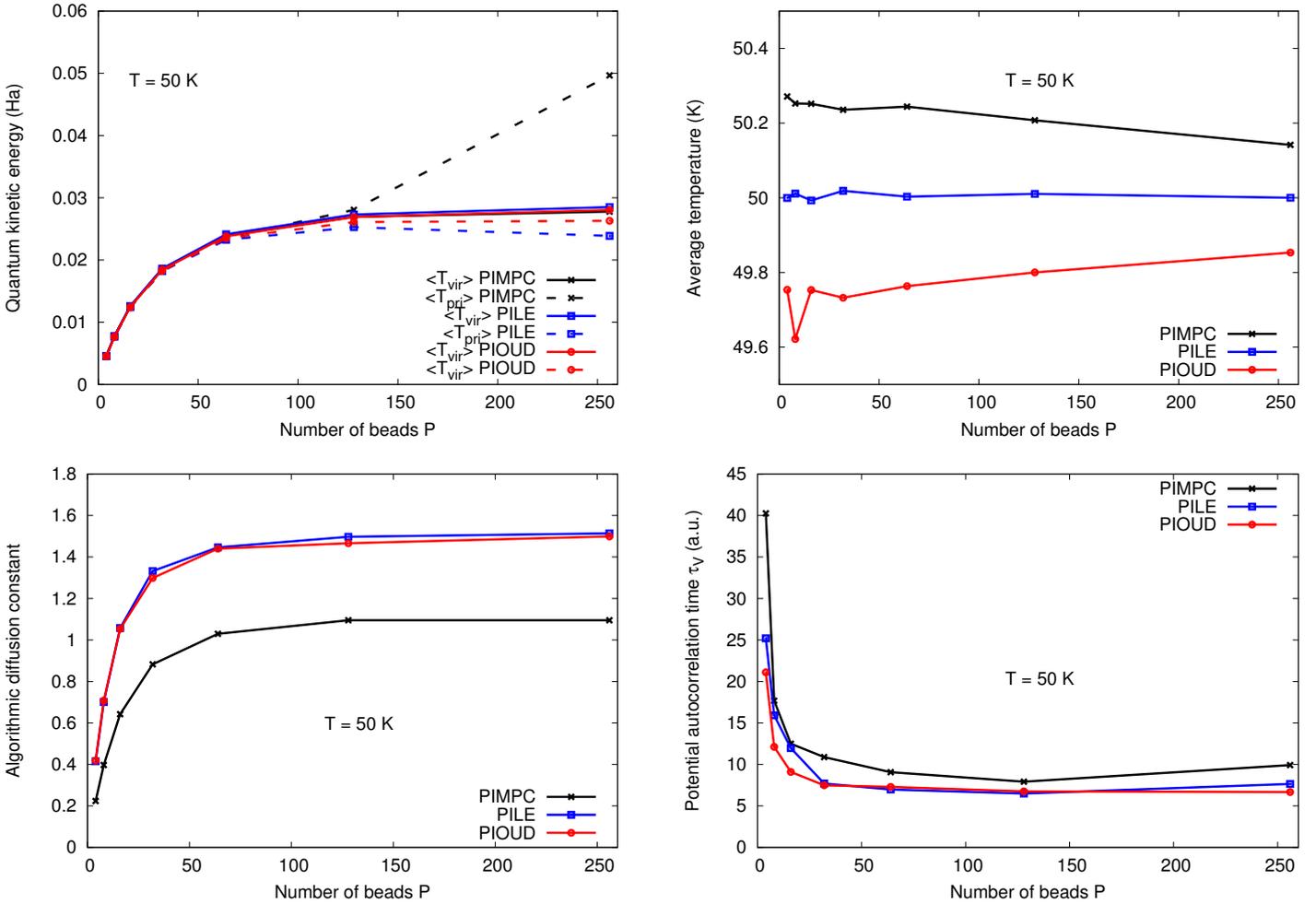


Figure 3.1 – Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the number of quantum replicas P evolving at $T = 50$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. The time step and the friction are respectively set to $\delta t = 0.5$ fs and $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).

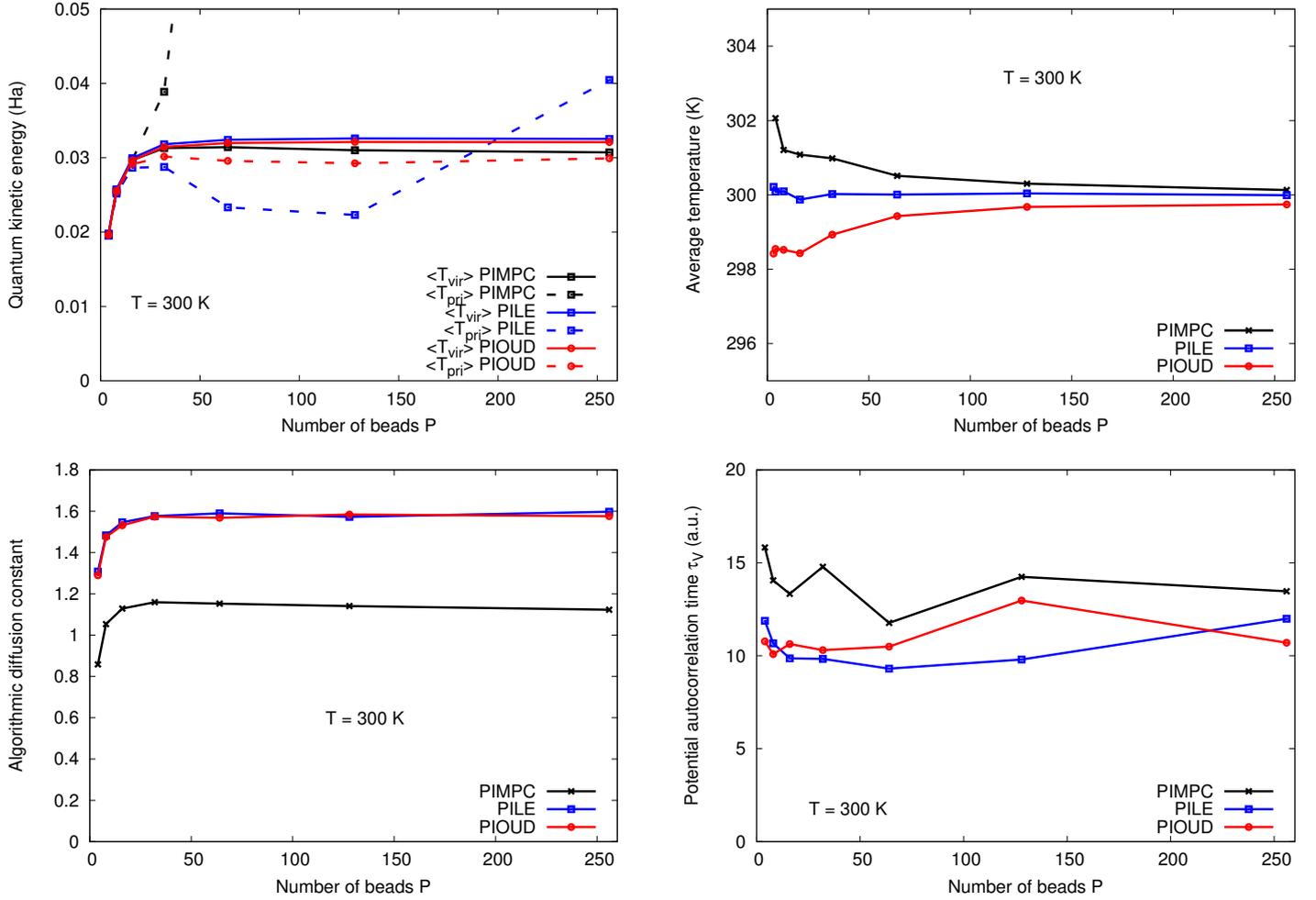


Figure 3.2 – Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the number of quantum replicas P evolving at $T = 300$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. The time step and the friction are respectively set to $\delta t = 0.5$ fs and $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).

In summary, the PIOUD propagator appears to be more efficient than PIMPC and PILE as the positions are better controlled, which enables us to work with a larger number of quantum replicas at a fixed time step δt .

3.1.4.2 Stability with respect to the time step δt

Thanks to the preliminary analysis discussed above, we know how many quantum replicas should be used to generate an efficient and converged PILD simulation of the Zundel ion at a given temperature. Moreover, stability for rather large time steps is crucial in the perspective of performing PILD calculations on larger systems using a PES evaluated by accurate but computationally demanding *ab initio* methods. We report here CC-PILD simulations at room temperature with $P = 32$ beads using different values of $\delta t = 0.1, 0.2, 0.3, 0.5, 0.75$ and 1 fs. The behavior of the observables used to evaluate the algorithm efficiency are plotted in Figures 3.3 and 3.4 as a function of the time step at 50 K and 300 K, respectively.

Similarly to the previous figures, the PILE propagator exhibits a remarkably stable average temperature when one increases the time step. On the contrary, our PIMPC and PIOUD algorithms are suffering from a time step error $\mathcal{O}(\delta t^3)$ arising from the Trotter factorization in the propagation of the equations of motion. In the PILE algorithm instead, the temperature is measured just after the $i\mathcal{L}_\gamma$ step so it is not contaminated by the time step error yet.

In order to quantify the bias induced by the time step error, we check the difference $|\langle T \rangle_{P,\text{vir}} - \langle T \rangle_{P,\text{pri}}|$ which gives us direct information on the accuracy of the positions sampling. Like in the previous tests, the PIOUD algorithm shows the smallest difference thanks to a good control of the primitive energy, due to the separation of the Liouvillian in physical and harmonic modes. The difference between these two kinetic energy estimators is more spectacular at room temperature. Indeed, the fluctuation-dissipation contributions in $i\mathcal{L}^{\text{harm}}$, related to damping and random forces in the dynamics, become more important as the temperature increases, while the BO forces are not so strongly affected by thermal effects.

The diffusion constant shows a very similar behavior as a function of time step for the PILE and the PIOUD propagators. As expected, the potential autocorrelation time decreases significantly with increasing the time step without relevant differences between the PILE and the PIOUD algorithms.

To conclude, we have derived a novel algorithm to efficiently integrate the Langevin equations of motion. The PIOUD integrator has been validated on an analytic (deterministic) CCSD(T) force field for the Zundel complex. Indeed, it has proven to be remarkably stable with respect to *both* the number of beads P and the time step δt without losing computational efficiency. We now wish to go further by applying the previous formalism to the case of stochastic PES and forces, such as the ones computed in a QMC framework, without reducing the efficiency of the phase space sampling.

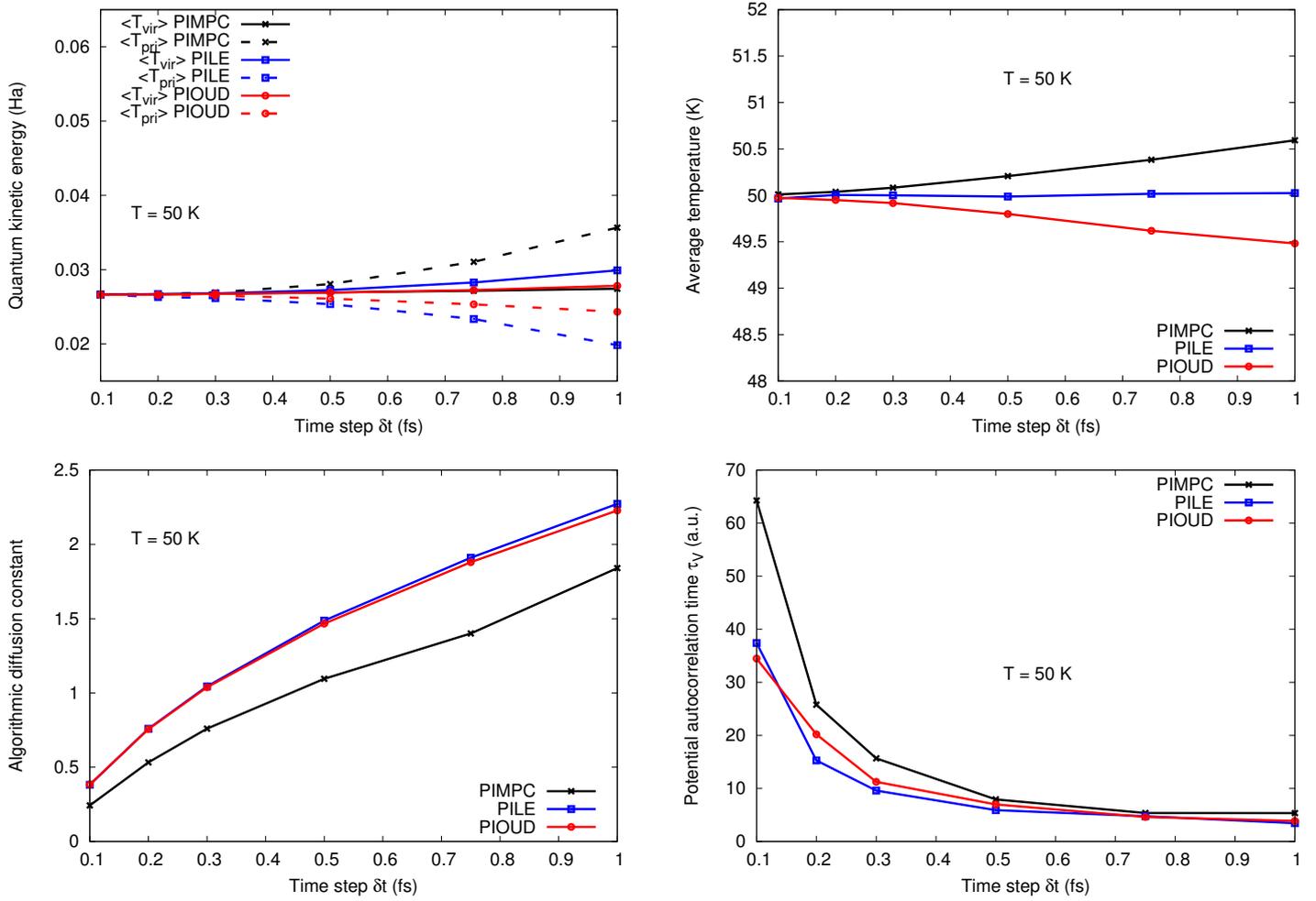


Figure 3.3 – Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the time step δt at $T = 50$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. $P = 128$ quantum replicas are used and the friction is set to $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).

3.1. Derivation of novel Langevin Dynamics integrators

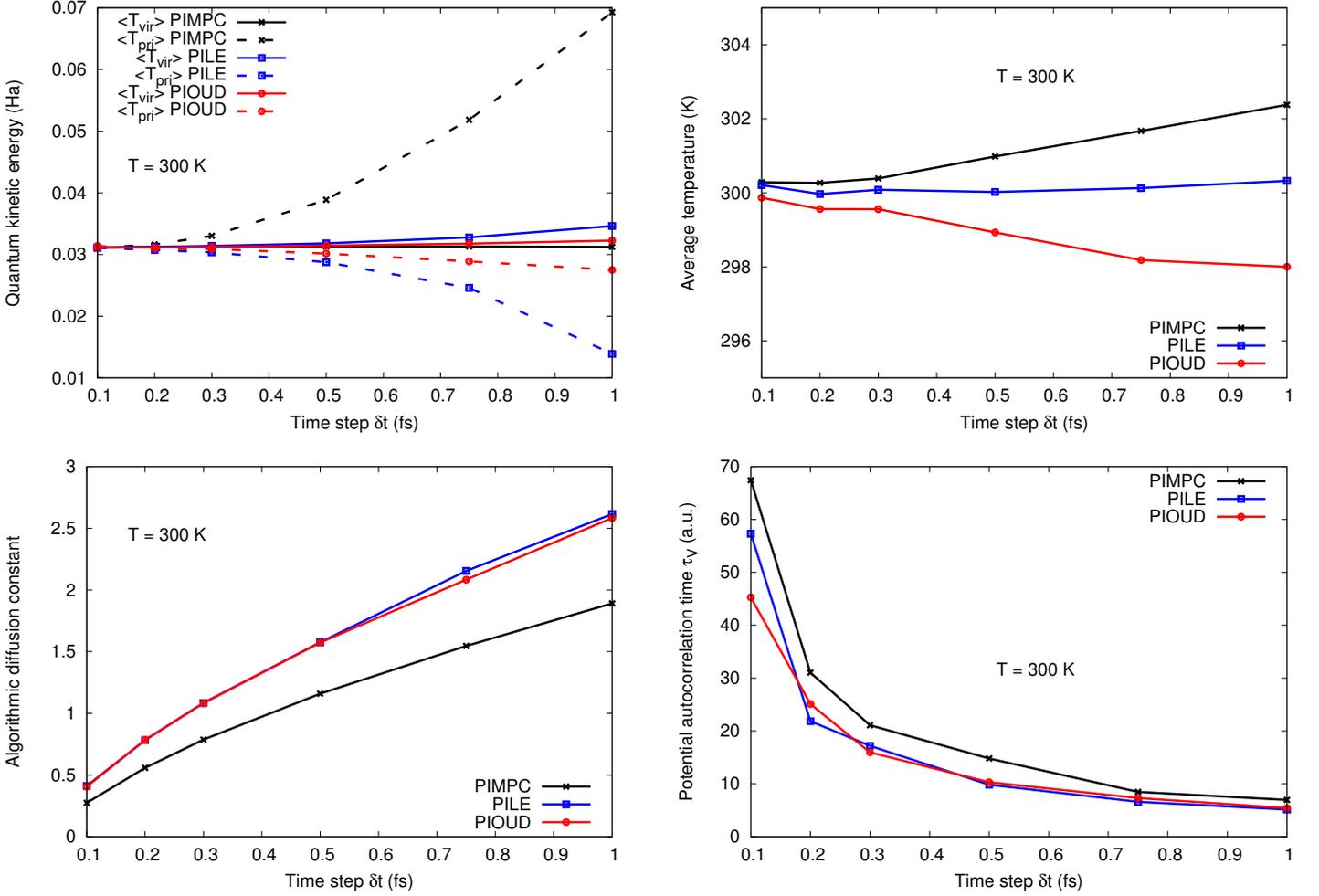


Figure 3.4 – Evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ (top left panel), the temperature T (top right panel), the algorithmic diffusion constant D (bottom left panel) and the potential autocorrelation time τ_V (bottom right panel) as a function of the time step δt at $T = 300$ K. Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. The color code indicates each algorithm: black for the PIMPC algorithm, blue for the PILE propagator and red for the PIOUD algorithm. $P = 32$ quantum replicas are used and the friction is set to $\gamma_0 = 1.46 \cdot 10^{-3}$ a.u. (γ_{BO} for the PIMPC algorithm).

3.2 Extension to the stochastic case: correlating the noise by Quantum Monte Carlo

In this thesis, our goal is to perform fully quantum LD simulations of small water clusters, with ionic forces computed by QMC. Due to the stochastic nature of the QMC approach, such forces are intrinsically *noisy*. We have tackled in the previous Section the integration of the Langevin equations of motion in the presence of VMC noisy forces, with an adapted break-up of the elementary operators of the Fokker-Planck Liouvillian. However, the question of controlling the temperature of the QMC-driven CPMC or the PIOUS simulations remains now to be solved. More generally, we will detail the employed procedure to perform QMC-driven LD simulations. In particular, we explain the various approximations employed to evaluate the WF between each nuclear iteration at an affordable computational cost.

3.2.1 Evolution of the Quantum Monte Carlo wave function

To describe the Zundel ion (Chapter 4) or the protonated water hexamer (Chapter 6), we use a Jastrow Antisymmetrized Geminal Power (JAGP) WF, whose expression is given in Eq. (1.36)². The JAGP WF is expanded over a *localized* gaussian basis set (Eq. (1.41)) for both the Jastrow and the AGP parts. In this thesis, we start from the optimal WF devised by Dagrada *et al.* for the Zundel ion at zero temperature and further technical details concerning the variational WF can be found in Ref. 72. We simply mention that we have employed the geminal embedding scheme^[202] to obtain $N_{\text{hyb}} < N_{\text{basis}}$ geminal embedding orbitals filling the AGP matrix. This strategy significantly reduces the total number of parameters describing the QMC WF. This point is very important, because in the present optimization methods^[77,203,204], the involved matrices in the iterative procedure have linear dimension equal to the number p of involved parameters. The used number of QMC samplings to stochastically characterize a $p \times p$ matrix should be much larger than p . Otherwise the matrix is biased if not rank deficient, and the QMC optimization methods no longer work. Thus, it is clear that the reduction of the number of parameters is very much needed at present, because it is roughly proportional to the computational cost of the QMC optimization.

Here, we use the O[8]H[2] hybrid basis set which has been proven to be the best compromise between accuracy and computational cost, due to a limited number (571) parameters describing the determinantal part of the WF^[72]. During the dynamics, the gaussian type orbitals exponents $\zeta_{l,n}$ (see Eq. (1.41)) in both the Jastrow and the AGP parts of the WF are kept frozen to make the simulation stable. At each new ionic configuration, the WF must be reoptimized. This is done by means of the Stochastic Reconfiguration (SR)^[77,203], or the optimization method with Hessian acceleration (SRH)^[204] (see Appendix C). As the ionic positions are smoothly connected to those of the previous MD time step, also the WF parameters will evolve continuously. After each nuclear iteration, the WF is optimized by few (five in our simulations of the Zundel ion) SR or SRH steps on the electronic parameters to ensure that the system is close enough to the BO surface, before continuing the propagation of the ion dynamics. Due to the continuity of the nuclear trajectories, the number of SR steps is significantly smaller than the one required for a WF optimization from scratch.

² $\Psi_D = \Psi_{AGP}$ in this case.

3.2.2 Bead-grouping approximation

Another advantage of our framework is represented by the local, ion-centered, basis set. We start from the observation that the most relevant dependence of the WF on the ionic positions \mathbf{q} comes directly (and explicitly) from the basis set, while the dependence of the electronic variational parameters is generally weaker. This is particularly true in quantum MD calculations, where each particle is represented by an P -bead necklace. In the first approximation, a necklace could share the same WF parameters, while the full \mathbf{q} dependence is provided by the basis set only, a dependence coming from the atomic centers -different in each bead- defining the basis set at each different (quantum) time slice. This approximation can be very effective to reduce the computational burden, because the main drawback encountered in PIMD and PILD calculations with respect to their classical counterparts is the factor-of- P increase of variational parameters ($P \times p$ if no approximation is employed). Conversely, if the number of variational parameters is restricted to the same number as the classical simulation, no computational overhead is expected to work with $P > 1$ in QMC. Indeed, the evaluation of the ionic forces can be done with a number of samples inversely proportional to P^3 , as the temperature in each time interval is increased by P and the Langevin equations require statistically less accurate -but nevertheless unbiased- forces, so that increasing P becomes essentially cost-free in QMC^[188].

When instead forces are computed with deterministic methods - *e.g.* DFT - the computational burden is necessarily proportional to P , and therefore several techniques have been recently developed to decrease the number of evaluations of the ionic forces by about an order of magnitude without missing any significant NQE. This is done by achieving smart interpolations and groupings of different possible paths^[205,206] or by applying a Ring Polymer Contraction (RPC) scheme^[207-209]. Otherwise, the number of quantum replicas can be reduced by working with a GLE including a colored noise mimicking the quantum fluctuations of the nuclei^[180,181,210,211].

Although these methods could be effectively incorporated into our QMC framework, we are forced to explore another approach, because, as we have discussed, in QMC the problem is just the large number of variational parameters, and not the large value of P . We introduce therefore a method which takes advantage of the explicit WF representation of the electronic problem. Indeed, each bead at each iteration has its own optimal WF $|\Psi_{\mathbf{q}}^{(k)}\rangle$, for $k = 1, \dots, P$, which minimizes the variational energy at the nuclear configuration $\mathbf{q}^{(k)}$. Consequently, we need to find the best variational parameters set $\{\boldsymbol{\lambda}^{(k)}\} = \{g_{\mu,\nu}^{a^{(k)},b^{(k)}}, \lambda_{\mu,\nu}^{a^{(k)},b^{(k)}}, b^{(k)}, \zeta_{l,n}^{(k)}, \dots\}$ for each WF. Despite the availability of efficient QMC optimization algorithms, that task is still computationally demanding and would limit the application of our method to very small systems. To overcome this major difficulty, we exploit the local nature of the gaussian basis sets used in the expansion of both the Jastrow and AGP factors as discussed above. As anticipated, we make the approximation of defining N_{groups} groups of neighboring beads and constraining the WF parameters to be equal for all beads in the same group. Since a group shares the same parameters, the corresponding energy gradients are then averaged over the

³Provided that the samples are independent, it is equivalent to generate N_{gen} samples to evaluate the ionic forces on a unique WF or $P \times \frac{N_{\text{gen}}}{P}$ samples (inversely proportional to P) on P WF describing each bead.

quantum replicas constituting the group. In this way, we improve the statistics by a factor of P/N_{groups} . We obtain less noisy parameters even though the resulting WF is not exactly optimized for each quantum replica. We mention that this is a controllable and systematically improvable approximation. Indeed, if one takes $N_{\text{groups}} = P$, the electronic result is exact, whereas $N_{\text{groups}} = 1$ constitutes the roughest approximation. In the latter case, one performs a fully quantum dynamics with almost the same statistics as the one with classical nuclei. We are going to test and validate this approximation for the Zundel ion in the following Chapter. In the case of the simple hydrogen molecule the approximation with $N_{\text{groups}} = 1$ allows one to recover 90% of the Zero Point Energy (ZPE).

3.2.3 Quantum Monte Carlo ionic forces and noise correction

In our approach, the potential energy landscape and thus the ionic forces acting on each particles are evaluated by VMC, according to Eq. (1.68). As we have seen at the very end of the Chapter 1, the expression of the QMC force (Eqs. (1.69) and (1.70)) is quite complex because it contains the implicit and explicit dependence of $V(\mathbf{q})$ on both the nuclear positions \mathbf{q} and the electronic parameters set $\{\boldsymbol{\lambda}\}$. Computing these forces with finite variance and in a fast way is of paramount importance to make a QMC-based MD and PIMD possible. This has become feasible, thanks to recent improvements, that have been discussed in the Chapter 1.

Another issue related to the stochastic nature of this method is the control of the statistical noise introduced into the dynamics. The latter must be kept under control if we want to have an unbiased sampling of the phase space during the propagation of the trajectory. Fortunately, the methodology described in the previous sections is particularly suited for this situation, as it deals with the most general case, where the friction γ and covariance $\boldsymbol{\alpha}$ matrices are not diagonal. This corresponds to the inclusion of a correlated noise into the dynamics, in contrast to the more usual white noise case. By construction, we can now make the assumption that there exists also a QMC contribution to the covariance matrix which becomes position-time dependent^[149]

$$\boldsymbol{\alpha}(\mathbf{q}) = \alpha_{\text{BO}}\mathbf{I} + \Delta_0\boldsymbol{\alpha}^{\text{QMC}}(\mathbf{q}). \quad (3.29)$$

$\alpha_{\text{BO}} = 2k_B T \gamma_{\text{BO}}$ is the white noise contribution and Δ_0 is a tunable parameter which is mainly set to make the covariance (and thus the friction) matrix positive definite⁴. $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are linked by the FDT, while $\boldsymbol{\alpha}^{\text{QMC}}$ is the QMC-force covariance matrix defined by

$$\alpha_{ij}^{\text{QMC}} = \langle \delta f_i \delta f_j \rangle, \quad (3.30)$$

where $\delta f_i = f_i - \langle f_i \rangle$ is the fluctuation of the i -th ionic force. Previous works have established that the QMC covariance matrix is roughly proportional to the dynamical matrix (it is exactly proportional for harmonic forces). Therefore, it carries information on the vibrational properties of the system^[19]. As a consequence, one could take advantage of an apparent drawback of the QMC approach by using the intrinsic noise to drive a dynamics in the phase space with nearly optimal sampling.

⁴To keep the covariance matrix positive definite in practice, the condition $\Delta_0 \geq \delta t$ must be fulfilled.^[149]

3.2. Extension to the stochastic case: correlating the noise by Quantum Monte Carlo

However, to fulfill the FDT, the QMC noise has to be disentangled from the total Langevin noise in the following way:

$$\eta_i = \eta_i^{\text{ext}} + \eta_i^{\text{QMC}} \quad (3.31)$$

such that $\alpha_{ij}^{\text{QMC}} = \langle \eta_i^{\text{QMC}} \eta_j^{\text{QMC}} \rangle$, namely η_i^{QMC} is the noise already present in the i -th QMC force component. Therefore, by using that the external noise is independent of the QMC one, we obtain

$$\langle \eta_i^{\text{ext}} \eta_j^{\text{ext}} \rangle = \alpha_{ij} - \alpha_{ij}^{\text{QMC}}. \quad (3.32)$$

The above equation is valid at the time instant t . However, in the schemes devised in Subsections 3.1.1, and 3.1.3, the equations of motion are discretized and integrated over a finite time step δt , and also the Langevin noise $\tilde{\eta}$ appearing in those equations is integrated over the same δt . This will slightly change the form in Eq. (3.32), leading to the following one:

$$\langle \tilde{\eta}_i^{\text{ext}} \tilde{\eta}_j^{\text{ext}} \rangle = \tilde{\alpha}_{ij} - \alpha_{ij}^{\text{QMC}}, \quad (3.33)$$

where $\tilde{\alpha}$ is the integrated noise correlator matrix, reported in Eq. (D.1) for the CMPC algorithm. α^{QMC} does not need to be modified, as the expectation value of the variance-covariance QMC-force matrix in Eq. (3.30) is evaluated in this case during the time step δt while sampling the electron coordinates. Since the r.h.s. of Eq. (3.33) is a positive definite matrix, it defines a corrected external noise which is compatible with the solution of the Langevin dynamics in both classical and quantum cases, as long as $|\mathbf{q}_{n+1} - \mathbf{q}_n|$ remains small, regardless of how large the friction is. As already mentioned, the proposed MPC schemes yield a correlated noise also affecting the conjugate variables (*i.e.* the nuclear positions \mathbf{q}). Therefore, the relation in Eq. (3.33) must be extended in the joint momentum-position coordinates. The full expression for the noise correction in the classical CPMC integration algorithm is detailed in Appendix D.2.

In the quantum PIOUS algorithm, the QMC noise correction must be applied exclusively in the BO step $i\mathcal{L}^{\text{BO}}$, where only the momenta are evolved according to the QMC ionic forces. Therefore, the external noise acts only on the momentum sector, and its variance-covariance is simply given by Eq. (3.33). Contrary to the CPMC case, there is no need to extend this relation to the position sector. In the harmonic step $i\mathcal{L}^{\text{harm}}$, the harmonic forces are noiseless, therefore the external noise coincides with the full noise ($\eta_i^{\text{QMC}} = 0$), and no correction is needed. The momentum and position components of the noise ($\tilde{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\eta}}$, respectively) are correlated according to Eqs. (D.12) and (D.13), reported in Appendix D.3.

3.2.4 Algorithm stability with QMC forces

In the following, we will present some calibration runs carried out with the PIOUS algorithm in the VMC framework, which will help us set the proper simulation parameters (γ_0 , δt) and show the remarkable stability of the quantum dynamics even with noisy VMC forces for large δt and large P . The results are obtained by performing QMC-PILD short tests (about 8.5 ps of dynamics) of the Zundel ion in the gas phase at room temperature (300 K). Hereafter, the additional PIOUS parameter γ_{BO} is set to zero, to avoid overdamping in the BO propagator. The Langevin thermostat in the $i\mathcal{L}^{\text{BO}}$ part plays solely the role of correcting the BO dynamics

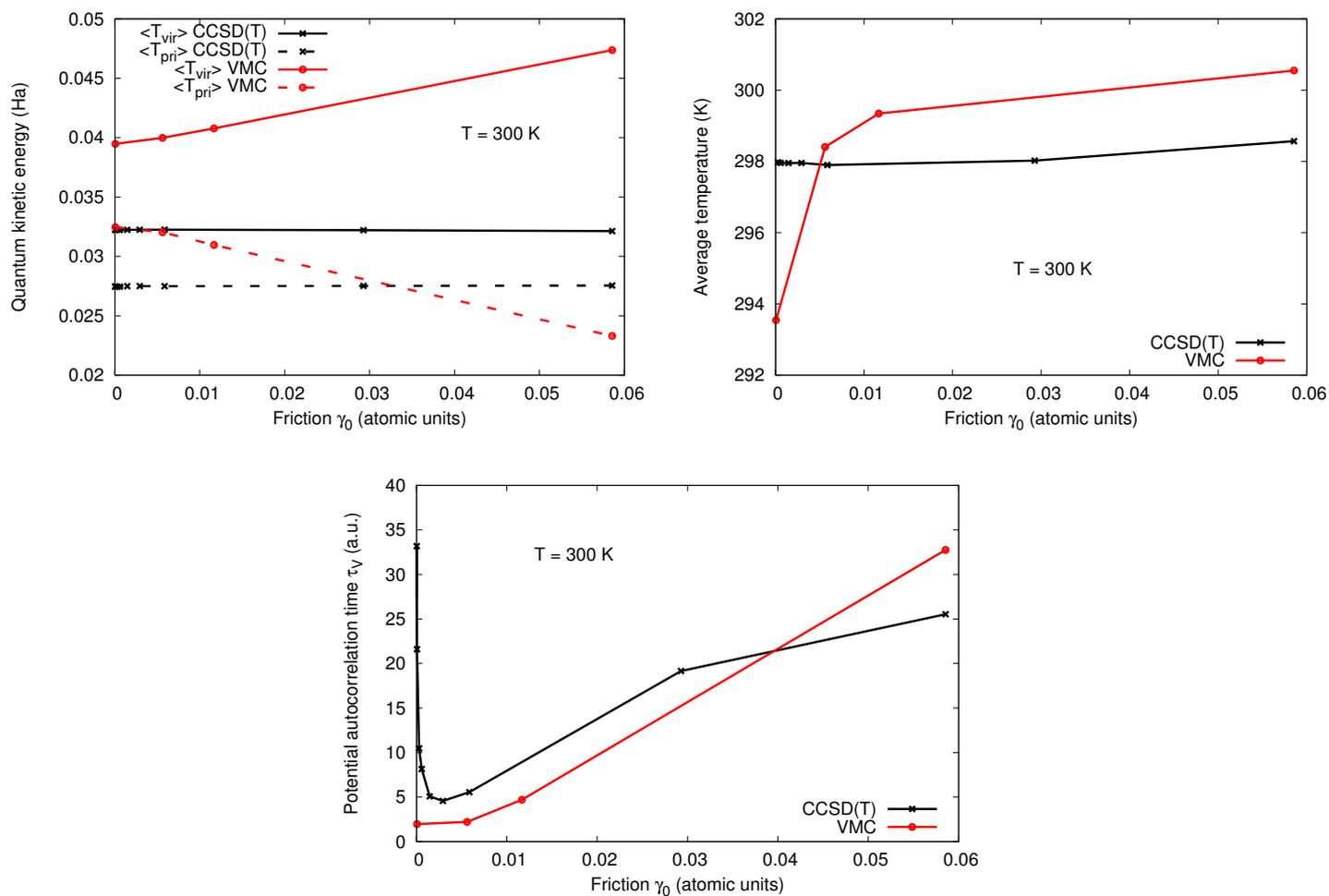


Figure 3.5 – PIOUD evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ (top panel), the temperature T (middle panel) and the potential autocorrelation time τ_V (bottom panel) as a function of the input friction γ_0 . Solid lines correspond to the virial estimator of the kinetic energy whereas the primitive estimator curves are dashed. Deterministic forces are represented in black whereas the noisy QMC forces are in red. The time step and the number of quantum replicas are respectively set to $\delta t = 1$ fs and $P = 32$.

3.2. Extension to the stochastic case: correlating the noise by Quantum Monte Carlo

for the intrinsic VMC noise. As we have already seen, for deterministic forces $i\mathcal{L}^{\text{BO}}$ reduces to $i\mathcal{L}_p$, i.e. it is a simple velocity step.

For each MD step we have evaluated forces and all the energy derivatives with $\approx 3.3 \times 10^5$ MC samples, much larger than the total number of variational parameters ($p \approx 2.4 \times 10^3$) of the WF. This allows reaching an accuracy of 1.5 mHa (0.94 kcal/mol) in the total energy per VMC energy minimization step. Between two MD steps, five QMC energy minimizations are performed with the Hessian (SRH) algorithm, in order to sample the correct BO surface. As already mentioned in Section 3.2.1, all variational parameters are evolved during the dynamics, except for the GTO exponents, which are kept frozen.

The optimal input friction γ_0 is chosen to minimize the potential autocorrelation time τ_V (Eq. (3.28)), and generate the most efficient phase space sampling during the dynamics. We will give a general protocol to find this optimal value in the case of stochastic VMC forces, where the situation can be more complicated since the FDT is now sensitive to the QMC intrinsic noise. In our initial tests, the time step is set to 1 fs, a large value, which guarantees a quick and effective exploration of the phase space. Moreover $\gamma_{\text{BO}} = 0$ and $\Delta_0 = \delta t$, as we discovered that the most efficient simulation is the one which minimizes the damping in the BO sector. Δ_0 is taken as the minimal value which provides a positive definite γ^{BO} . In the top panel of Figure 3.5, we first observe that in the VMC case, the virial and the primitive kinetic energy estimators deteriorate when the applied input friction is too large ($> 0.02 \times 10^{-3}$ a.u.). In this situation, the additional QMC noise makes the coupling between the system and the thermostat too large to be fully controlled for such time step values ($\delta t = 1$ fs). On the other extreme, at very small γ_0 , we see that the presence of a QMC-correlated noise tends to flatten the sharp and deep minimum of the potential autocorrelation time obtained with deterministic CCSD(T) forces (bottom panel). This indicates that, contrary to the deterministic case where there is a clear advantage to set the input friction γ_0 to its optimal value, there is more freedom to choose this parameter in VMC. Indeed, the autocorrelation time divergence shown in the CCSD(T) case for small values of γ_0 disappears with VMC forces. This is due to the implicit low-value cutoff in the γ matrix provided by the intrinsic QMC noise, once the QMC-force covariance matrix is converted into an effective friction, according to Eqs.(2.32) and (3.29). In practice however, the value of γ_0 cannot be too small either, in order to avoid too cold temperatures shown in the middle panel of Figure 3.5. Consequently, we need to take the largest γ_0 before the increase of the potential autocorrelation time τ_V due to the soft-modes overdamping. This will also let us recover an acceptable target temperature (see middle panel). Therefore, $\gamma_0 = 1.46 \times 10^{-3}$ a.u seems to be a very good compromise between autocorrelation time τ_V , effective temperature, and quality of the phase space sampling revealed by the kinetic energy estimators. All subsequent runs will be performed with that value. It is interesting to remark that this is optimal for both deterministic and stochastic forces. Similarly to the tests performed in Subsection 3.1.4 with analytic CCSD(T) forces, we check here the robustness of our novel PIOUD algorithm with respect to the time step δt and to the number of quantum replicas P in the presence of noisy QMC forces. In the upper panel of Figure 3.6, the difference between the virial and the primitive kinetic energy estimators remains very reasonable with increasing values of the time step δt , even though the time step error is more important in the stochastic case once compared to the deterministic

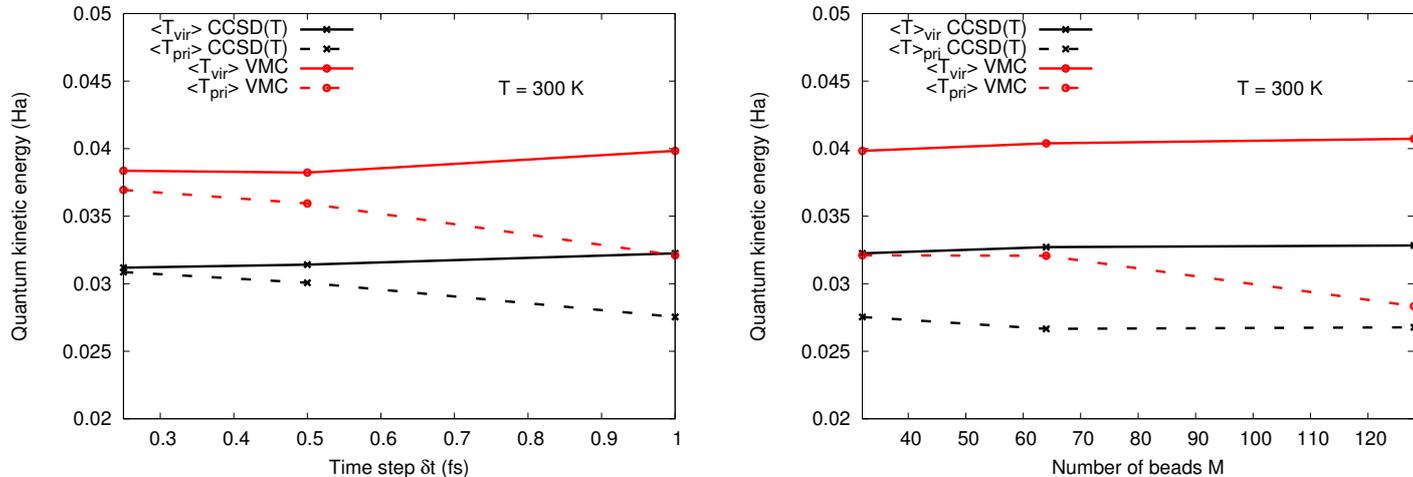


Figure 3.6 – PIOUS evolution of the quantum kinetic energy estimators $\langle T_{\text{vir}/\text{pri}} \rangle$ as a function of the time step δt (top panel) and the number of quantum replicas P (bottom panel). Colors and symbols are the same as in Figure 3.5. The input friction was set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u. fs. The default values of the time step and the number of quantum replicas are respectively $\delta t = 1$ fs and $P = 32$.

one. The PIOUS propagator exhibits a smaller difference between these two estimators with QMC forces than the PILE algorithm in the deterministic case. The superior performances of PIOUS will allow us to use large time steps δt .

Finally, we also check the stability of the PIOUS integration scheme with increasing number of quantum replicas P . As we can see in the bottom panel of the Figure 3.6, the difference between the kinetic and the primitive energy estimators is well controlled up to $P = 64$ beads in the stochastic case. On the contrary, the PILE propagator already exhibits signs of instability at this value with deterministic forces (see upper left panel of Figure 3.2). This is a further proof of the robustness of the PIOUS integrator, which we thus recommend when one wants to perform a PILD simulation with a large number of beads with any force field, deterministic or not.

Thanks to the technical developments presented in this Chapter, we have paved the way to perform fully quantum simulations of small protonated water clusters, with highly accurate ionic forces evaluated within the VMC approach. The very last step before performing such calculations on realistic systems such as the protonated water hexamer, is to check the reliability of our novel methodology on a benchmark system, as we will do in the next Chapter.

The Zundel ion: a benchmark system

Contents

4.1	Zero temperature results	88
4.1.1	Potential energy landscape	88
4.1.2	Equilibrium geometries	90
4.1.3	Accuracy of the Quantum Monte Carlo approach	95
4.2	Benchmark calculations	98
4.2.1	Validation of the classical dynamics	99
4.2.2	Validation of the quantum dynamics	100
4.3	Proton transfer in the Zundel ion	104

ONE OF THE MAIN ISSUES which sets back from a complete understanding of proton transfer (PT) in water, is related to the very sensitive thermal behavior of PT. The required Potential Energy Surface (PES) precision, of the order of a few tenths of kcal/mol, has been reached only recently by state-of-the-art computational methods beyond DFT, such as MP4^[212], Coupled Cluster (CC)^[213] or multi-reference configuration interaction (MRCI)^[214] methods. They are however characterized by a much poorer scalability with respect to DFT, and therefore they do not allow simulation of sufficiently large molecular clusters. In the previous Chapter, we have set the stage to perform efficient Path Integral Langevin Dynamics (PILD) simulations of such systems with accurate and noisy Quantum Monte Carlo (QMC) forces, exploiting the mild scaling of the VMC approach with the system size.

To validate our methodological developments, we first apply them to H_5O_2^+ , namely the Zundel ion^[215], widely used as a benchmark system. Indeed, it is the smallest charged water cluster to exhibit a non trivial proton transfer and its reduced size makes a comprehensive and systematic study of the problem easier. More importantly, there is a huge amount of data to compare our results to, because the description of excess proton in water has been widely studied both theoretically and experimentally in the last fifty years. On the one hand, the fast development of spectroscopical instruments allowed to probe experimentally vibrational properties of ionic species and therefore many studies have been published^[17,216] on the H_5O_2^+ ion. On the other hand, several accurate theoretical works have appeared on the Zundel ion to study its structure and energetics^[110,216–225]. We can cite for instance the extremely accurate Potential Energy Surface (PES) generated by Bowman and coworkers from almost exact CCSD(T) calculations^[65], the MS-EVB methods^[218,219,226,227], or more recently the LEWIS model developed by Herzfeld^[228]. Moreover, because of the great importance of the Nuclear Quantum Effects (NQE) in the Zundel cation, the latter is almost always used to test and validate new approaches^[229,230].

In this Chapter, we first assess the ability of the QMC approach to accurately describe both the energetics and the geometric properties of the protonated water dimer in Section 4.1. Later, we perform in Section 4.2 benchmark calculations at finite temperature to ensure that our QMC-based methodology provides an accurate description of the Zundel ion. To that purpose, we compare our results to classical or quantum Molecular Dynamics simulations carried out with an analytical CCSD(T) force field to ensure the QMC intrinsic noise does not spoil the protonated water dimer dynamics. In Section 4.3, we finally investigate the impact of the temperature and NQE on the PT processes occurring in the Zundel cation.

4.1 Zero temperature results

The protonated water dimer, represented in Figure 4.1, is constituted by an excess proton H^+ surrounded by two neighboring water molecules. The nature of the minimum energy structure of the $H_5O_2^+$ ion has been debated in the literature. There are two candidates with competing energies: a C_2 symmetric structure (left hand side of Figure 4.1), commonly known as the Zundel configuration, with the proton evenly shared between the two oxygen atoms, and a C_s -Inv one (right hand side of Figure 4.1) with the proton slightly closer to one H_2O molecule. Accurate highly correlated studies^[65,222–224] have confirmed that the global minimum is C_2 symmetric. In this Section, we will explore the zero temperature properties of the protonated water dimer using different computational methods. In particular, we focus on the description of the PES and the geometric properties which are of paramount importance to predict the possible configurations the $H_5O_2^+$ ion may adopt at finite temperature.

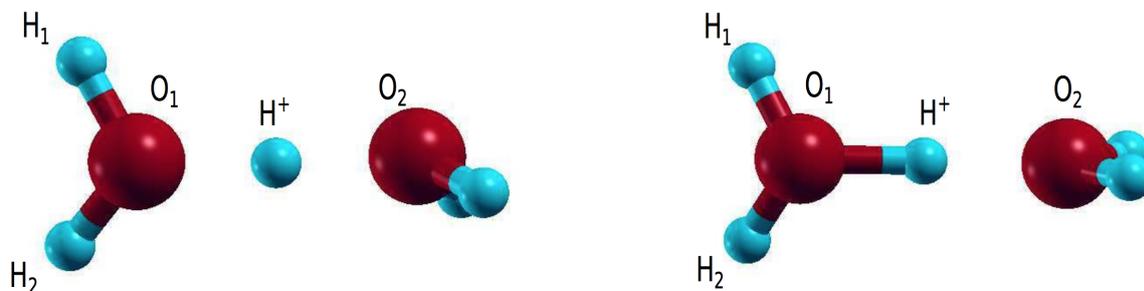


Figure 4.1 – QMC optimized geometries for global C_2 minimum (left) and for C_s -Inv local minimum (right) of the Zundel ion^[72]. Atom labels for the analysis of the Zundel properties are also indicated.

4.1.1 Potential energy landscape

In order to explore the protonated water dimer PES, we compute the $H_5O_2^+$ electronic energy as a function the central inter-oxygen distance $d_{O_1O_2}$ (atom labels are given in Figure 4.1). The results are plotted in Figure 4.2, where different electronic structure methods are compared. The solid curves are obtained using DFT (dark green for PBE and magenta for DF2), while the QMC results are respectively represented by blue triangles (VMC with pseudopotential), brown squares (VMC all electron calculations) and black squares (LRDMC). Finally, the

quantum chemistry calculations, performed using the triple zeta (TZ) basis set, are represented by red triangles (MP2) and orange circles (CCSD(T)).

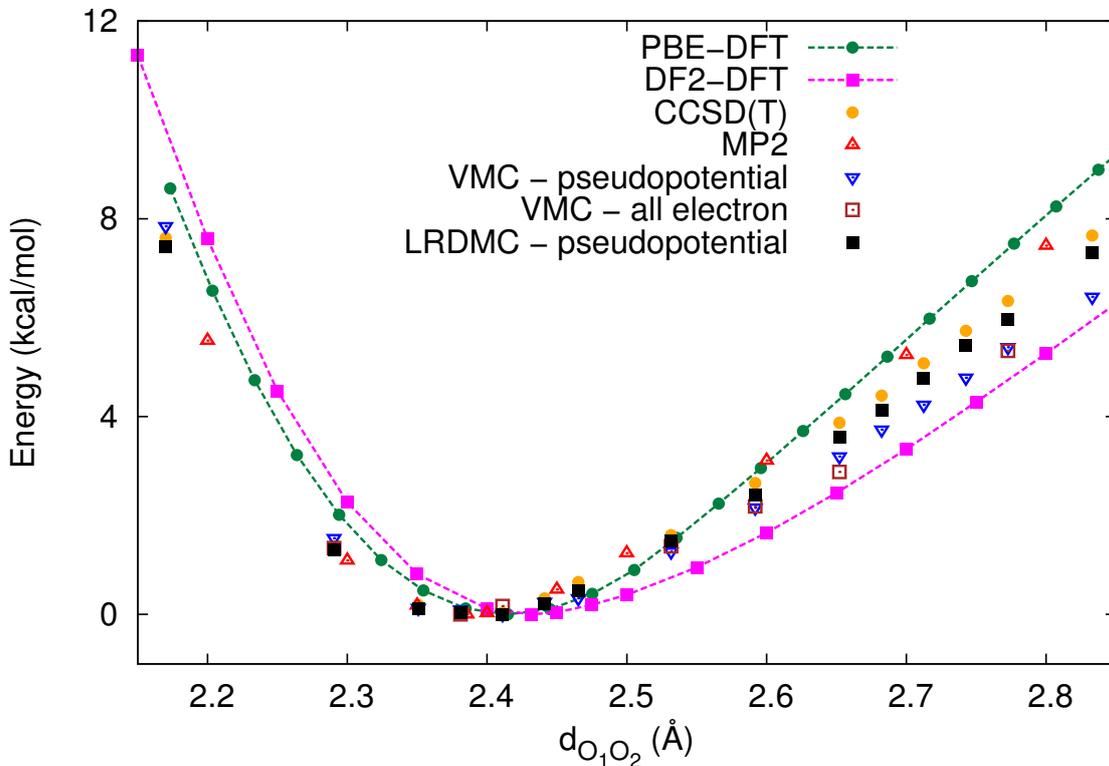


Figure 4.2 – Potential energy curve (kcal/mol) of the Zundel ion projected on the central \overline{OO} distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory, except CCSD(T). Each curve has its minimum as reference point.

First of all, we notice that the PBE functional tends to overestimate the electronic energy by about 1 kcal/mol in both the short (~ 2.25 Å) and large (~ 2.7 Å) \overline{OO} distance regions. Moreover, the PBE functional predicts a slightly too large $d_{O_1O_2}$ distance (by about 0.02 Å) for the minimum geometry configuration. Therefore, we do not expect this functional to be accurate enough to describe properly the PT physics in the protonated water dimer. The accuracy of the DF2 functional is even worse than the PBE approach since DF2 underestimates by 1.5 kcal/mol the electronic energy for large $d_{O_1O_2}$ values (2.7 – 2.9 Å). The DF2 functional certainly overestimates the polarizability of the Zundel ion when one stretches the \overline{OO} distance. Besides, the whole PES obtained by DF2 is remarkably shifted to larger inter-oxygen distances, suggesting that this functional might attribute a too large number of C_s -Inv configuration during the Zundel dynamics. Consequently, the DF2 functional cannot provide the necessary accuracy to correctly describe the PT process in the $H_5O_2^+$ ion. For the QMC simulations, we used the Jastrow Antisymmetrized Geminal Power wave function (JAGP WF, in Eqs. (1.38) and (1.40) devised by Dagrada and coworkers^[72]. This QMC WF is built as follows. On the one hand, the basis set describing the AGP matrix contains 12 hybrids orbitals: 8 for the oxygen atoms and 2 for the hydrogen atoms. On the other hand, the orbitals in the Jastrow factor are developed on a primitive basis set, with 9 basis functions

for the oxygen atoms and 3 for the hydrogen atoms. A detailed description of the WF will be provided in Section 5.1, to study the water dimer. We checked first the quality of the Burkatzki-Filippi-Dolg (BFD) pseudopotential^[80] of the oxygen atom which is, similarly to the norm-conserving pseudopotentials used in DFT, used to reduce the computational cost of our calculations. To that purpose, we carried out some VMC energy optimizations of the JAGP WF either taking into account explicitly all the electrons of the Zundel ion or using the BFD pseudopotential. The agreement between these two approaches further assesses the quality of the pseudopotential that will be safely used for all the subsequent QMC calculations. The VMC results are close to (within 1 kcal/mol) the reference CCSD(T) calculations, despite a slight underestimation of the slope of the curve for large \overline{OO} distances (≥ 2.55 Å). We emphasize that this deviation of the VMC method with respect to the reference CCSD(T) result is observed in a region that will be rarely explored during the Zundel dynamics, with no major consequences on the description of PT. The LRDMC approach, which yields the best QMC correlation energy, shows a remarkable agreement with the state-of-the-art CCSD(T) values which all are in the range of ~ 0.3 kcal/mol, hence very close to the targeted chemical accuracy. The MP2 approach is also, as expected, in very good agreement (within 0.5 kcal/mol) with the CCSD(T) energy estimates for any oxygen-oxygen distance. We highlight that, at variance with the DFT results, all these highly correlated techniques predict the same value of the equilibrium \overline{OO} distance, indicating that the region around 2.39 Å is very sensitive to the electronic correlation of the system. Therefore, it is clear that an extremely accurate description of the electronic correlation is required to properly describe the PT physics in the protonated water dimer, and by extension, of water clusters and the bulk liquid as well.

We have demonstrated the ability of the QMC approach, and more particularly its VMC variant, to describe accurately the potential energy landscape of the protonated water dimer as a function of the inter-oxygen distance. This is not the case of the DFT approach where the position of the minimum energy configuration is not correctly predicted with any employed functional. The next step is to ensure that beyond the good description of the Zundel PES, the QMC approach is also able to properly describe the geometric properties of this cation.

4.1.2 Equilibrium geometries

The slope of the potential energy curve presented in Figure 4.2 is related to the behavior of the excess proton H^+ in the Zundel ion. We elucidate this property by performing a series of structural optimizations at various oxygen-oxygen distances $d_{O_1O_2}$ using different electronic structure methods.

The results are reported in Figure 4.3 where we represent the separations between the excess proton and the two central oxygen atoms O_1 and O_2 (atom labels are indicated in Figure 4.1). The solid curves represent the DFT geometries obtained with the PBE (dark green) and DF2 (magenta) functionals, respectively. The QMC geometries, obtained at the VMC level, are represented by blue circles, while the quantum chemistry configurations are indicated by black crosses and red squares for the MP2 and the CCSD(T) methods, respectively. First of all, let us emphasize that all the employed theories, including the DFT approach, predict a fully symmetric C_2 minimum for the Zundel cation. Nevertheless, as previously mentioned, the

predicted values of the equilibrium oxygen-oxygen distance $d_{\text{O}_1\text{O}_2,\text{eq}}$ strongly depends on the quality of the treatment of the electronic correlation present in the system. Indeed, the blue, red and black vertical dashed lines indicating the equilibrium oxygen-oxygen distance $d_{\text{O}_1\text{O}_2}$ almost superimpose while the dark green and magenta ones are shifted to larger \overline{OO} distances.

Using Figure 4.3, let us interpret the shape of the curves describing the geometry of the H_5O_2^+ ion when one shortens or stretches the oxygen-oxygen distance $d_{\text{O}_1\text{O}_2}$ around its equilibrium value. At short \overline{OO} distances, the repulsive Coulomb interaction between the excess proton and its neighboring oxygen atoms is large, leading to an increase of the total electronic energy of the system. Anyway, to minimize this unfavorable interaction, the central proton leaves out of the straight line defined by the two oxygen atoms. This explains the slower decrease of the corresponding oxygen-proton distances $\overline{O_1H^+}$ and $\overline{H^+O_2}$ around $d_{\text{O}_1\text{O}_2} = 2.1 - 2.2$ Å. On the contrary, when the oxygen-oxygen distance \overline{OO} reaches a critical value d_c , the excess proton is no longer evenly shared between its two oxygen neighbors. In this situation, the localized proton is covalently bound to the O_1 or the O_2 atom and the cluster no longer belongs to the C_2 symmetry group. Indeed, at large \overline{OO} distances (≥ 2.55 Å), the protonated water dimer is appropriately described by a $\text{H}_3\text{O}^+ + \text{H}_2\text{O}$ complex which is C_s -Inv.

The key point is the dependence of the value of the critical distance d_c , defining the frontier between the C_2 -symmetric Zundel and the $\text{H}_3\text{O}^+ + \text{H}_2\text{O}$ regions, with the employed level of theory for the electronic description of the problem. The PBE functional overestimates the correct critical oxygen-oxygen distance d_c by about 0.1 Å with respect to the CCSD(T) result, due to an overestimation of the PES slope at such \overline{OO} distances. We therefore expect this functional to underestimate the PT rate in water clusters and aqueous systems. Including Van der Waals (VdW) and dispersions effects via the DF2 functional improves the protonated water dimer geometries, with a more accurate value of $d_{c,\text{DF2}} = 2.46$ Å. The PBE deficiencies are however not fully recovered by the DF2 functional.

For accurate highly correlated methods such as MP2, QMC, and CCSD(T) approaches, the obtained geometries are in excellent agreement for any oxygen-oxygen distance, even in the critical region $\overline{OO} = 2.4 - 2.45$ Å defining the transition between the Zundel complex and the hydrated hydronium ion. In particular, the QMC (JAGP) and the CCSD(T) perfectly match to each other and both theories predict the critical oxygen-oxygen distance $d_c = 2.39$ Å. We also performed a series of VMC structural relaxations with the simpler JSD WF (Eq. (1.37)) to verify that the accurate QMC description of the protonated water dimer geometric properties does not significantly depend on the QMC WF form. The MP2 approach, although accurate, displays very minor discrepancies (≤ 0.03 Å) with respect to the QMC and the CCSD(T) results for $\overline{OO} = 2.4 - 2.45$ Å. It thus confirms that the QMC approach constitutes the most promising candidate to provide a correct description of the H_5O_2^+ geometric properties at a moderate computational cost.

In Table 4.1, we report the equilibrium oxygen-proton $\overline{OH^+}$ and the oxygen-hydrogen intramolecular \overline{OH} distances for various levels of theory. For some of them (PBE- and BLYP-DFT, MP2 and CCSD(T)), we compare the results with those obtained in literature to ensure the reliability of our structural optimizations. We first stress that all tested methods, included

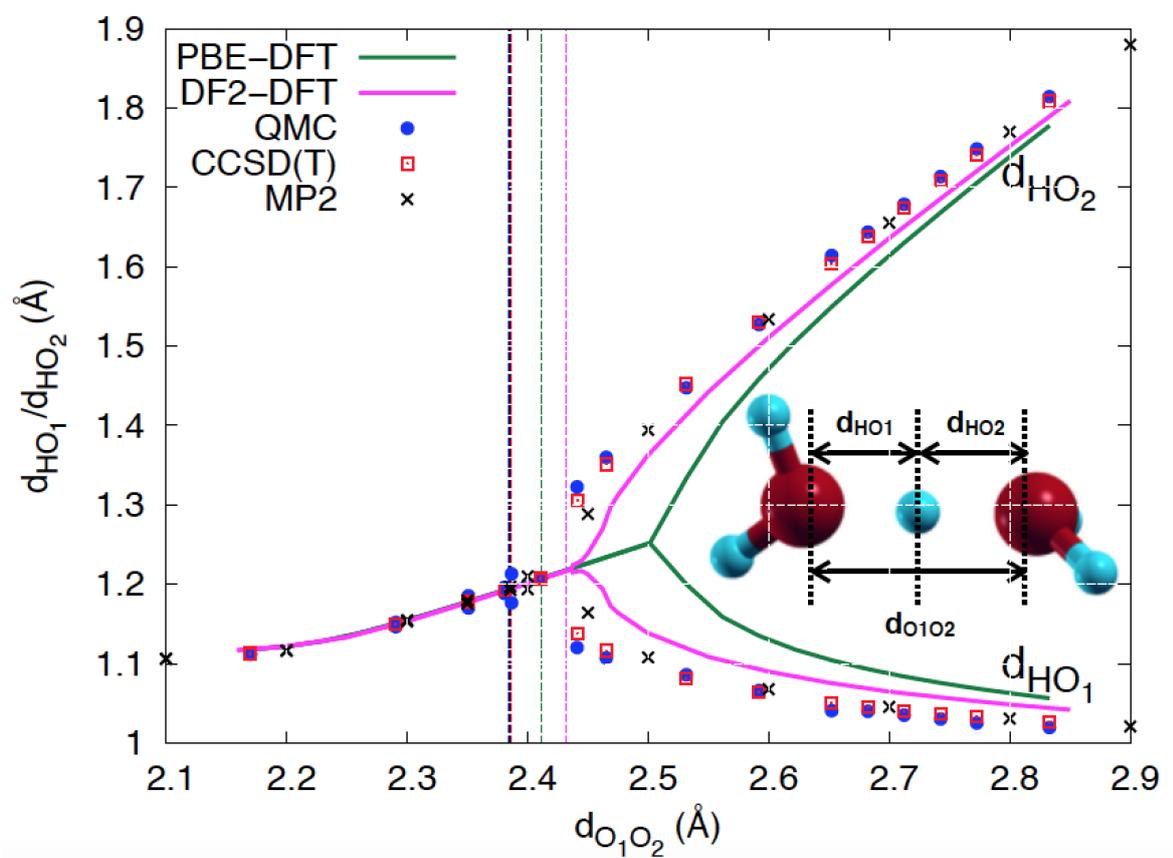


Figure 4.3 – Separations (Å) between the two central oxygen atoms and the excess proton as a function of the reaction coordinate $d_{O_1O_2}$ for different computational methods. Vertical dashed lines indicate the equilibrium $d_{O_1O_2}$ for each method.

the DFT approach, predict a C_2 -symmetric minimum for the protonated water dimer, with an excess proton evenly shared between its two neighboring oxygen atoms. We also note that all our results are in very close agreement (within 0.01 Å) with previous studies, making us confident in the accuracy of our electronic structure calculations. To further verify the convergence of quantum chemistry calculations (DFT, MP2 and CC) with the basis set size, we carried out a series of structural relaxations at the DFT level with the quadruple zeta (QZ) basis set. In this case, no noticeable difference with the TZ results has been found (distances are all converged within 0.0001 Å). The very minor discrepancies between our calculations and the ones reported in the literature can thus presumably be attributed to the different geometric structure optimization algorithms.

As already mentioned, the equilibrium $\overline{O_1O_2}$ distances predicted by DFT are overestimated by about 0.02 Å, 0.035 Å and 0.045 Å, for the PBE, BLYP and DF2 functionals, respectively. We remark that the minimum geometry computed using the B3LYP functional is in an overall better agreement with the reference CCSD(T) result if compared to the other functionals. This result is however less accurate than those obtained with more advanced techniques employed to deal with the electronic correlation, such as the QMC and the MP2 methods. Indeed, the latter ones display an excellent agreement with the CC golden standard, in particular for the equilibrium oxygen-oxygen distance $\overline{O_1O_2}$ whose values coincides with the CCSD(T) estimate, within a 0.005 Å error. The intramolecular \overline{OH} distances are also accurately reproduced in the QMC and MP2 approaches, while they tend to be slightly overestimated in the DFT calculations, except for the DF2 functional.

Theory	$\overline{O_1O_2}$	$\overline{O_1H^+}$	$\overline{H^+O_2}$	$\overline{O_1H_1}$	$\overline{O_1H_2}$
DFT-PBE ^[72]	2.4111	1.2074	1.2074	0.9697	0.9691
DFT-PBE	2.4093	1.2061	1.2061	0.9738	0.9732
DFT-DF2	2.4315	1.2166	1.2166	0.9673	0.9668
DFT-BLYP ^[223]			1.2172	0.9784	0.9778
DFT-BLYP	2.4216	1.2121	1.2121	0.9761	0.9755
DFT-B3LYP	2.3973	1.1997	1.1997	0.9676	0.9669
QMC JAGP with pseudo ^[72]	2.3847(5)	1.1930(5)	1.1942(8)	0.9605(8)	0.9650(8)
QMC JAGP all electron ^[72]	2.3905(4)	1.1944(6)	1.1989(5)	0.9630(7)	0.9628(6)
CCSD	2.3802	1.1918	1.1918	0.9646	0.9640
CCSD(T) ^[65]	2.3864	1.1950	1.1950	0.9686	0.9682
CCSD(T)	2.3853	1.1945	1.1945	0.9672	0.9667
MP2 ^[223]			1.1934	0.9706	0.9701
MP2	2.3859	1.1949	1.1949	0.9676	0.9670

Table 4.1 – Geometric properties (distances in Å) of the C_2 -symmetry minimum of the protonated water dimer, comparison between different computational methods.

To go further with the geometry analysis, we also report in Table 4.2 the characteristic angles of the protonated water dimer, namely: the intermolecular $\angle O_1H^+O_2$, the intramolecular $\angle H_1O_1H_2$ and the two dihedral ω - $H_1O_1H^+O_2$ and ω - $H_2O_1H^+O_2$ angles. Similarly to the DF2 functional, the B3LYP calculation predicts a too large value for the intermolecular $\angle O_1H^+O_2$ angle. Since this angle is clearly related to the position of the excess proton with respect to

the straight line defined by the two oxygen atoms, it suggests that the effective Coulomb interaction might be overestimated. We therefore expect the B3LYP functional to exhibit a too large slope of its PES for short oxygen-oxygen distances, as for the DF2 functional. The intramolecular $\angle H_1O_1H_2$ angles are correctly reproduced within the DF2 approach, except the BLYP/B3LYP functionals. A good agreement (within 0.3°) with the reference CCSD(T) result is also found for both the MP2 and the QMC techniques. The observables that depend the most on the employed theory to describe the electronic structure of the protonated water dimer are the dihedral angles $\omega\text{-}H_1O_1H^+O_2$ and $\omega\text{-}H_2O_1H^+O_2$. Indeed, such quantities are related to the soft vibration modes of the system than can strongly differ from one theory to another. In particular, the DF2 functional predicts dihedral angles that are dramatically too large, making this functional unreliable for a correct description of the geometric properties of the $H_5O_2^+$ ion around its minimum energy configuration. Similarly to PBE and other GGA functionals such as BLYPP, DF2 cannot be used to provide an accurate picture of the PT physics in the protonated water dimer and larger water clusters. The dihedral angles predicted within a QMC framework are quite accurate, if compared to the DFT results, with respect to the CCSD(T) benchmark values. We attribute this discrepancy to the intrinsic noise that affects the structural relaxation in the QMC approach which biases the correct sampling of the softest vibration modes of the system. However, the overall ground state geometry of the Zundel ion is appropriately described by QMC techniques and better than using DFT.

Theory	$\angle O_1H^+O_2$	$\angle H_1O_1H_2$	$\omega\text{-}H_1O_1H^+O_2$	$\omega\text{-}H_2O_1H^+O_2$
DFT-PBE ^[72]	173.661	109.161	295.690	163.809
DFT-PBE	174.359	109.887	297.562	161.459
DFT-DF2	175.7947	109.243	307.593	177.981
DFT-BLYP ^[223]	173.6			
DFT-BLYP	174.694	110.161	298.863	160.642
DFT-B3LYP	175.214	110.525	300.118	159.303
QMC with pseudo ^[72]	174.71(7)	109.16(9)	293.5	158.5
QMC all electron ^[72]	174.43(9)	109.40(7)	296.6	162.0
CCSD	173.949	109.143	296.648	162.718
CCSD(T) ^[65]	173.730	108.8	295.3	163.6
CCSD(T)	173.6780	108.850	295.731	163.507
MP2 ^[223]	173.7			
MP2	173.5157	108.815	295.780	163.445

Table 4.2 – Geometric properties (angles in $^\circ$) of the C_2 -symmetry minimum of the protonated water dimer, comparison between different computational methods.

To summarize, we have demonstrated that at variance with DFT, the QMC approach is able to accurately describe not only the energetic properties but also the evolution of the protonated water geometry as a function of the oxygen-oxygen distance. Indeed, a very close agreement with the reference CCSD(T) results is observed, particularly in the critical region defining the frontier between the C_2 -symmetric Zundel complex and the asymmetric one. Despite giving very satisfactory results, the MP2 theory exhibits a poorer scaling with the system size than the VMC approach, for a comparable accuracy. In the following, we will therefore use the VMC framework to study the PT processes at finite temperature in the protonated dimer,

and later in the hexamer (Chapter 6).

4.1.3 Accuracy of the Quantum Monte Carlo approach

In the previous Subsections, we have established that the accuracy of the QMC approach is certainly very close to the reference CCSD(T) method. Nevertheless, at some points such as the description of the two dihedral angles, the QMC estimates are far from being perfect. Our goal is now to objectively determine the place of the QMC approach in the hierarchy of the advanced electronic structure methods for the specific case of the water clusters.

To assess the accuracy and the reliability of the QMC geometries with respect to those obtained via the PBE and the DF2 functionals, we decided to carry out a series of CCSD and CCSD(T) energy gradient estimations, keeping the geometry of the system frozen. In principle, when one performs such a calculation on the equilibrium geometry of the corresponding theory (CCSD or CCSD(T) geometries in our case), the obtained gradients should be equal to zero¹. Therefore, the finite values of the computed gradients give us precious information on the quality of the input configuration. The more accurate the geometry, the lower the energy gradients. The PBE, DF2 and QMC optimal geometries employed to perform such CC calculations are chosen within the range $d_{\text{O}_1\text{O}_2} = 2.25 - 2.65 \text{ \AA}$, which corresponds to the very large majority of the configurations that will be explored by the H_5O_2^+ ion during its dynamics. The computed CCSD (top) and CCSD(T) (bottom) gradients are reported in Figure 4.4. We isolate in the left panels the gradient module acting on the excess proton whereas the average of the gradient module applied on each atom, excluding the two oxygen atoms, of the Zundel cation is represented in the right panels. In practice, the energy gradients are estimated numerically thanks to an option of the MOLPRO suite of codes that generates symmetric displacements of each considered atom. Then, the energy gradients are straightforwardly computed, using finite differences, namely:

$$\frac{\partial E}{\partial q} = \frac{E(q + \delta q) - E(q)}{\delta q}, \quad (4.1)$$

with the user-specified² elementary displacement is set to $\delta q = 0.01 \text{ \AA}$. In Figure 4.4, the dark green curves are the obtained gradients with some PBE geometries, while the DF2 results are plotted in black and the QMC ones in red.

When looking at the upper panels of Figure 4.4, it is clear that, working at the CCSD level of theory, the QMC geometries are much more accurate than the configurations predicted by the PBE and the DF2 functionals. Indeed, the energy gradients computed using the QMC geometries never exceed 1 mHa/\AA (0.6 kcal/mol/\AA) for the excess proton and 3.5 mHa/\AA (2.2 kcal/mol/\AA) for the other atoms, which is particularly small. As expected, PBE geometries lead to larger values of the CCSD energy gradients than those obtained in DF2, confirming the $\text{QMC} > \text{DF2} > \text{PBE}$ hierarchy already established in the previous Subsection.

¹at full convergence.

²We verified on a single geometry that the result remains unchanged by setting $\delta q = 0.005 - 0.2 \text{ \AA}$.

4.1. Zero temperature results

The analysis of the bottom panels of Figure 4.4 is interesting since it tells us that the aforementioned hierarchy is completely shaken within the CCSD(T) approach. Indeed, the CCSD(T) energy gradients computed on the QMC geometries are no longer the smallest ones, at variance with the CCSD result. The average QMC gradients values are larger than the ones computed within the DFT framework, by about 3 mHa/Å (1.9 kcal/mol/Å) per atom. Nevertheless, the PBE, DFT and QMC excess proton gradients display very similar values (bottom left panel), indicating that the major bias in the QMC geometries certainly comes from the description of the intramolecular \overline{OH} distances. To justify this remark already reported during the analysis of Table 4.1, we argued that this discrepancy with the reference CCSD(T) result might come from the size of the employed basis to perform CCSD(T) calculations.

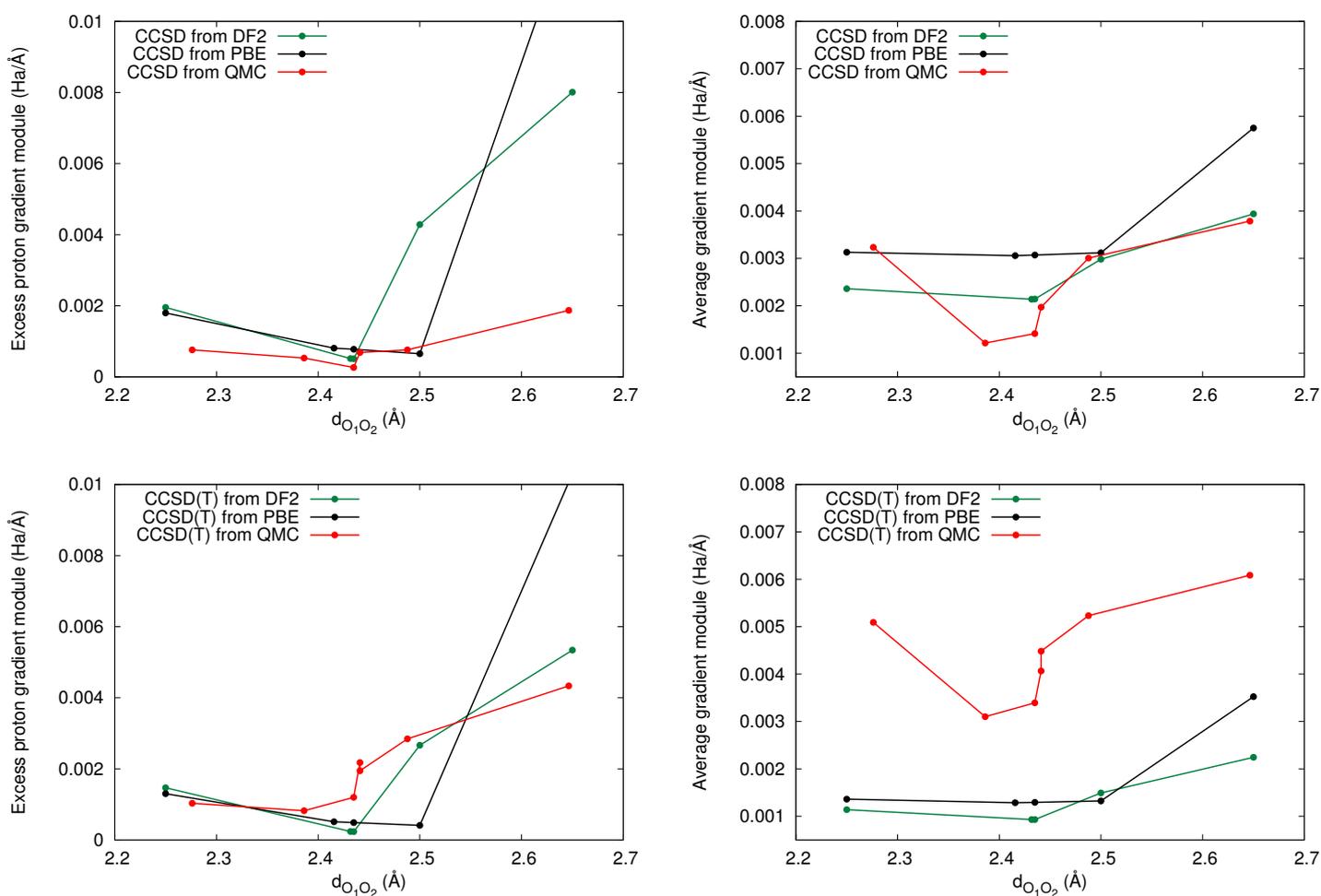


Figure 4.4 – Excess proton (left) and average (right) gradients calculated for the Zundel ion at the CCSD (top) or CCSD(T) (bottom) level of theory for various oxygen-oxygen distances $d_{O_1O_2}$. Results are obtained with the TZ basis set.

To verify this assumption, we performed a series of CCSD and CCSD(T) energy estimations of the obtained QMC and DF2 geometries at $d_{O_1O_2} = 2.65$ Å with increasing the basis set size from the TZ³ to the larger Quintuple Zeta (5Z). The computed energy gradients are

³We recall to the reader that when it is not specified, the TZ basis set is the one we employed to carry out

plotted in Figure 4.5 as a function of the basis set size, growing from TZ to 5Z. First of all, we remark that the gradients applied to the excess proton or to the other atoms (except for the two oxygen atoms which are pinned) tend to vanish for the QMC geometry when the basis set size increases for both CCSD and CCSD(T) calculations. The behavior of the computed gradients on the DF2 geometry is opposite to its QMC counterparts, the CCSD and CCSD(T) gradients being larger with an increasing number of gaussians. It is clear from the left panel of Figure 4.5 that the predicted position of the central proton is better predicted within the QMC approach than using the DF2 approach, as already discussed in the previous Subsection. Moreover, when inspecting the right panel of Figure 4.5, we notice that, in the complete basis set limit, the configurations predicted by the QMC approach are certainly closer to both CCSD and CCSD(T) geometries than the DF2 ones. This suggests once again that, despite its slightly less accurate description of the intramolecular \overline{OH} distances, the QMC approach is trustable to accurately describe PT in water since it provides here an excellent description of the proton localization. Last but not least, we remark that at variance with the DF2 results, the CCSD(T) gradients on the QMC geometry at $d_{O_1O_2} = 2.65 \text{ \AA}$ are larger than the CCSD values. We can deduce that the QMC geometries are closer to the CCSD ones than to those obtained by CCSD(T). Based on this result together with Tables 4.1 and 4.2, we can thus establish that the VMC method provides a very accurate description of the geometric properties of the protonated water dimer, with very close agreement to reference CCSD/CCSD(T) calculations. As we will see in Chapter 6, this accuracy is preserved when one increases the cluster size, making the VMC approach a promising candidate to study liquid water or aqueous systems.

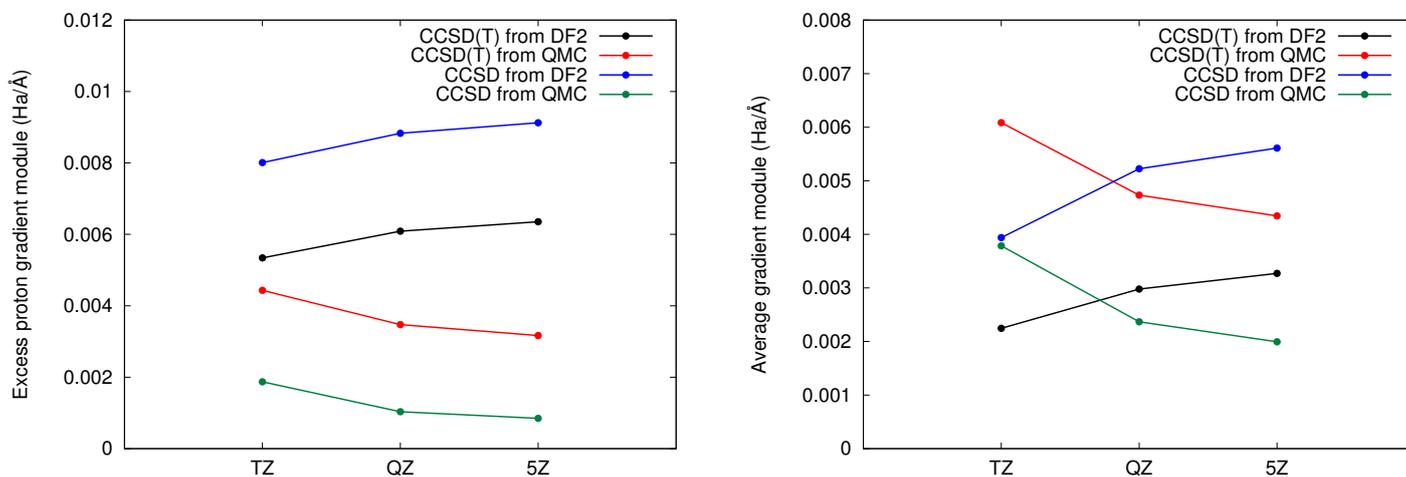


Figure 4.5 – Excess proton (left) and average (right) gradients calculated for the Zundel ion at the CCSD/CCSD(T) levels of theory with various basis sets. For the two tested geometries, the oxygen-oxygen distance is $d_{O_1O_2} = 2.65 \text{ \AA}$.

To summarize, thanks to a detailed energy gradients analysis, we have confirmed that the QMC - and more specifically the VMC - approach provides an accuracy close, or almost equivalent to the golden reference CC method, with a milder scaling with the system size. Using to this very significant advantage of the QMC approach and the methodological de-

all our quantum chemistry calculations.

velopments introduced in Chapter 3, we are now almost ready to explore the impact of both thermal and nuclear quantum effects on the proton mobility in small protonated water clusters. Indeed, the very last step before carrying out finite temperature Langevin simulations of such systems, with noisy ionic QMC forces, consists in checking that this intrinsic noise does not bias the dynamics.

4.2 Benchmark calculations

In this Section, we apply the methodology developed in Chapter 3 to perform QMC-based Langevin Dynamics (LD) or Path Integral LD (PILD) simulations of the Zundel complex in the gas phase at low ($T = 50$ K) and room ($T = 300$ K) temperature, with and without NQE. A schematic representation of our approach is given in Figure 4.6. We emphasize that we perform here two distinct stochastic dynamics within the same system: the first one concerns the electrons whose positions are sampled by a QMC random walk in the configurational space of the electronic coordinates (blue points). The ions (oxygen atoms in red and hydrogen atoms in white) also move on the electronic PES according to a Markovian (Langevin) Dynamics to keep the system at a constant temperature T .

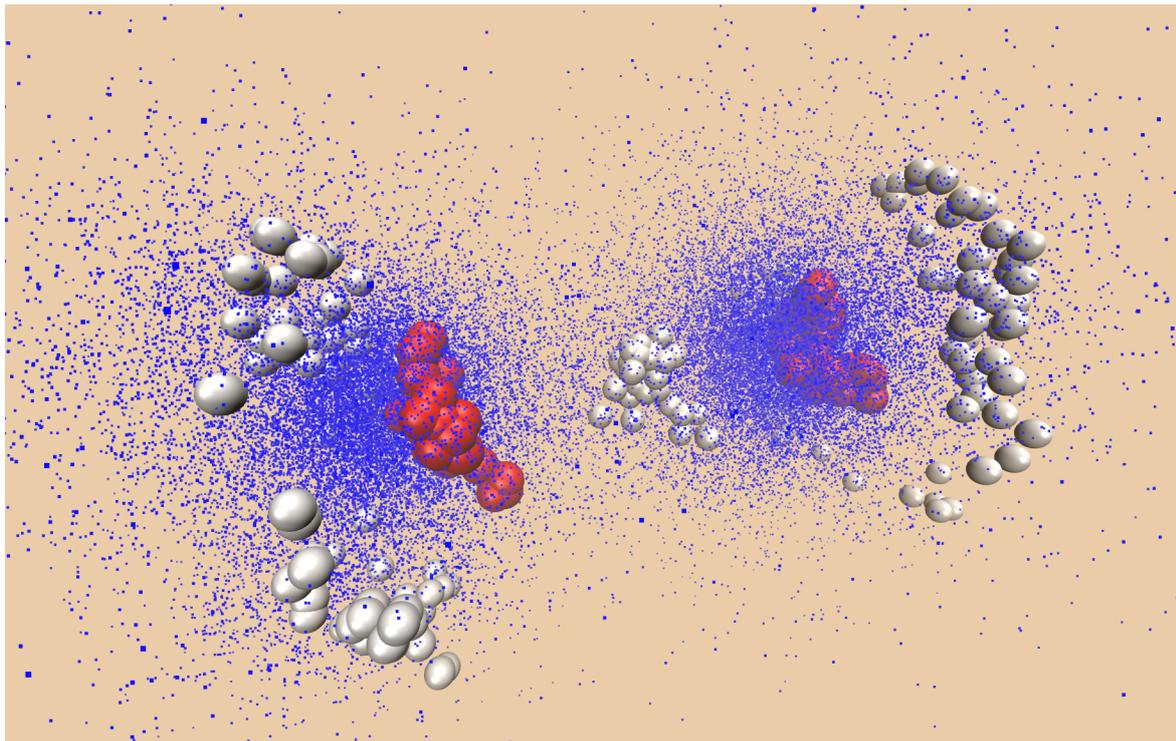


Figure 4.6 – Intuitive representation of the QMC-driven (Path Integral) Langevin Dynamics developed in this thesis and applied to the Zundel ion. The electronic configurations adopted by the QMC walkers are depicted by the blue points, while the oxygen and the hydrogen atoms are represented by red and white spheres, respectively.

To benchmark our results, we first carried out Classical Momentum-Position Correlator (CPMC) and Path Integral Ornstein Uhlenbeck (PIOUD) calculations using the analytical CCSD(T) PES provided by Huang *et al.*^[65]. In that case, the ionic forces are computed as finite dif-

ferences of the potential energy, with an increment $\delta q = 10^{-4}$ Bohr in the geometry. The dynamics, with either QMC forces or CCSD(T)-parametrized PES, is propagated with a time step $\delta t = 1$ fs until we obtain fully converged atomic distributions. As previously discussed (Section 3.2.4), the input friction (γ_{BO} for CMPC, γ_0 for PIOUD) is set to 1.46×10^{-3} a.u., in order to both ensure a good thermalization of the system and minimize the potential energy autocorrelation time τ_V . Consequently, that value also maximizes the diffusion of the nuclei between two subsequent LD iterations, and guarantees an efficient phase space sampling. We use $P = 128$ beads at low temperature whereas $P = 32$ beads are enough to fully recover NQE at room temperature, which is in agreement with Ref. 201. We stopped the production run after 20 ps of dynamics, which is the minimum required to obtain converged static properties^[111]. At the end of each simulation, the average temperature, the virial and primitive kinetic energy estimators, the energy fluctuations, and the evolution of the energy gradients with respect to the electronic parameters are checked to ensure the reliability of the simulation. In the electronic QMC part, the correct BO PES is provided by the full convergence of the WF parameters at each ionic configuration. On the other hand, to follow as close as possible the PES during the ion dynamics, the values of the energy gradients with respect to the WF parameters must be lower than 3-4 times their standard deviation.

4.2.1 Validation of the classical dynamics

We first start with the simpler case of the classical Zundel. We present in Figure 4.7 the normalized oxygen-oxygen (g_{OO}) (top panels) and oxygen-(excess) proton (g_{OH}) (bottom panels) distributions as a function of the inter-oxygen distance. These Radial Distribution Functions (RDFs) are obtained at low (left panels) and room (right panels) temperature for classical nuclei. Both the VMC and benchmark CCSD(T) results are shown.

We observe a good overall agreement⁴ between the Pair Correlation Functions (PCFs) g_{OO} and g_{OH} obtained with the reference CCSD(T) calculations and our VMC-based simulations at the two considered temperatures, proving that our simulations are free of biases. This is very encouraging in the perspective of simulating larger systems for which there is no available parametrized PES on highly accurate quantum chemistry calculations. This is for instance the case of the protonated water hexamer whose fascinating properties will be studied in Chapter 6 at an unprecedented level of accuracy.

Looking at Figure 4.7, we notice a broadening of the oxygen-oxygen and oxygen-proton RDFs when the temperature of the Zundel ion increases. Indeed, at low temperature ($T = 50$ K), the protonated water dimer is almost frozen and moves a little around its minimum energy configuration, leading to high and sharp peaks in the g_{OO} and g_{OH} RDFs. On the contrary, at room temperature, the thermal energy brought to the system is transformed into kinetic energy. Therefore, the molecular vibrations are more important and the motion of all the atoms constituting the H_5O_2^+ ion, in particular the excess proton, is enhanced. This is characterized by a more diffuse peak in the g_{OO} and g_{OH} RDFs. We also remark that the peak of the oxygen-(excess) proton RDF g_{OH} is symmetric at $T = 50$ K, which is no longer the case

⁴We are aware that the agreement could however be improved, especially as regards as the g_{OH} distributions, because of the shortness of the QMC-MD trajectories.

around ambient conditions ($T = 300$ K). This suggests that, within a classical picture, the central proton is *localized* near an oxygen atom, and the protonated water dimer is better described as a $\text{H}_3\text{O}^+ + \text{H}_2\text{O}$ complex (hydrated hydronium) than by a Zundel complex. We will see in the following that this description is not complete, since one should take into account NQE to properly understand the temperature-induced behavior of the excess proton in this system.

Finally, we highlight that in the classical case, the intrinsic noise present in the QMC forces tends to spoil the sampling of the instantaneous oxygen-oxygen and oxygen-proton distances, which has to be accurate because the resulting distributions are extremely sharp. This explains the presence of small but noticeable discrepancies between the RDFs extracted from VMC-driven LD simulations and those obtained via CCSD(T) forces. To reach a perfect agreement, one should in principle prolongate the MD trajectories on large timescales (about $t_{\text{traj}} = 50$ ps). We did not generate such long LD trajectories to save computational resources⁵ to study more complex systems, such as the protonated hexamer (Chapter 6), with a richer and unexplored physics than in the already well-understood Zundel ion.

Our methodological development to perform QMC-driven classical LD passed the stringent test to provide almost exactly the same RDFs as those extracted from reference simulations with an analytical CCSD(T) PES. The following step is to ensure that the machinery (both the PIOUD algorithm and the bead-grouping technique) set in Chapter 3 to perform fully quantum simulations of any water cluster with QMC forces also produces the correct distributions.

4.2.2 Validation of the quantum dynamics

In this Subsection, we will focus our attention on two issues. The first one, is to ascertain that our novel PIOUD integrator works efficiently with the noisy ionic QMC forces, with no irreversible biases in the ion dynamics, as we did in the classical case for the CMPC integrator. Afterwards, we will discuss the validity of the bead-grouping technique. Indeed, this approximation to reduce the computational cost of our QMC-driven simulations, has only been validated on a simple system, the H_2 molecule, and needs to be verified on a system with different chemical elements. To that purpose, we performed an additional simulation at low temperature ($T = 50$ K) by taking $N_{\text{groups}} = 16$. As explained in Subsection 3.2.2, this implies that instead of considering P distinct WFs $\Psi_{\lambda_{\text{av}}}(\mathbf{R}; \mathbf{q}^{(k)})_{k=1, \dots, P}$ whose electronic parameters $\{\lambda_{\text{av}}\}$ are averaged over the $P = 128$ beads constituting the necklace, the quantum system is described by averaging the parameters $\{\lambda_{\text{av}}^{(l)}\}_{l=1 \dots N_{\text{groups}}}$ describing each QMC WF among $N_{\text{groups}} = 16$ different groups of 8 beads. The corresponding dynamics is in principle more accurate but also more noisy with respect to the $N_{\text{groups}} = 1$ case, so probably less convenient to exploit, disregarding the fact that, in this case, the computational cost is about $N_{\text{groups}} = 16$ times larger.

In Figure 4.8 we present the normalized oxygen-oxygen g_{OO} (top panels) and oxygen-(excess)

⁵Computational expenses to generate such QMC-driven trajectories are discussed at the very end of the next Subsection.

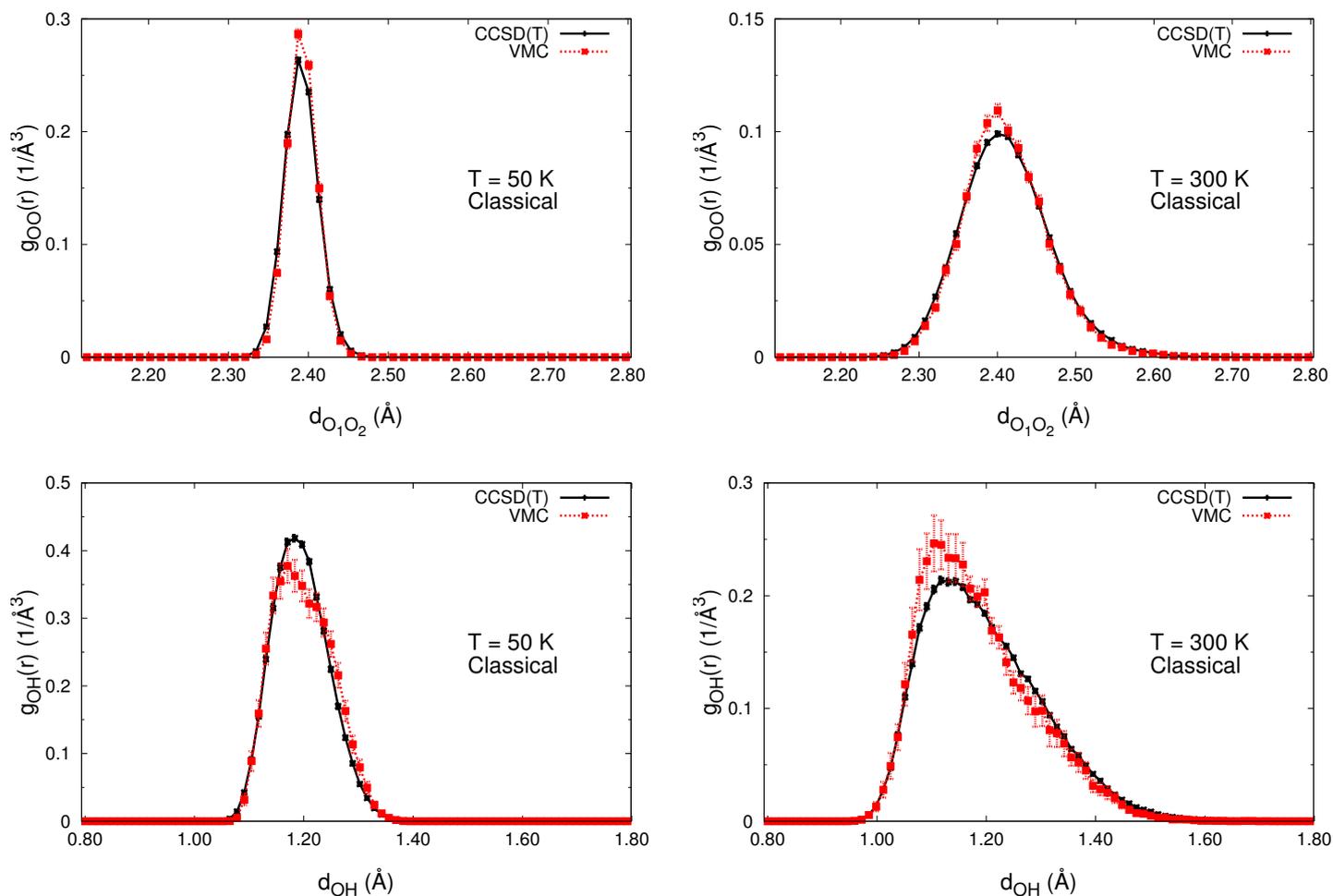


Figure 4.7 – Normalized oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions obtained by CMPC-LD simulations at low temperature $T = 50$ K (left) and $T = 300$ K (right). The black curves represent the distributions obtained with analytic CCSD(T) forces, whereas the red curves correspond to VMC-LD dynamics. The friction is set to $\gamma_{\text{BO}} = 1.46 \times 10^{-3}$ a.u.

proton g_{OH} (bottom panels) distributions as a function of the inter-oxygen distance, as in the classical case. These RDFs are obtained at low (left panels) and room (right panels) temperature for quantum nuclei. We first notice that at $T = 50$ K, the quantum effects significantly broaden the g_{OO} and g_{OH} RDFs. Consequently, the g_{OO} distribution is renormalized by a factor of 3, while the height of the g_{OH} peak is divided by 2 with respect to its classical counterpart. The NQE have thus a huge effect in the oxygen-oxygen distribution function because the classical system is almost frozen around its zero temperature equilibrium configuration, whereas the ZPE leads to strong quantum fluctuations even at low temperature. At $T = 300$ K, even though NQE are much less significant for the oxygen-oxygen distribution, they are still very important for the oxygen-proton correlation function, where the shapes of the distributions obtained in the classical and in the quantum case are very different. Indeed, the classical g_{OH} distribution is asymmetric around the equilibrium distance, whereas the quantum correlation function is symmetric with much longer tails, indicating the possibility of instantaneous proton hops by quantum tunneling. On the contrary, NQE are less dramatic at room temperature for the g_{OO} radial distribution, as expected from the greater mass of the oxygen atoms which shortens its thermal de Broglie wavelength (Eq. (2.47)).

Our benchmark system is ideal also to check out the quality of the bead-grouping technique described in Subsection 3.2.2 for the electronic parameters in *ab initio* VMC-PIOUD simulations. At room temperature the bead-grouping with $N_{\text{groups}} = 1$ works very well, giving results on the top of the CCSD(T) reference. At low temperature, the agreement between the CCSD(T) reference and the VMC-PIOUD results with $N_{\text{groups}} = 1$ is still good, although some minor discrepancies appear in the tails of the oxygen-oxygen and oxygen-proton distributions. The strongest bias, though still quantitatively acceptable, is present in the g_{OH} function, as this pair distribution is the most sensitive to quantum delocalization effects. By increasing N_{groups} to 16, we improve the peak positions of both g_{OH} and g_{OO} , and the error made in their tails is significantly reduced. This can be simply interpreted by considering the quantum-to-classical isomorphism of Eq. (2.59): hydrogen atoms have a light mass, so the corresponding ring polymers are much more spread than the ones mimicking the quantum nuclei of oxygen atoms. Consequently, the bead-grouping approximation on the optimal electronic parameters $\{\lambda_{\text{av}}^{(l)}\}_{l=1,\dots,N_{\text{groups}}}$ is more severe for hydrogen than for oxygen. The resulting potential energy landscape is thus affected, and displays a larger curvature around its minimum, due to the energy penalty given by the non fully-optimized WFs, being the worst for those beads which are the farthest from the centroid. This effect is apparent in the slightly shorter tails of the PCF, since the corresponding ionic configurations are less visited as they have higher energies. Therefore, a compromise must be found by minimizing the total amount of CPU time spent for a simulation and the desired target accuracy on the structural and static properties of the system. While at 50 K an $N_{\text{groups}} > 1$ should be chosen, at 300 K $N_{\text{groups}} = 1$ gives very accurate results. Thus, at room temperature we are able to carry out a reliable and very accurate fully-quantum dynamics of the Zundel cation in almost the same CPU time as for classical nuclei. More generally, since the hydrogen atom is the lightest chemical element with the strongest NQE because of its very short thermal de Broglie wavelength, we claim that the bead-grouping technique can reasonably be used for the simulation of any extended system with heteroelements (*e.g.* carbon or nitrogen to cite a few).

We also remark that, at variance with the classical case, the QMC intrinsic noise is here helpful in improving the quantum delocalization of the nuclei in the desired regions of the phase space. The phase space sampling efficiency seems to be enhanced in the quantum case with respect to the classical one. The quantum results are unexpectedly easier to converge, and they yield radial distributions with reduced error bars compared to their classical counterparts.

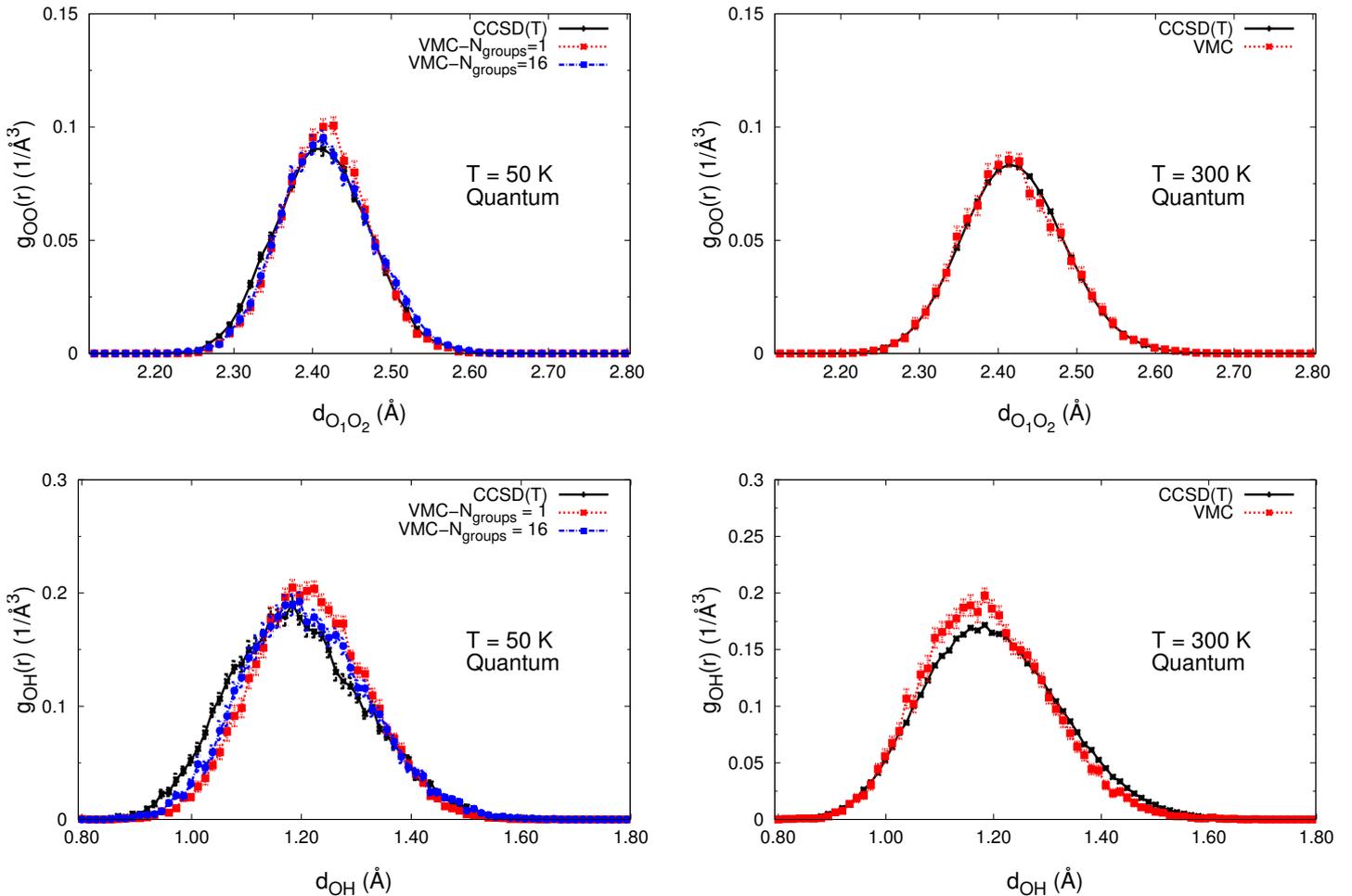


Figure 4.8 – Normalized oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions obtained by PIOUS-LD simulations at low temperature $T = 50$ K (left) and $T = 300$ K (right). The black curves represent the distributions obtained with analytic CCSD(T) forces, whereas the red and blue curves correspond to VMC-PILD dynamics with $N_{\text{groups}} = 1$ and $N_{\text{groups}} = 16$, respectively. Quantum simulations are performed with $P = 128$ beads at $T = 50$ K, and $P = 32$ beads at $T = 300$ K. The friction is set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u.

Last but not least, we discuss the computational cost of our VMC-(PI)LD simulations. The CPU time required on HPC Marconi at CINECA (2.3 GHz 2 x 18-cores Intel Xeon E5-2697 v4 processors) to carry out a 20 ps trajectory of the Zundel ion using our VMC-(PI)LD approach is about 270k core hours. There is no significant difference between classical and quantum

simulations, as long as the path integral dynamics is performed with the bead-grouping approximation with $N_{\text{groups}} = 1$. The target statistics has been detailed in Chapter 3 and is enough to have stable and converged simulations in both electronic and ionic parts. At the given target statistics, relaxing the bead-grouping approximation in the quantum case implies a total computational time increased by a factor $1 \leq N_{\text{groups}} \leq P$ with respect to the fastest case of $N_{\text{groups}} = 1$. In the perspective of applying our novel methodology to larger protonated water clusters such as the protonated water hexamer in Chapter 6, it is worth mentioning that for single-point calculations of equilibrium geometry at zero temperature, it has been estimated^[72] that the six-molecule complex represents a crossing point in the relative efficiency between VMC and CCSD(T) methods. By considering the excellent performances of QMC for parallel computations, this technique is already competitive for small systems as far as the elapsed computational time is concerned. Moreover, it is obvious that a further increase in the cluster size would make the QMC approach considerably favored even in terms of total computational demand, thanks to its milder scaling with the system size.

To conclude, we have made a step forward in the development of a fully quantum dynamics of small protonated water clusters since we demonstrated the ability of our method to properly reproduce accurate results obtained with the most modern quantum chemistry techniques, such as the CCSD(T) method, at a reasonable computational cost. Indeed, a good agreement in the computed RDFs by means of QMC or CCSD(T) approaches is found, for both a classical and a quantum description of the nuclei. We did not however study in detail the impact of thermal and nuclear quantum effects on the excess proton mobility in the Zundel cation yet. This point is discussed in the following Section.

4.3 Proton transfer in the Zundel ion

It is interesting to quantify NQE versus thermal effects, by observing the evolution of the PCFs at increasing temperature or when the quantum delocalization of the nuclei is taken into account. To investigate the proton mobility in the very high temperature regime, we decided to perform two supplemental simulations at $T = 900$ K, with the analytical CCSD(T) PES provided by Bowman and his colleagues^[65]: one with classical ions using the CMPC algorithm and a quantum trajectory generated for $P = 4$ beads thanks to the PIOUD algorithm.

A direct comparison of the evolution of the g_{OO} and g_{OH} RDFs as a function of the temperature is reported in Figure 4.9 for both a classical and a quantum description of the nuclei. For classical particles, there is a clear broadening of the g_{OO} and g_{OH} distribution with increasing temperature, as we can see on the left panels of Figure 4.9. Indeed, the entropy is increased and the instantaneous oxygen-oxygen and oxygen-proton distances are subject to enhanced fluctuations. This is confirmed by examining the upper left panel of Figure 4.10, where we plot the oxygen-proton distance for the two oxygen sites as a function of the inter-oxygen distance. At $T = 50$ K, the classical proton remains extremely localized around the C_2 -symmetry geometric minimum which leads to a very sharp distribution, reported in the top panels of Figure 4.7. At $T = 300$ K, the system has more thermal energy to visit asymmetric configurations with longer oxygen-oxygen distance and the excess proton sitting closer to one oxygen than to the other. This is represented by the two branches in the middle

left panel of Figure 4.10. Even in this situation the symmetric configurations are more often explored than the phase regions forming the two wings. In the very high temperature limit ($T = 900$ K), the thermal energy is so important that the protonated water dimer is sometimes close to the dissociation limit, at large \overline{OO} distances (bottom left panel).

On the contrary, thermal effects are much less important when NQE are taken into account, as apparent from the right panels of Figures 4.9 and 4.10, which show very similar correlation functions at $T = 50$ K and $T = 300$ K. Indeed, quantum fluctuations make the proton able to easily jump to a neighbor oxygen site by quantum tunneling and Zero Point Energy (ZPE), recovering a more symmetric behavior. This is characterized by longer tails in the g_{OH} distribution function compared with the classical ones at $T = 300$ K, and by the absence of wings in Fig. 4.10. Moreover, the 2D distributions are nearly the same at low and room temperature, as we can see in the top right and middle right panels of Figure 4.10. These conclusions are in agreement with previous studies on this system^[172,201]. Interestingly, when one increases further the cluster temperature up to $T = 900$ K, thermal effects start to be significant. Indeed, inspecting the right panels of Figure 4.9, we observe that the RDFs are no longer similar to their low- and room temperature counterparts. More particularly, the quantum g_{OH} RDF is no longer symmetric at $T = 900$ K, indicating that the proton is now localized near its closest oxygen neighbor, as in the classical picture. Indeed, the quantum g_{OH} RDF is very similar to its classical counterpart in the very high temperature limit. This is also confirmed looking at the oxygen-oxygen and oxygen-protons bidimensional distributions obtained at $T = 900$ K, represented in the bottom panels of Figure 4.10, whose shape is very similar for classical and quantum ions. This last observation confirms that NQE are essential to fully understand the microscopic mechanisms involving the proton dynamics in liquid water at *ambient conditions*.

To put it in a nutshell, thanks to the methodological developments introduced in Chapter 3, we have been able to propose a complete investigation of the PT physics in the Zundel ion. Indeed, we have confirmed the paramount importance of NQE to properly describe the excess proton mobility, especially around ambient conditions, which are the most frequent ones in chemical or biological systems. At higher temperatures, the proton tends to be localized by forming a covalent bond with its closest oxygen neighbor. However, despite its utility to benchmark novel approaches as we did in this Chapter, the Zundel ion is not a realistic system to model PT in the bulk water or in aqueous systems. Indeed, this cation is *isolated* and there is no solvation effects due to the presence of surrounding water molecules. To make a further step in the comprehension of PT in water, we should apply our methodology to larger and more realistic systems, such as the protonated hexamer. The latter, that will be extensively studied in Chapter 6, is composed by a Zundel-like core and 4 solvating molecules forming 4 H-bonds with the inner part of the complex. To first provide an accurate description of the interaction between the core and the solvation shell, we will study the water dimer in the next Chapter.

4.3. Proton transfer in the Zundel ion

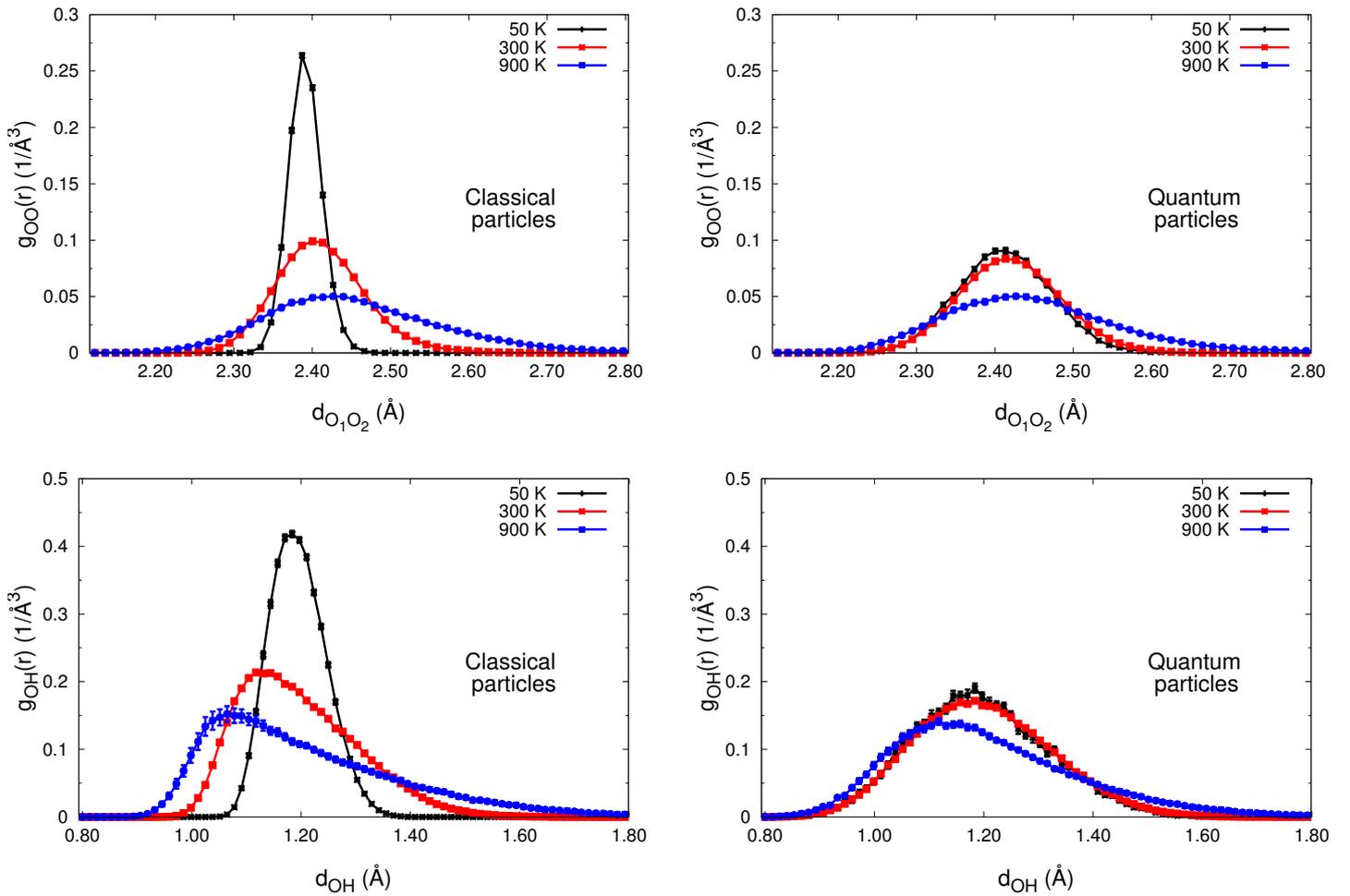


Figure 4.9 – Evolution of the oxygen-oxygen (top) and oxygen-(excess) proton (bottom) distributions as a function of the temperature for the classical (left) or the quantum (right) Zundel ion.

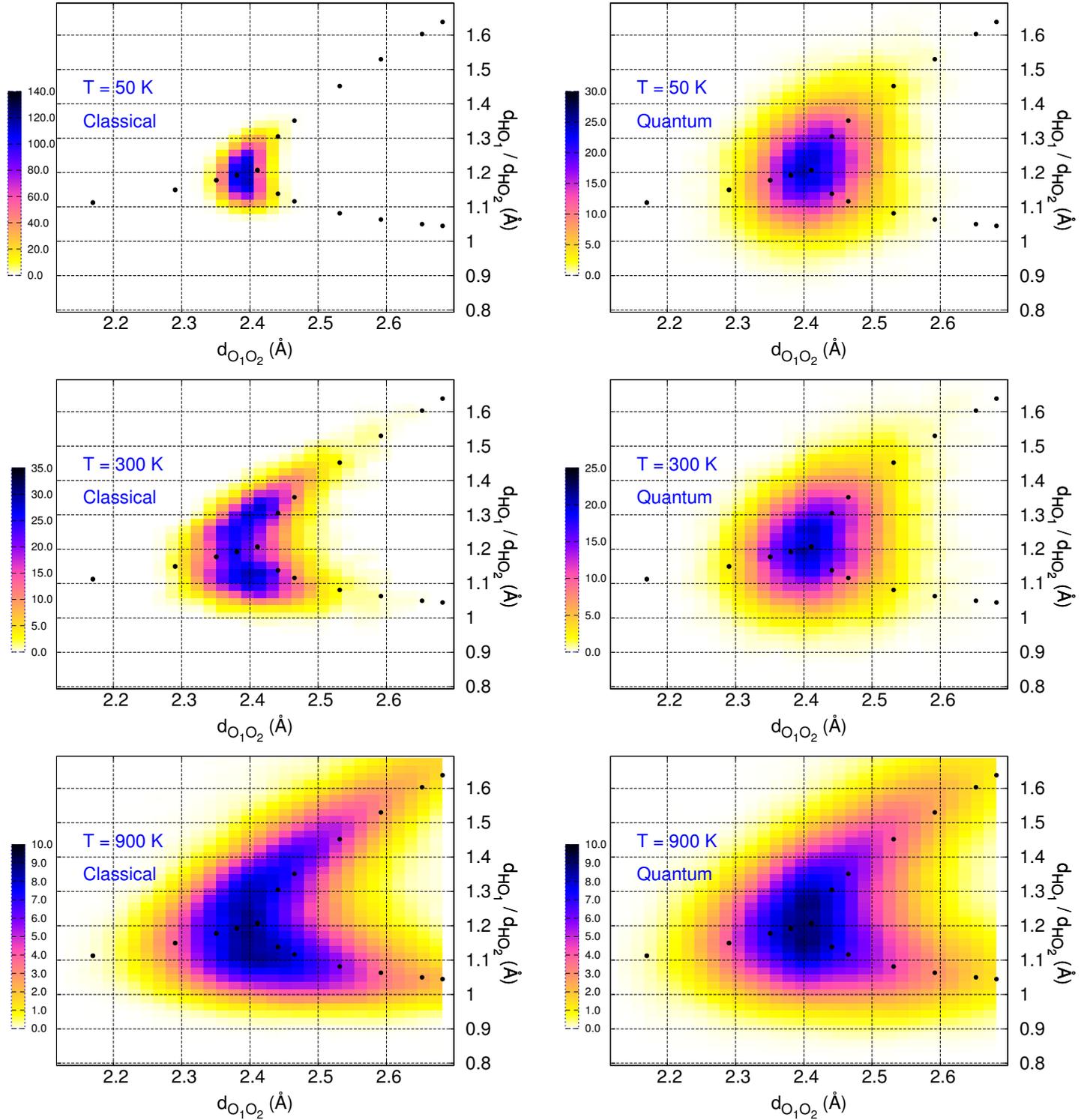


Figure 4.10 – Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC- or CCSD(T)- driven CMPC-LD (left) and PIOUD (right) simulations at $T = 50$ K (top), $T = 300$ K (middle) and $T = 900$ K (bottom). The black circles correspond to the equilibrium geometries of the Zundel ion at zero temperature obtained by CCSD(T) calculations^[72]. Quantum simulations are performed with $P = 128$ beads at $T = 50$ K, $P = 32$ beads at $T = 300$ K and $P = 4$ beads at $T = 900$ K. The friction is set to $\gamma_0 = 1.46 \times 10^{-3}$ a.u.

Part III

On the way to simulate fully quantum water

Interlude: the water dimer

Contents

5.1	Non-bonding water dimer	112
5.2	Bonding water dimer	115
5.2.1	Binding energy	115
5.2.2	Geometric properties	119

AS A CENTRAL PROTOTYPE for hydrogen bonding, the water dimer $(\text{H}_2\text{O})_2$ has been one of the most studied molecular clusters by quantum chemistry calculations since 1968^[231]. Various electronic structure methods (Hartree-Fock, Density Functional Theory, Møller-Plesset and Coupled Cluster approaches to cite but a few) have been used to obtain both the geometry and the Potential Energy Surface (PES) of the dimer. In particular, the computed geometric parameters have been compared with reported experimental results^[232–234]. High-level electronic structure benchmarks are often used to assess the accuracy of lower scaling methods such as new DFT functionals. One of the main thrusts of these theoretical investigations was to estimate the hydrogen bond energy and its decomposition into different components (polarization, exchange, induction and dispersion). The water dimer stationary points and PES are of critical importance in the development of transferable many-body potentials for water. We refer the reader to the reviews of Mukhopadhyay *et al.* for more information about recent experimental^[235] and theoretical advances^[236] on this topic.

In this Chapter, we present some properties (PES and geometric parameters) of the water dimer using different techniques such as the Density Functional Theory (DFT), the Coupled Cluster (CCSD/CCSD(T)) method or the Quantum Monte Carlo (QMC) approach. More particularly, we will benefit from the reduced size of this system to perform, at a very reasonable computational cost, a number of tests to improve the quality of the QMC wave function (WF) in the perspective of simulating larger clusters. The results presented here are necessary to build a reliable QMC WF that is simple enough to be employed on larger scales calculations, at it will be done in Chapter 6.

The Chapter is organized as follows. First, we will check the ability of QMC methods to describe accurately the PES of the non-bonding water dimer depicted in Figure 5.1a. Second, the accuracy of various QMC WF will be tested in the prototypical bonding water dimer (Figure 5.1b). In particular, the role of the mathematical form of the QMC WF and the selected basis sets to describe both Jastrow and determinantal parts of the WF will be discussed.

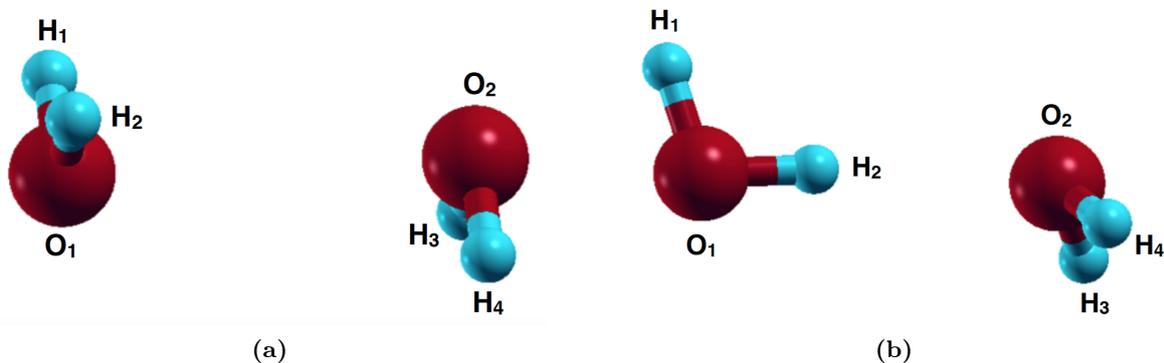


Figure 5.1 – Different configurations of the water dimer $(\text{H}_2\text{O})_2$. (a) non-bonding water dimer. (b) Bonding water dimer.

5.1 Non-bonding water dimer

In the non-bonding water dimer (Figure 5.1a), there is no significant stabilization of the cluster due to the presence of chemical (covalent or H-) bonds between the two water molecules of the dimer. Therefore, only Van der Waals (VdW) and dispersion interactions remain, making the non-bonding water dimer an interesting system to benchmark the ability of various electronic structure methods to properly capture such effects. As already mentioned, we need to verify that the QMC approach is able to describe properly the VdW interactions in water, in the perspective of simulating larger waters clusters, or even the bulk liquid.

To that purpose, we perform a series of structural relaxations of the non-bonding water dimer at various oxygen-oxygen distances \overline{OO} for various computational methods. Being aware that, by essence, this configuration is not the more stable one of the $(\text{H}_2\text{O})_2$ complex, we need to apply to the system some constraints during the structural optimization. At a given oxygen-oxygen distance \overline{OO} , we imposed the position of the 4 hydrogen atoms to be symmetric with respect to the inversion point of the water dimer, defined by the midpoint of the \overline{OO} distance. This symmetry choice is not unique, as Lin and coworkers studied the non-bonding water dimer VdW properties by imposing the $\angle \text{OHO}$ angles to be equal to 180° and calculating electronic energies at various oxygen-oxygen distances \overline{OO} ^[237].

In Figure 5.2, we represent the binding energy E_{bind} as a function of the oxygen-oxygen distance \overline{OO} . The water dimer binding energy, a key quantity in this Chapter, is defined as

$$E_{\text{bind}} = E_{\text{dimer}} - 2E_{\text{monomer}}, \quad (5.1)$$

where E_{dimer} and E_{monomer} are the water dimer and monomer energies, respectively. The LDA binding energy, represented by the black solid line, confirms that this functional provides a very severe overstructuring of the water dimer, as already discussed in the first Chapter. Even worse, this functional gives a very poor position of the non-bonding dimer minimum $\overline{OO}_{\text{LDA}} \simeq 3.2 \text{ \AA}$ more than 0.5 \AA shorter than the converged reference CCSD(T) value $\overline{OO}_{\text{CCSD(T)}} = 3.8 \text{ \AA}$ (brown curve). The PBE result, given by the dark green plot, is in remarkable agreement with the reference CCSD(T) calculation whereas the *a priori* more

Chapter 5. Interlude: the water dimer

sophisticated B3LYP functional underestimates the binding energy by a few (~ 0.2) kcal/mol. We mention that we verified the convergence of our DFT calculations with the basis set size by performing the same analysis using a quadruple zeta (QZ) basis set. The results, not shown here, superimpose with the curves obtained with the triple zeta (TZ) basis set, still used as our reference basis set in this thesis.

We want to investigate here the ability of the QMC approach to capture the VdW and dispersion interactions in the water dimer, disregarding the possible deficiencies arising from an eventual simplification of the WF. Consequently, we performed several VMC calculations using either a JSD WF expanded on a primitive basis set, which is composed of:

- 9 (6+3) basis functions for the Jastrow part, being: $3s, 2p, 1d$ for the oxygens and $2s, 1p$ for the hydrogens;
- 18 (12+6) basis functions for the determinantal part, being: $5s, 5p, 2d$ for the oxygens and $4s, 2p$ for the hydrogens.

We point out that this primitive basis set is exactly the same one as we employed to build the JAGP WF describing the Zundel ion in the two preceding Chapters. As in Chapter 3, we also applied the Geminal embedding scheme^[202] on the determinantal part of the QMC WF to reduce its number p of parameters. According to the Ref. 72, the best compromise between accuracy and computational cost is found to use 8 hybrid orbitals for the oxygen atoms¹ and only 2 for the hydrogen atoms. Looking at Figure 5.2, the agreement of the VMC binding energies computed on a primitive JSD WF² (red points) and the CCSD(T) reference values is excellent. As expected, the JSD WF with hybrid orbitals in its Slater determinant is slightly less accurate, with binding energies that are underestimated by about $\sim 0.1 - 0.2$ kcal/mol (blue points), within the error bar. Even though this QMC WF is less sophisticated than the primitive one, it is already as accurate as that of the hybrid functionals, such as B3LYP (magenta). This last remark further proves the ability of the QMC approach, via the use of a Jastrow term in the QMC WF (Eq. (1.36)), to properly take account for VdW contributions, at variance with most of the DFT functionals.

Once again, we have verified the robustness of the QMC approach, that is able to correctly describe the non-bonding water dimer PES, at any oxygen-oxygen distance. A very good agreement with the CC calculations is found, at variance with most functionals (LDA and B3LYP). We also stress that the PBE functional works remarkably well for the description of the non-bonding water dimer energetics. Although interesting as test case, the non-bonding dimer is not the most representative configuration of the $(\text{H}_2\text{O})_2$ complex in the bulk water or in aqueous systems. Indeed, due to the presence of a well-structured H-bond network, the bonding water dimer is a much more frequent configuration. Its properties will be investigated in the following Section.

¹instead of 12 primitive gaussian type orbitals (GTOs), as described a few lines above.

²for the sake of simplicity, we name 'primitive' a WF whose orbitals are expanded over a primitive basis set.

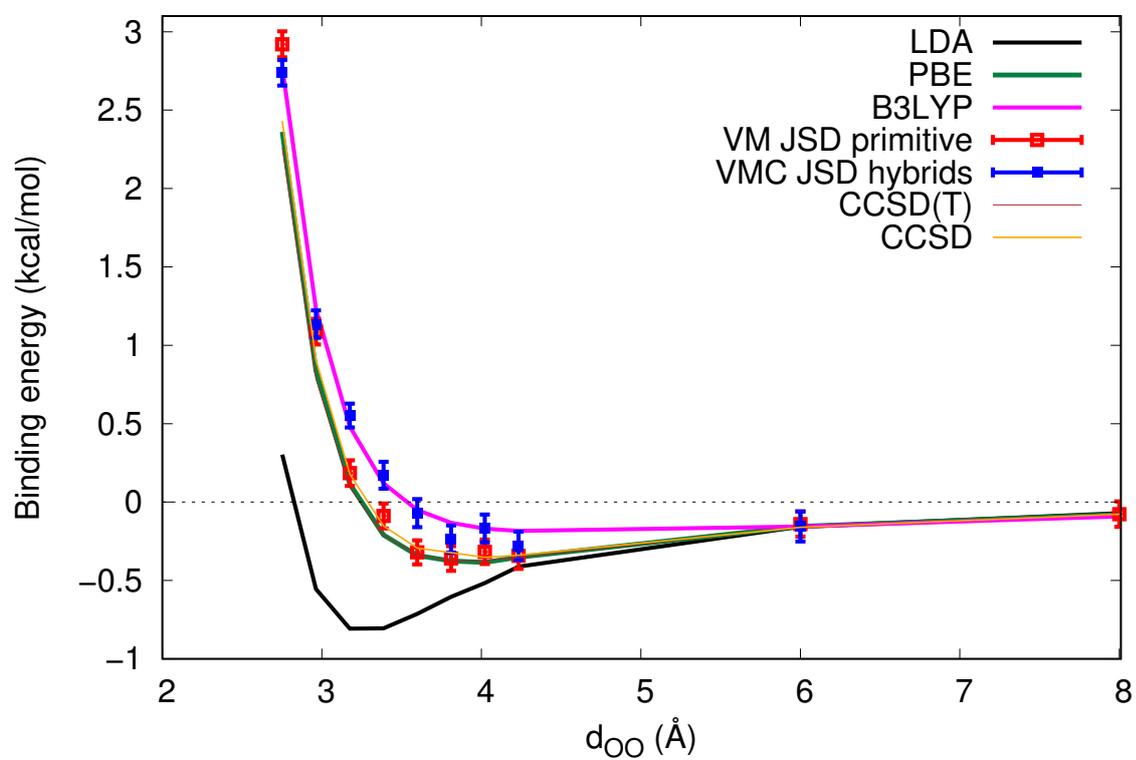


Figure 5.2 – Binding energy (kcal/mol) of the non-bonding water dimer (H_2O)₂ as a function of the OO distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory.

5.2 Bonding water dimer

In this Section, we focus our attention on the bonding water dimer properties (Figure 5.1b). As discussed in Chapter 1, the binding properties of this prototypical system have already been studied by means of QMC methods^[88,89]. Indeed, Sterpone *et al.* first proved the ability of the QMC approach, and more specifically its LRDMC variant, to properly reproduce the dissociation energy curve of the bonding $(\text{H}_2\text{O})_2$ complex. They found the LRDMC binding energy to be close to the experimental value, albeit slightly underestimated (by about ~ 0.5 kcal/mol)^[88]. Zen and his colleagues have then proposed a complete and systematic convergence study of the bonding water dimer properties as a function of the chosen ansatz and basis sets to describe the QMC WF. In the following, our goal is not to reproduce such a detailed analysis but we aim at testing some QMC WF that could be employed for the study of other water clusters and aqueous systems, such as the protonated hexamer.

5.2.1 Binding energy

We first start with the exploration of the bonding water dimer PES as a function of the oxygen-oxygen distance \overline{OO} . The results of our structural optimizations are presented in Figure 5.3, where different functionals are tested within the DFT framework (left panel). The QMC dissociation energy curves, obtained by both VMC and LRDMC, are plotted in the right panel for two different WFs.

Let us start with the discussion of the DFT results. As we already noticed in the case of the non-bonding water dimer, we confirm that the LDA functional (black) significantly overestimates the water dimer binding energy. We found $E_{\text{bind,LDA}} = 8.74$ kcal/mol, in correct agreement with $E_{\text{bind,LDA}} = 9.02$ kcal/mol, predicted in Ref. 43. The observed minor discrepancy can be explained by the coarse oxygen-oxygen distance grid used in the dissociation energy calculations (left panel of Figure 5.3). The GGA and the hybrid functionals such as PBE (dark green) and B3LYP (magenta) provide much more accurate dissociation curves of the bonding water dimer, the results being significantly closer to the reference CCSD(T) calculations. To be more quantitative, we found $E_{\text{bind,PBE}} = 5.28$ kcal/mol and $E_{\text{bind,B3LYP}} = 4.66$ kcal/mol, in very close agreement with the tabulated values in literature (5.2 and 4.57 kcal/mol, respectively^[43]). The predicted CCSD(T) binding energy is slightly too large (5.2 against 5.02 kcal/mol in Ref. 43). This can be attributed to the fact that we took the B3LYP geometries for our CCSD(T) energy calculations without relaxing the geometry at the same level of theory.

In the right panel of Figure 5.3, we report the binding dimer PES as a function of the oxygen-oxygen distance \overline{OO} , computed using the two QMC WFs tested in the previous Section, by both VMC and LRDMC. We notice that the QMC approach is able to properly reproduce the dissociation curve of the binding water dimer, giving correct values of the binding energy at any oxygen-oxygen distance \overline{OO} . As expected, the LRDMC result is even more accurate than its VMC counterpart, with an overall agreement within ~ 0.3 kcal/mol with the reference CCSD(T) curve. A more accurate and detailed study of the QMC binding energies will be presented in Table 5.1.

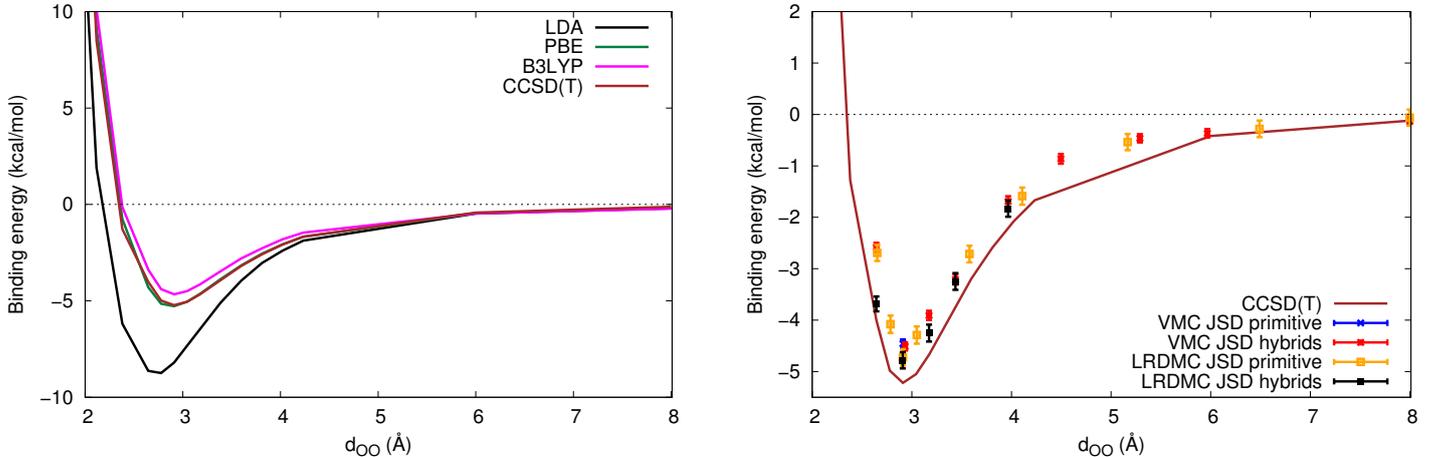


Figure 5.3 – Binding energy (kcal/mol) of the bonding water dimer (H_2O)₂ as a function of the \overline{OO} distance. Comparison between different computational methods. Left panel: DFT results. Right panel: QMC results. Structural relaxation is performed at each level of theory, except for the CCSD(T) and the LRDMC approaches. The LRDMC lattice parameter is set to $a = 0.125$.

In the perspective of simulating larger water clusters such as the protonated water hexamer in Chapter 6, we need to tackle the problem of the number of variational parameters p defining the QMC WF. Indeed, as discussed in Chapter 3, the computational time spent for a single QMC evaluation of the electronic energy is roughly proportional to p . Therefore, we plan to apply also the Geminal embedding scheme to the orbitals in the Jastrow factor. To the best of our knowledge, this is the very first time that such a strategy is employed for the Jastrow part of the QMC WF in water. We applied 3 distinct contractions schemes, referred as $\text{O}[m]\text{H}[n]$, with the integers m and n defining the number of employed hybrid orbitals for each atom. For instance, the $\text{O}[4]\text{H}[1]$ WF is the most simplified one, since it contains only 4 hybrid orbitals defining the oxygen atom in the Jastrow, whereas a single orbital is used for the hydrogen atom. As one can see in the second column of Table 5.1, the Geminal embedding scheme, applied to both the Jastrow and the determinantal part of the WF, drastically reduces the number of parameters p describing the QMC WF. Indeed, only 1283 parameters are required for the $\text{O}[6]\text{H}[2]$ QMC WF, against 6303 for the fully primitive WF (both in the Jastrow and the determinantal parts). Therefore, the statistical gain on the subsequent QMC calculations is at least of a factor 5 and may be even more important in larger systems such as the protonated water hexamer.

We need however to ensure that this increase of computational efficiency of the QMC calculations is not obtained at the price of a significant deterioration of the QMC WF, losing for instance a significant proportion of the QMC binding energy. To address that issue, we performed a series of structural relaxations of the bonding water dimer at the VMC level, using 3 different contraction schemes for the Jastrow orbitals, namely $\text{O}[4]\text{H}[1]$, $\text{O}[6]\text{H}[1]$ and $\text{O}[6]\text{H}[2]$. The corresponding results are represented in Figure 5.4, where the binding energies obtained for each QMC WF are given in the insert. We highlight that the computed binding energy the $\text{O}[4]\text{H}[1]$ and $\text{O}[6]\text{H}[1]$ hybrid WFs are affected by a significant basis set superposition

error (BSSE). Indeed, we applied a correction³ of 0.68 and 0.2 kcal/mol per water molecule to the O[4]H[1] and O[6]H[1] WFs, respectively. At variance with the two aforementioned hybrid WFs, the BSSE of the O[6]H[2] WF is reduced to 0.01 kcal/mol and can reasonably be neglected. As we can observe in Figure 5.4, the O[6]H[2] contraction scheme seems to be the most adapted to properly describe the overall PES of the bonding water dimer at a moderate computational cost. Indeed, a correct agreement (within ~ 0.3 kcal/mol) is found with the primitive JSD result.

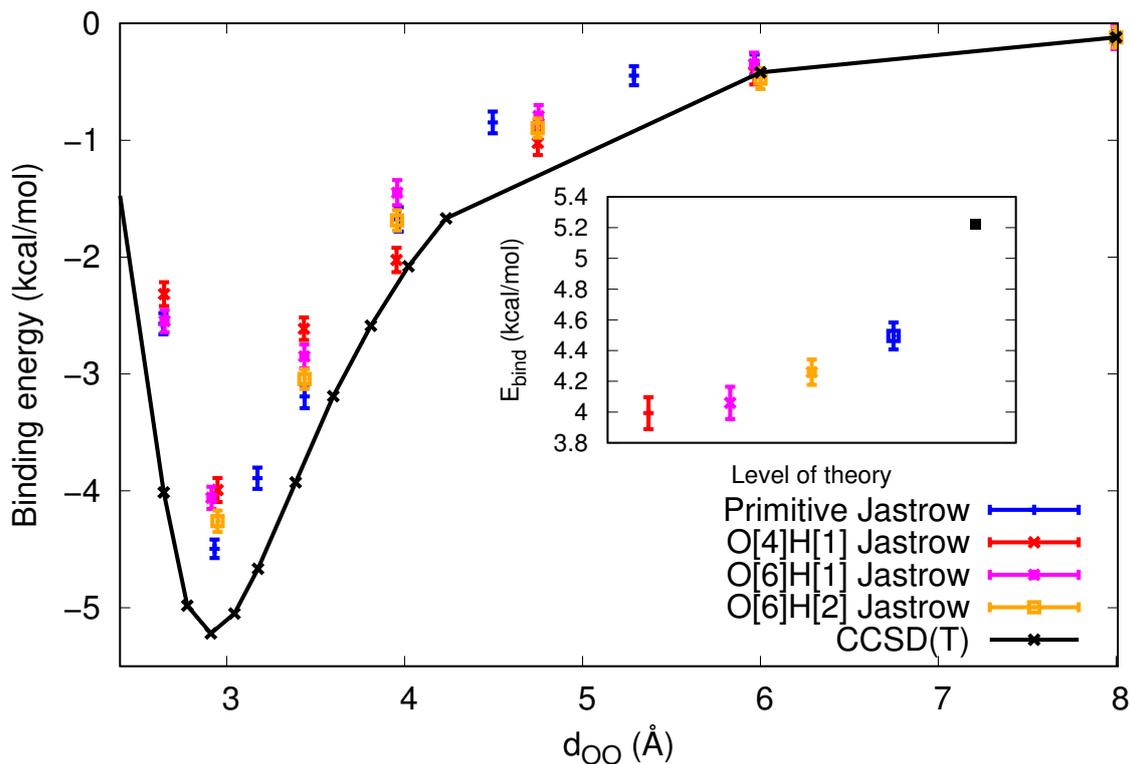


Figure 5.4 – Binding energy (kcal/mol) of the bonding water dimer (H_2O)₂ as a function of the \overline{OO} distance. Comparison between different Jastrow contraction schemes of the QMC WF. Structural relaxation is performed for each employed basis set.

To pursue our discussion on the binding properties of the water dimer, we report in Table 5.1 the values of the water dimer, monomer and binding energies for various QMC WFs. The number of parameters p describing the QMC WF is also given in the second column. We first emphasize that the value of the primitive binding energy is presumably not converged, since it represents roughly only 86% of the correct value. In the meantime, the simpler WF with hybrid orbitals in the Slater determinant (SD) recovers 85% of the reference CCSD(T) binding energy, as it is slightly less accurate than the primitive JSD. When the geminal embedding scheme is applied to both the Jastrow and the SD, the corresponding binding energy is found to be 4.05(8) kcal/mol, which is very reasonable ($\sim 78\%$ of accuracy), considering the

³The correction is defined as the energy difference between twice the energy of a single molecule and the water dimer calculated at a very large distance.

reduced number of parameters employed to describe the QMC WF. To go beyond the VMC approach and try to recover a more important fraction of the overall binding energy of the water dimer, we carried out some LRDMC calculations. In LRDMC, the lattice parameter $a = 0.125$ gives unbiased Fixed-Node (FN) energy differences⁴. We will employ this value for all LRDMC calculation carried out in this thesis. The LRDMC binding energy is estimated around 4.7 – 4.8 kcal/mol, in agreement with the value provided by Sterpone in Ref. 88. The LRDMC value is close to the reference CCSD(T), proving once again its remarkable accuracy.

We also tested another analytic form of the QMC WF, by taking an Antisymmetric Geminal Power matrix (AGP) instead of a SD to describe the fermionic part of the QMC WF. We find that the JAGP WF is, for a given basis set, less accurate than its JSD counterpart to describe the energetics of the bonding water dimer. Indeed, the JAGP binding energy is about ~ 0.7 kcal/mol lower than the one predicted with the JSD WF. We have however to investigate the quality of the water dimer geometric parameters, as it will be done in the following Subsection, before claiming that the JSD WF is more accurate than the JAGP one, or vice versa.

Last but not least, when one includes spin terms in the Jastrow factor, the binding energy is significantly improved (by about ~ 0.6 kcal/mol) for both JSD and JAGP WF. This significant improvement of the water dimer binding energy suggests that the spin-charge and spin-spin correlations are not so small in water, and should be taken into account. To the best of our knowledge, this is the first time that the possible influence of the spin on the binding interactions between water molecules is discussed so far. This remark is however preliminary and further analysis should be done to confirm the role of spin-spin correlations in water.

Method	p	E_{dimer} (Ha)	E_{monomer} (Ha)	E_{bind} (kcal/mol)
VMC JSD primitive	6303	-34.50372(8)	-17.24823(4)	4.46(8)
VMC JSD hybrids det	2089	-34.50337(9)	-17.24818(5)	4.40(8)
VMC JSD hybrids Jas + det	1283	-34.49876(8)	-17.24615(5)	4.05(8)
VMC JSD hybrids det + spin	5713	-34.51471(9)	-17.25390(4)	4.34(8)
VMC JSD hybrids Jas + det + spin	1953	-34.51069(7)	-17.25184(4)	4.40(7)
VMC JAGP hybrids Jas + det	1283	-34.50689(9)	-17.25089(7)	3.21(9)
VMC JAGP hybrids Jas + det + spin	1953	-34.51713(6)	-17.25551(5)	3.83(8)
LRDMC JSD primitive ($a = 0.25$)	6303	-34.5341(1)	-17.26360(8)	4.3(1)
LRDMC JSD primitive ($a = 0.125$)	6303	-34.5333(1)	-17.2629(1)	4.7(2)
LRDMC JSD hybrids det ($a = 0.125$)	2089	-34.5336(1)	-17.2630(1)	4.78(9)

Table 5.1 – Water dimer, water monomer and binding energies for various QMC WF. The number of parameters p of the WF employed for the calculations is also indicated.

To summarize, we have demonstrated that the QMC binding energies of both the non-bonding and the bonding water dimer are in good agreement with reference CCSD(T) calculations, albeit they are slightly underestimated. Indeed, describing such binding energies at the QMC level is a complex task since a subtle compromise must be found between extreme accuracy

⁴The corresponding simulation is more time-consuming (by a factor ~ 2) than if one sets $a = 0.25$, which leads to partially converged results.

and computational efficiency. Thanks to an original application of the Geminal embedding scheme to the Jastrow orbitals, we have devised a robust QMC WF that properly describes the water dimer energetics (80% of the CCSD(T) binding energy⁵), with a moderate number of parameters. This development allows us to use such a WF to study larger water clusters, such as the protonated water hexamer (Chapter 6). However, we have to ensure that this encouraging result is accompanied by an accurate description of the $(\text{H}_2\text{O})_2$ equilibrium geometry.

5.2.2 Geometric properties

In this Subsection, we provide a detailed analysis of the bonding water dimer geometric parameters of its minimum energy configuration for various QMC WFs. The results are reported in Tables 5.2 and 5.3, where DFT and reference CCSDTQ^[238] values are also indicated for comparison. A particular attention will be dedicated to the two most sensitive geometric parameters to properly describe the bonding water dimer minimum, namely: the oxygen-oxygen distance $\overline{O_1O_2}$, a key parameter for the proton localization in any water cluster (see Chapters 4 and 6), and the oxygen-(central) proton distance $\overline{O_1H_2}$. The correct description of the two aforementioned parameters is of paramount importance since it implies an accurate treatment of the electronic structure around the H- and covalent bonds in which the central proton is involved. For the sake of readability, these quantities are reported in bold characters.

By inspecting Table 5.2, we notice that the QMC approach predicts an accurate equilibrium oxygen-oxygen distance $\overline{O_1O_2}$ which is found to be 0.005 Å larger for the JSD WF, with respect to the reference CCSDTQ geometry^[238]. Looking at this observable, the JAGP WF is much less accurate than its JSD counterpart, as the $\overline{O_1O_2}$ parameter is overestimated by ~ 0.04 Å. This is also in relation with a weaker binding energy (Table 5.1). On the other hand, the QMC values of the $\overline{O_1H_2}$ distances are all compatible with the reference CCSDTQ result, within a 0.004 Å error bar, at variance with DFT results which predict too large oxygen-(central) proton distance. We notice that the geminal embedding scheme applied to the orbitals describing the Jastrow factor does not significantly spoil the overall description of the water dimer geometry. We also highlight that the inclusion of spin terms into the Jastrow factor improves the results obtained by the JAGP WF, as the oxygen-oxygen distance shortens, and the oxygen-(central) proton gets very close to the CCSDTQ value. This improvement, is much less noticeable in the case of the JSD WF, suggesting that only the JAGP WF should be employed to capture spin-spin correlations in the water dimer, and thus in the bulk liquid.

Looking at Table 5.3, the $\angle \text{H}_1\text{O}_1\text{H}_2$ and $\angle \text{H}_3\text{O}_2\text{H}_4$ are very accurately described by the QMC approach, as their values remain very close (within a 0.2° error) to the reference CCSDTQ values^[238] for any QMC WF. The inclusion of spin terms in the Jastrow factor does not significantly modify the intramolecular angles, even though the $\angle \text{H}_3\text{O}_2\text{H}_4$ is slightly better described in this case. We are however convinced that the spin-spin correlations in water are not so small, looking at the obtained water dimer binding energies when the spin is taken

⁵The CCSD(T) and VMC binding energies are in agreement within 0.5 kcal/mol, less than $1 k_B T$ at $T = 300$ K.

5.2. Bonding water dimer

Theory	$\overline{O_1O_2}$	$\overline{O_1H_1}$	$\overline{O_1H_2}$	$\overline{O_2H_3}$	$\overline{O_2H_4}$
PBE-DFT	2.9105*	0.9657	0.9772	0.9682	0.9682
B3LYP-DFT	2.9105*	0.9596	0.9685	0.9616	0.9616
QMC JSD primitive	2.9105*	0.9536(2)	0.9600(2)	0.9561(2)	0.9547(2)
QMC JSD hybrids det	2.9165(4)	0.9539(2)	0.9614(2)	0.9553(2)	0.9557(2)
QMC JSD hybrids Jas + det	2.9232(4)	0.9537(2)	0.9602(2)	0.9553(2)	0.9549(2)
QMC JSD hybrids det + spin	2.9197(4)	0.9543(2)	0.9622(2)	0.9563(2)	0.9556(2)
QMC JSD hybrids Jas + det + spin	2.9271(4)	0.9539(2)	0.9616(2)	0.9550(2)	0.9558(2)
QMC JAGP hybrids Jas + det	2.9600(2)	0.9535(2)	0.9606(2)	0.9557(2)	0.9558(2)
QMC JAGP hybrids Jas + det + spin	2.9535(4)	0.9543(2)	0.9626(2)	0.9563(2)	0.9562(2)
CCSDTQ ^[238]	2.90916	0.95685	0.96414	0.95843	0.95843

Table 5.2 – Geometric properties (distances in Å) of the bonding water dimer minimum, comparison between different computational methods. The * indicates that we kept the distance $\overline{O_1O_2}$ constant during the structural optimization.

Theory	$\angle H_1O_1H_2$	$\angle H_3O_2H_4$
PBE-DFT	105.148	105.037
B3LYP-DFT	105.807	105.649
QMC JSD primitive	105.145(5)	105.006(5)
QMC JSD hybrids det	105.101(5)	104.968(5)
QMC JSD hybrids Jas + det	104.979(5)	105.015(5)
QMC JSD hybrids det + spin	104.973(5)	104.950(5)
QMC JSD hybrids Jas + det + spin	105.004(5)	104.865(5)
QMC JAGP hybrids Jas + det	104.694(5)	104.729(5)
QMC JAGP hybrids Jas + det + spin	104.509(5)	104.810(5)
CCSDTQ ^[238]	104.854	104.945

Table 5.3 – Geometric properties (angles in °) of the bonding water dimer minimum, comparison between different computational methods.

Chapter 5. Interlude: the water dimer

into account in the Jastrow factor. We initiated the analysis of other molecular properties, such as the charge density, the dipole moment and their fluctuations with or without spin correlations in the QMC WF. They suggest that the spin-correlations tend to increase the water dimer dipole, by increasing the partial charges of each oxygen atoms but this analysis remains very qualitative and at a too early stage to provide clear conclusions.

To conclude, we have demonstrated throughout this Chapter, the ability of the QMC method to accurately reproduce both the energetics and the geometry of the bonding water dimer, at a reasonable computational cost. Furthermore, the preliminary study of the non-bonding water dimer also proved that the QMC approach is able to capture subtle VdW effects. By applying a novel embedding scheme to the orbitals filling the Jastrow factor, we also provided a useful simplification of the QMC WF, without significantly deteriorating its overall accuracy, as the water dimer bonding energy and its geometric parameters are correctly reproduced. This technical development will allow us to use such a WF to perform larger scale calculations to study more realistic problems, starting with the protonated water hexamer that will be studied in the next Chapter. Finally, we stress that despite the remarkable accuracy of the JSD WF to describe the water dimer properties, we keep using the JAGP WF to study charged water clusters (as we did for the Zundel ion in Chapter 4). Indeed, this WF ansatz allows larger charge fluctuations in the system, and is therefore more appropriate to describe the creation of charge defects in water clusters such as the protonated hexamer.

The protonated water hexamer

Contents

6.1	Zero temperature results	125
6.1.1	Potential energy landscape	125
6.1.2	Equilibrium geometries	129
6.1.3	Proton transfer static barriers	133
6.2	Classical hexamer dynamics	135
6.2.1	Radial distribution functions	136
6.2.2	Proton displacement	139
6.3	Quantum hexamer dynamics	144
6.3.1	Radial distribution functions	144
6.3.2	Proton displacement	147
6.3.3	Thermal effects versus Nuclear Quantum Effects	154
6.4	Proton transfer in the protonated hexamer	157
6.4.1	Classical counting	157
6.4.2	Quantum counting: the instanton theory	160

FOR MORE THAN 200 YEARS and the seminal work of von Grothaus^[8], the correct molecular description of the properties of the *aqueous* or *hydrated* proton $\text{H}_{(\text{aq})}^+$ has intrigued the scientific community^[108,239]. Despite significant advances, the exact role of the solvated proton in proton transfer (PT) reactions in chemical and biological systems is not fully elucidated yet. The common picture is that the hydrated proton exists as the classical hydronium cation H_3O^+ but it looks more appropriately described as a delocalized electronic charge defect between multiple molecules. The spread of this charge defect blurs the identity of the excess proton among two limiting structures, namely the Zundel^[10] and the Eigen^[9] ions. Indeed, the hydrated proton infrared (IR) spectrum displays a combination of few discrete absorption bands on top of a continuous broad absorption across the entire IR spectrum. Neither the symmetrically solvated hydronium ion H_9O_4^+ (Eigen), nor the equally shared proton in the H_5O_2^+ (Zundel) ion models involving fast interconversions between these two configurations^[240] can individually rationalize this characteristic IR fingerprint.

To deal with this issue, Stoyanov *et al.*^[241] have introduced the stable $\text{H}^+(\text{H}_2\text{O})_6$ (or equivalently $\text{H}_{13}\text{O}_6^+$) species, which is *Zundel-type* in the sense that the excess proton is equally shared between two water molecules. There is however more charge delocalization and, consequently, the core of the cluster is characterized by an unusually long central oxygen-oxygen distance. On the other hand, recent Molecular Dynamics (MD) simulations suggest the existence of a distorted, nonsymmetric *Eigen-type* cation, remaining at the heart of a dynamical electronic charge defect spanning multiple water molecules^[242]. This justifies our choice to

study the protonated water hexamer during this thesis, since it represents the smallest protonated water cluster for which both of these characteristic binding motifs coexist^[22,23]. The two main protonated hexamer configurations are represented in the left and right panels of Figure 6.1 for the Zundel-like and the Eigen-like forms, respectively.

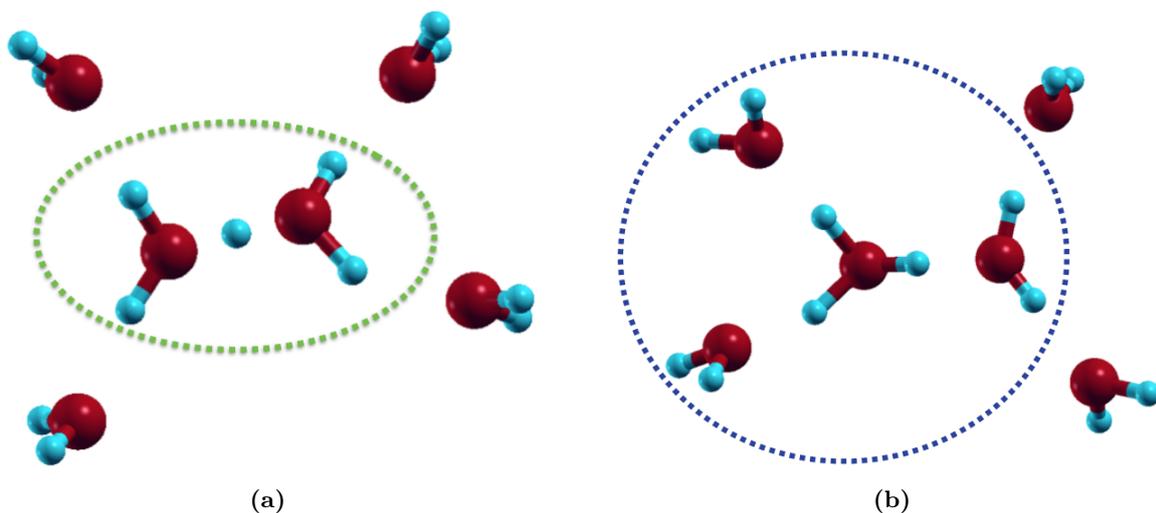


Figure 6.1 – Different configurations of the protonated water hexamer $\text{H}_{13}\text{O}_6^+$. (a) Zundel-like configuration with a Zundel center (H_5O_2^+ , green dashed line) and its first solvation shell (4 H_2O). (b) Eigen configuration with an Eigen cation (H_9O_4^+ , blue dashed line) on the left accompanied by two solvating water molecules (2 H_2O) on the right.

The protonated hexamer Potential Energy Surface (PES) has been partially explored by IR spectroscopic measurements^[216,243,244] and electronic structure calculations performed within the DFT and the MP2 approaches^[22,216,243,245], confirming that the two structures presented above are the lowest energy isomers. Nevertheless, to the best of our knowledge, there is no extensive study of the proton dynamics in the protonated hexamer at finite temperature. Furthermore, the rare theoretical investigations attempting to model the PT in water clusters are mostly performed with classical particles evolving in a PES evaluated at the DFT level^[242,246].

Thanks to the tools described in the previous Chapters, we are now able to perform the very first fully quantum dynamics of the protonated hexamer, with almost exact ionic forces computed within a QMC framework. In this Chapter, we hope to shed some light on the fundamental mechanisms that drive the dynamical behavior of the hydrated proton in water clusters. It is organized as follows. First, we study the zero temperature properties of the protonated water hexamer in Section 6.1. In particular, several electronic structure methods are compared for the description of the PES (Subsection 6.1.1) and the geometry of the system (Subsection 6.1.2). Their ability to predict the correct static barriers for proton transfer (PT) are also discussed in Subsection 6.1.3. Afterwards, we present the results obtained for the dynamics of the $\text{H}_{13}\text{O}_6^+$ ion with classical particles in Section 6.2. They are later compared with those obtained in the quantum case (Section 6.3), where the competition between thermal and Nuclear Quantum effects (NQE) is investigated (Subsection 6.3.3). In the last part of the Chapter, we attempt to quantify the proton jumps occurring during the protonated

water hexamer dynamics in both classical and quantum cases (Section 6.4).

6.1 Zero temperature results

Analogously to the case of the Zundel ion (see Chapter 4), exploring the zero temperature properties of the protonated water hexamer is of paramount importance. This constitutes a preliminary step to predict the possible proton transfer scenarios which may occur at finite temperature. Indeed, the physics of PT is certainly richer in the $\text{H}_{13}\text{O}_6^+$ ion than in the Zundel cation. As we can see in Figure 6.2, this is due to the presence of a solvation shell that interacts with a Zundel-like core, and *a priori* modifies its preferred configurations.

We will explore the impact of the solvation on the static properties (PES, equilibrium geometries and proton transfer static barriers) of the protonated water hexamer, using various computational methods to describe its electronic structure. The results will be compared to those obtained in Chapter 4 for the Zundel ion to accurately evaluate the impact of the solvation on the equilibrium properties of small water clusters.

6.1.1 Potential energy landscape

In order to explore the protonated water hexamer PES, we compute the $\text{H}_{13}\text{O}_6^+$ electronic energy as a function the central inter-oxygen distance $d_{\text{O}_1\text{O}_2}$ (atom labels are given in Figure 6.2) similarly to the calculations carried out on the Zundel ion. The results are plotted in Figure 6.3, where different electronic structure methods are compared together. The solid curves are obtained using DFT (dark green for PBE and magenta for DF2), while the QMC results are respectively represented by blue triangles (VMC) and black squares (LRDMC). Finally, the quantum chemistry calculations, performed using the triple zeta (TZ) basis set, are represented by red triangles (MP2), brown squares (CCSD) and orange circles (CCSD(T)). Except for CCSD/CCSD(T) calculations, the structural relaxations to find the equilibrium geometries are performed at each level of theory. The Coupled-Cluster calculations require much memory (more than 7 Gigabytes per core), which make them quite complex to perform. As a consequence, we decided to compute the CCSD/CCSD(T) energies using the MP2 geometries which, as we will see in the next Subsection, seem to provide an accurate description of the system. Let us emphasize that carrying out structural relaxations using CC techniques is already extremely complex, even for moderate size clusters such as $\text{H}_{13}\text{O}_6^+$. This again proves that, despite their remarkable accuracy, it is not possible to employ such calculations, at the present stage of development, to perform CC-driven MD simulations of extended systems such as the bulk liquid. It thus further motivates to use the QMC approach to deal with this issue because of its mild scaling with the system size.

The obtained results are qualitatively similar to those obtained for the Zundel complex. Indeed, the PBE functional tends to overestimate the slope of the PES by about 2 kcal/mol in the Coulombic repulsion regime, for short \overline{OO} distances. For larger values of $d_{\text{O}_1\text{O}_2}$, there is a good agreement with accurate quantum chemistry or QMC results. However, PBE predicts a slightly too large $d_{\text{O}_1\text{O}_2}$ distance (by about 0.05 Å) for the minimum geometry configuration, that could spoil the description of PT in the hexamer. The DF2 functional provides a less accurate description of the PES than PBE. Indeed, it underestimates by 2 kcal/mol

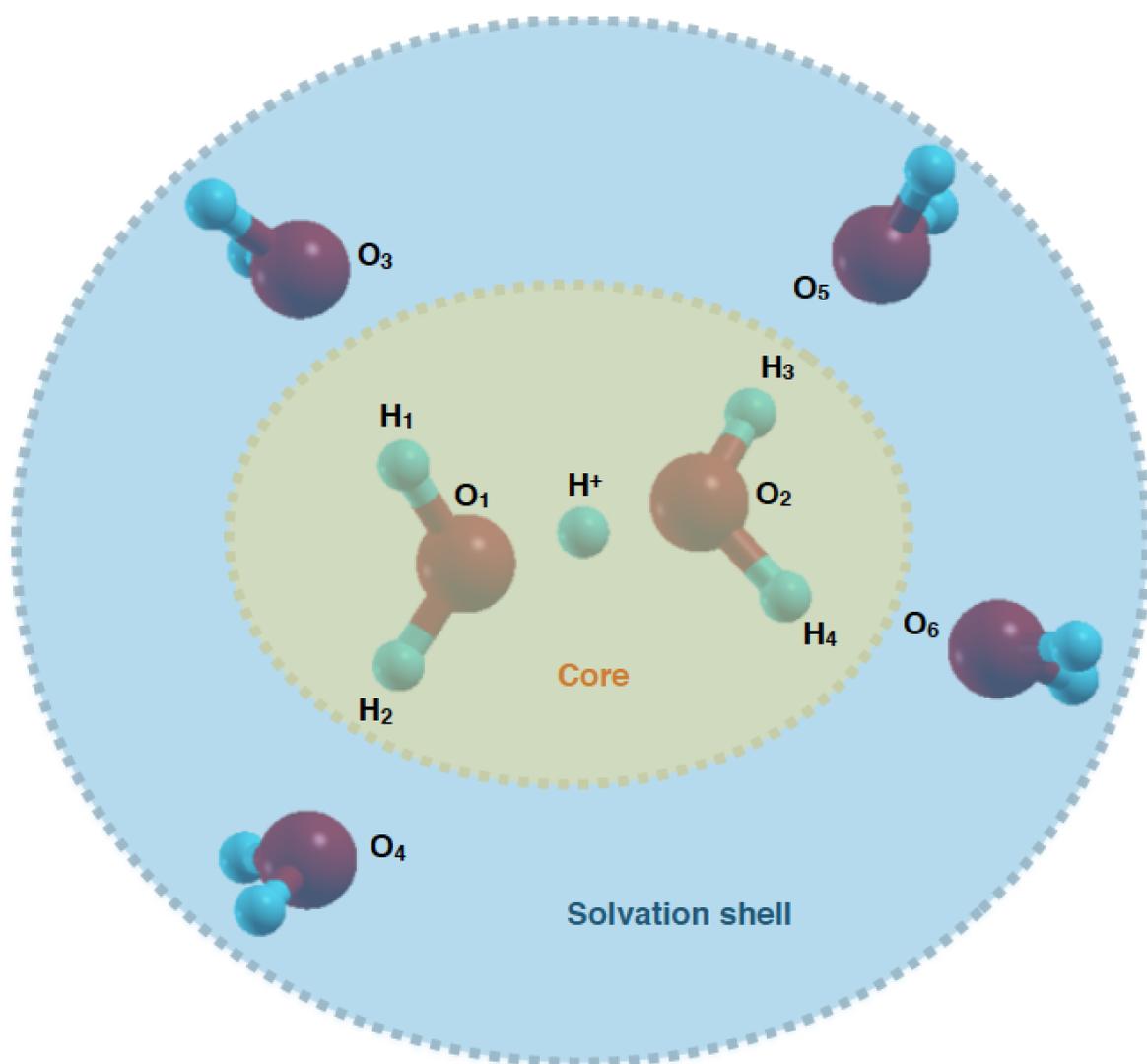


Figure 6.2 – Atom labels employed for the analysis of the protonated water hexamer properties. This cation is composed of a Zundel-like core with 4 surrounding water molecules in the solvation shell.

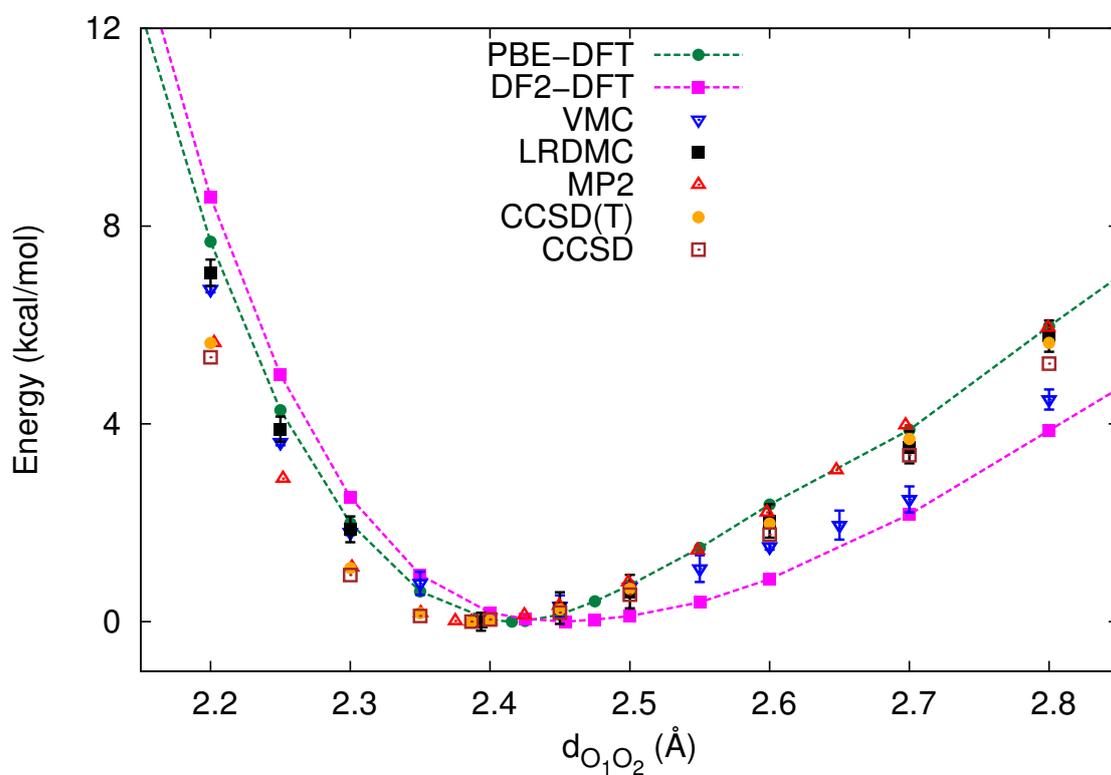


Figure 6.3 – Potential energy curve (kcal/mol) of the protonated water hexamer projected on the central \overline{OO} distance. Comparison between different computational methods. Structural relaxation is performed at each level of theory, except for CCSD/CCSD(T). Each curve has its minimum as reference point.

the electronic energy for large $d_{\text{O}_1\text{O}_2}$ values, presumably because it overestimates the cluster polarizability. Even worse, the whole PES obtained by DF2 is remarkably shifted to larger inter-oxygen distances, indicating that this functional might overestimate the \overline{OO} distances in larger aqueous systems such as bulk water. Consequently, the DF2 functional cannot provide the necessary accuracy to capture all the PT physics in the protonated water hexamer. The VMC results are, within the error bar, in good agreement with the reference CC calculations. Nevertheless, even though it is more accurate than the DF2 functional, the VMC approach also slightly overestimates the dispersion effects at large \overline{OO} distances. Furthermore, the LRDMC results, which are simply an improvement of the VMC ones since they are obtained with the same wave function (WF), are in perfect agreement with CCSD(T) results for the long-range part. However, for short \overline{OO} distances, the QMC (both VMC and LRDMC) energies are too large. This indicates that the orbitals of the QMC WF are probably too diffuse, which leads to an overestimation of the Coulomb repulsion between the central proton and its two surrounding water molecules.

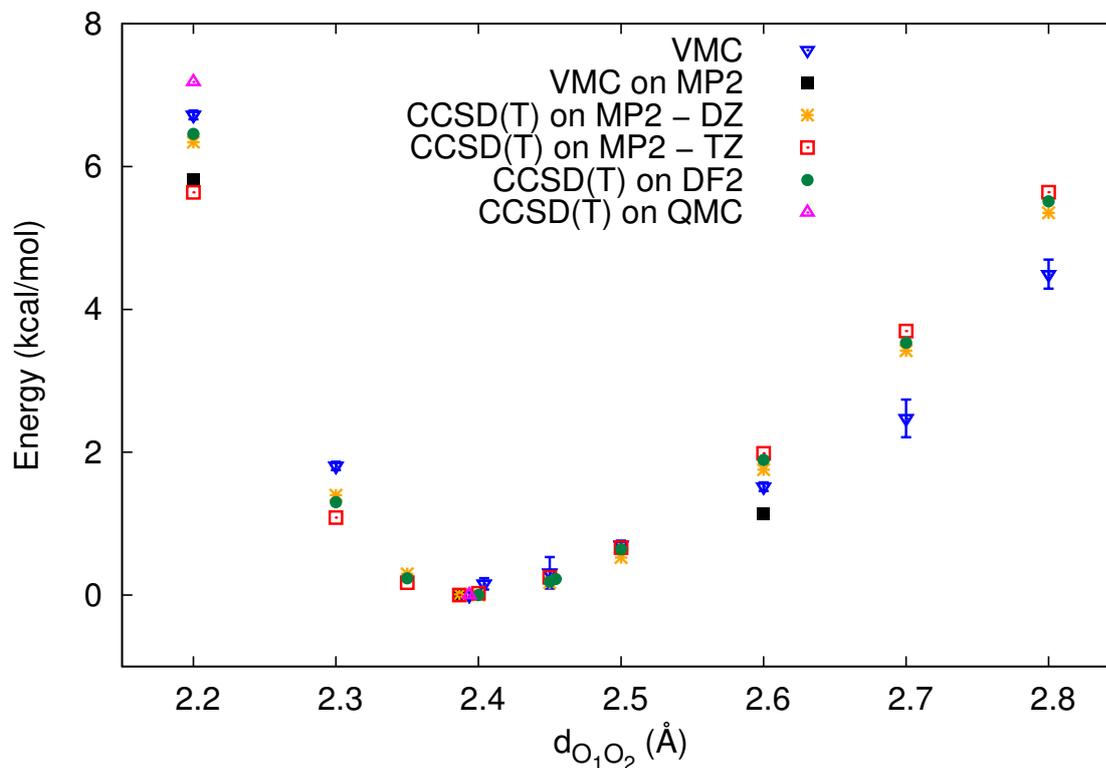


Figure 6.4 – Potential energy curve (kcal/mol) of the protonated water hexamer projected on the central \overline{OO} distance. Comparison between different computational point methods for various relaxed geometries and basis sets. Each curve has its minimum as reference point.

This last hypothesis should however be verified since, in such small systems, the electronic energy strongly depends on the geometry of the system. To that purpose, we decided to perform several calculations where we do not necessarily select the same level of theory for the structural relaxation and the electronic energy evaluation. The corresponding results are plotted in Figure 6.4, where CCSD(T) calculations on DF2 geometries are represented by

dark green circles, while the magenta triangles are the obtained results on QMC geometries and the red squares are the reference MP2 calculations. At $d_{O_1O_2} = 2.2 \text{ \AA}$, the obtained values differ each other by 1 kcal/mol, confirming the sensibility of such results with respect to the employed geometry. We also performed VMC calculations on MP2 geometries and the black squares are close to the red ones, indicating that a good agreement is recovered between MP2 and VMC energy profiles. Last but not least, we tested the influence of the basis set size (double zeta or triple zeta) for the CCSD(T) calculations. The obtained values with the double zeta (orange circles) basis set are in very good agreement with those computed using the TZ one, meaning that basis set size effects can be neglected as a source of discrepancy.

To summarize, we have demonstrated that the PBE and DF2 functionals do not accurately describe the protonated hexamer PES along the reaction coordinate \overline{OO} , the situation being worse for the DF2 functional. We have also verified the accuracy of the QMC approach for the protonated hexamer energetics. Indeed, the QMC energies are in good agreement with reference quantum chemistry calculations as soon as they are performed on the same geometries. The quality and the accuracy of the protonated hexamer geometries is discussed in the following Subsection.

6.1.2 Equilibrium geometries

We have established in Chapter 4 that an accurate description of the equilibrium geometries is essential to provide a good localization of the excess proton during its dynamics. In particular, PT processes are known to be very sensitive to the inter-oxygen distance between water molecules^[242]. Indeed, the central proton can hop, more or less frequently, from one water molecule to another, depending on the average distance \overline{OO} .

In the Figure 6.5, we report the separations between the excess proton H^+ and the two central oxygen atoms O_1 and O_2 , to directly compare the results with those obtained for the Zundel ion. The solid curves represent the DFT geometries obtained with the PBE (dark green) and DF2 (magenta) functionals, respectively. The QMC geometries, obtained at the VMC level, are represented by blue circles whereas the red triangles indicate the MP2 results. First of all, at variance with the Zundel cation, the protonated water hexamer minimum energy structure is *asymmetric*¹. This fundamental symmetry modification of the system is due to the presence of the 4 solvating molecules, which tend to stabilize the hexamer into its Eigen-like configuration (right panel of Figure 6.1). This result is in agreement with Ref. 247 that predicts an Eigen structure for the hydrated proton. The PBE functional however predicts a Zundel-like symmetric minimum, which is erroneous. More generally, this functional gives a poor description of the proton location when one stretches the \overline{OO} distance. Despite its reasonable accuracy for the protonated hexamer energetics, the PBE functional cannot be used to accurately probe PT processes in water clusters, and more generally in the bulk liquid. When one includes dispersion effects using the DF2 functional, the geometric properties of the hexamer are significantly improved, displaying a better agreement with the reference quan-

¹As we have seen in Chapter 4, an asymmetric structure is characterized by a noticeable difference between the $\overline{O_1H^+}$ and $\overline{H^+O_2}$ distances in the cluster, at variance with the C_2 -symmetric cluster where $\overline{O_1H^+} = \overline{H^+O_2}$.

6.1. Zero temperature results

tum chemistry calculations. Nevertheless, the predicted equilibrium oxygen-oxygen distance is too large, which leads to an overly asymmetric cluster. Therefore, we expect the DF2 PT static barriers to be inaccurate, which is problematic in the perspective of studying PT at finite temperature. By contrast with the DFT, the QMC geometries are in excellent agreement with those obtained via MP2, especially around $d_{\text{O}_1\text{O}_2} = 2.4 \text{ \AA}$. Indeed, it is crucial to accurately describe this critical region since it represents the frontier between the symmetric Zundel-like, and the asymmetric Eigen-like configurations discussed in this Chapter. Let us also note that Figure 6.5 is very similar to Figure 4.3, meaning that despite the presence of solvating molecules, the hierarchy between the different levels of theory is preserved. Indeed, we have seen in Chapter 1 that the positions of the oxygen-oxygen or oxygen-proton RDFs peaks in the bulk water are more accurate within a QMC framework than using MP2 or DFT.

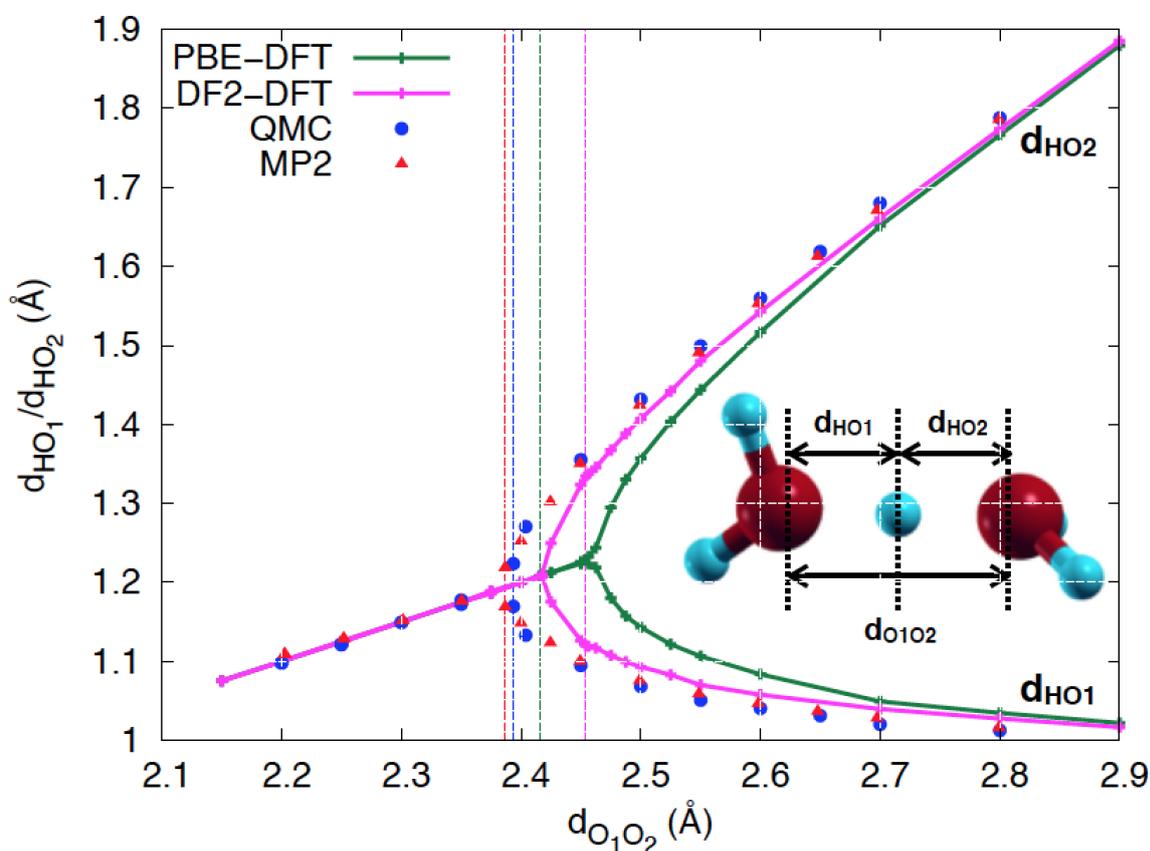


Figure 6.5 – Separations (\AA) between the two central oxygen atoms and the excess proton as a function of the reaction coordinate $d_{\text{O}_1\text{O}_2}$ for different computational methods. Vertical dashed lines indicate the equilibrium $d_{\text{O}_1\text{O}_2}$ for each method.

We provide in the following a more detailed analysis of the geometries, focusing first on the core (Tables 6.1 and 6.2) and later on the solvation shell (Table 6.3). Detailed QMC geometries are also given in the Appendix E (Sections E.1 and E.2 for the equilibrium geometry and other configurations, respectively). To further explore the accuracy of the DFT functionals, we also performed structural relaxations using the BLYP and B3LYP functionals. To help the

reader to visualize the overall symmetry of the protonated hexamer at each level of theory, we employ a color code. Red values refer to a fully symmetric cluster, with a Zundel core surrounded by a symmetric solvation shell (PBE functional). The blue data indicate a partial asymmetry of the cluster, composed of a Eigen-like core and an almost symmetric outer shell (BLYP and B3LYP functionals). Finally, when both the core and the solvation shell of the $\text{H}^+(\text{H}_2\text{O})_6$ ion are asymmetric, no coloration is employed (DF2 functional, MP2 and VMC). We emphasize that the lowest energy configuration of the protonated water hexamer is *asymmetric* according to the most accurate electronic structure theories (MP2; VMC and CCSD/CCSD(T)).

In Table 6.1, we report the equilibrium \overline{OH} distances for various levels of theory, where atoms labels are indicated in Figure 6.2. First of all, the PBE results strikingly differ from those obtained with other methods since the core of the system has a C_2 -symmetry, as in the Zundel complex. Indeed, the \overline{OH} intramolecular distances are perfectly equal in the left and right parts of the cluster, which is not the case for all the other theories. This errors does not come from the equilibrium $\overline{O_1O_2}$ distance since the B3LYP functional predicts a shorter $\overline{O_1O_2}$ value and displays an asymmetric geometry at the minimum. The B3LYP results are in rather good agreement with reference MP2 calculations, in spite of a 0.02 Å error on the equilibrium $\overline{O_1H_2}$ distance, which spoils the excess proton localization. The BLYP results compare with those obtained with the PBE functional but the cluster asymmetry at its minimum is well-recovered. As already mentioned, the predicted minimum by the DF2 functional is too asymmetric, due to an exceedingly large \overline{OO} distance. While this functional provides a quite accurate description of the global geometry of the cluster for $d_{OO} = 2.2 - 2.9$ Å, it should not be employed to probe the geometric properties around the zero temperature minimum. The VMC minimum geometry is in very good agreement with the one obtained in MP2, with an accurate description of the excess proton localization, as previously seen in Figure 6.5. In particular, the VMC and MP2 equilibrium oxygen-oxygen distances $\overline{O_1O_2}$ are in excellent agreement with the experimental value $\overline{O_1O_{2\text{exp}}} = 2.39 \pm 0.02$ Å measured by X-ray diffraction^[248]. The \overline{OH} intramolecular distances are however slightly too short in VMC, as already noticed in the case of the neutral and protonated water dimer (Chapters 4 and 5). This error is attributed to the reduced basis sets and the contraction schemes to simplify the QMC wave function and can be neglected because it does not affect the rest of the whole geometry.

Finally, let us note that, except for the \overline{OH} intramolecular distances that are slightly elongated due to the presence of the solvation shell, the geometric parameters are very close to those reported in Table 4.1. This suggests that, in spite of a subtle change of the cluster symmetry due to solvation effects, the core of the water hexamer can be considered as a *distorted* Zundel complex.

To go further with the geometry analysis, we report in Table 6.2 the obtained $\angle HOH$ angles for the 2 water molecules surrounding the excess proton in the Zundel-like complex. We report also the values of the $\angle O_1H^+O_2$ angle, that indicates whether the central proton belongs or not to the straight line defined by the O_1 and O_2 atoms. Looking at the $\angle HOH$ angles, the cluster asymmetry at its minimum energy configuration is less clear, except for

6.1. Zero temperature results

Theory	$\overline{O_1O_2}$	$\overline{O_1H^+}$	$\overline{H^+O_2}$	$\overline{O_1H_1}$	$\overline{O_1H_2}$	$\overline{O_2H_3}$	$\overline{O_2H_4}$
DFT-PBE	2.4156	1.2078	1.2078	0.9935	0.9935	0.9935	0.9935
DFT-DF2	2.4541	1.1196	1.3346	0.9913	0.9911	0.9797	0.9797
DFT-BLYP	2.4271	1.2061	1.2210	0.9982	0.9982	0.9973	0.9973
DFT-B3LYP	2.4025	1.1894	1.2131	0.9879	0.9879	0.9866	0.9866
VMC	2.3930(5)	1.1555(5)	1.2375(5)	0.9800(8)	0.9798(8)	0.9752(8)	0.9748(8)
MP2	2.3867	1.1690	1.2188	0.9877	0.9878	0.9847	0.9848

Table 6.1 – Geometric properties (distances in Å) of the core of the protonated water hexamer minimum, comparison between different computational methods.

the DF2 functional. These parameters seem to be less impacted by the presence of the surrounding water molecules, as they increase only slightly with respect to the values obtained in the Zundel case (Table 4.2). Still comparing to the Zundel cation, $\angle HOH$ angles are also generally (apart from the PBE functional) overestimated in the DFT approach while they are very well reproduced within the QMC framework. The situation for the $\angle O_1H^+O_2$ angle is completely different since its average value spectacularly changes when the Zundel complex is solvated. Indeed, except for the MP2 approach, the protonated hexamer core is found to be flat, since the central proton remains in the plane defined by its 2 neighboring oxygen atoms ($\angle O_1H^+O_2 \sim 180^\circ$). The protonated hexamer is certainly less flexible than the Zundel cation, due to the presence of more numerous H-bonds between the core and the solvation shell that strongly structure the cluster. Consequently, the 2 central H-bonds involving the central proton are more directional, preventing it to leave from the O_1O_2 straight line. The MP2 exception might come from an underestimation of the Coulomb repulsion between electrons, since the slope of the MP2 electronic PES is softer for short \overline{OO} distances (see Figure 6.3).

Theory	$\angle O_1H^+O_2$	$\angle H_1O_1H_2$	$\angle H_3O_2H_4$
DFT-PBE	180	109.6	109.6
DFT-DF2	179	110.7	109.3
DFT-BLYP	179.9	110.3	110.1
DFT-B3LYP	179.8	110.6	110.4
VMC	179.7(1)	109.2(2)	108.6(2)
MP2	176.6	109.4	109

Table 6.2 – Geometric properties (angles in $^\circ$) of the core of the protonated water hexamer minimum, comparison between different computational methods.

To conclude our study of the geometric properties of the protonated water hexamer, we report in Table 6.3 the \overline{OO} distances between the two core oxygen atoms and those belonging to the surrounding water molecules. We first note that the obtained distances are about 0.25 Å larger than the typical value of the oxygen-oxygen distance $\overline{O_1O_2}$ in the Zundel core. This implies that there is *a priori* no other possible PT scenarios that could occur inside the protonated hexamer. Indeed, one could think that the protons involved in the H-bonds between the side water molecules of the core might, in principle, jump to the lone pair of the neighboring oxygen atom of the solvation shell. In practice, this phenomenon is never observed here since the corresponding final state would be completely unstable. If one is eager

to multiply the possibilities of proton hopping between the Zundel-like core and its shell, an additional set of water molecules should be added to the cluster, to form another solvation shell stabilizing the edifice. The (erroneous) full-symmetry predicted by the PBE functional is confirmed here, where all the \overline{OO} distances are equal, and similar to the MP2 result. The DF2, BLYP and B3LYP functionals predict, as expected, the correct cluster asymmetry, with shorter \overline{OO} distance on the left part of the system with respect to the right one. Indeed, the central proton is slightly closer to the left oxygen atom O_1 , which must be compensated by the surrounding water molecules that move closer to the left part of the core. Let us also remark that for the BLYP and B3LYP functionals, $\overline{O_1O_3} = \overline{O_1O_4}$ and $\overline{O_2O_5} = \overline{O_2O_6}$, which is not the case for the DF2 functional, neither for the QMC and MP2 approaches. These functionals do not reproduce the exact chemical environment around the hydrated proton, so cannot be used to perform quantitative calculations to evaluate PT rates in water. The QMC approach still gives satisfactory results, in spite of a non-negligible overestimation by ~ 0.05 Å of the \overline{OO} distances. As already mentioned, it comes from the employed WF to carry out the structural relaxations at the QMC level. Fortunately, the $H_{13}O_6^+$ core is well-reproduced within the QMC framework, and no significant physics should be missed.

Theory	$\overline{O_1O_3}$	$\overline{O_1O_4}$	$\overline{O_2O_5}$	$\overline{O_2O_6}$
DFT-PBE	2.6594	2.6594	2.6594	2.6594
DFT-DF2	2.6765	2.6771	2.7487	2.7523
DFT-BLYP	2.6889	2.6890	2.6947	2.6949
DFT-B3LYP	2.6698	2.6696	2.6790	2.6792
VMC	2.6926(6)	2.6873(6)	2.7167(6)	2.7204(6)
MP2	2.6485	2.6465	2.6645	2.6667

Table 6.3 – Geometric properties (distances in Å) of the solvation shell of the protonated water hexamer minimum, comparison between different computational methods.

To put it in a nutshell, we have demonstrated the ability of the QMC method to properly capture subtle geometry changes for the protonated water hexamer with respect to the simpler Zundel case. Indeed, this cluster adopts a *distorted* Zundel configuration at its minimum, where the hydrated proton is no longer delocalized, but trapped by a static barrier. The latter represents the energy one should provide to the $H_{13}O_6^+$ ion to make the central proton freely hop from one side water molecule to another, as in the Zundel picture. The PT static barrier value thus strongly depends on the selected method to describe the protonated hexamer PES, as we will see in the next Subsection.

6.1.3 Proton transfer static barriers

In the perspective of probing the microscopic rearrangements triggering PT processes within the protonated water hexamer at finite temperature, it is necessary to fully determine the relationship between the zero temperature PES and the geometric properties of the system. In the first two Subsections of this Chapter, we have established that the protonated hexamer minimum energy configuration is *asymmetric*, where the hydrated proton is *not* equally shared between the 2 central water molecules. Consequently, its motion from one side water molecule to another is conditioned by the necessary energy it should acquire to go across the

6.1. Zero temperature results

PT static barrier. In practice, this energy is provided by the molecular vibrations that cause fluctuations of the covalent and H-bonds network. Thus, the height of the barrier gives us an indication about the temperature at which we expect to observe significant thermal effects on the hydrated proton dynamics, in both the classical and quantum cases.

To that purpose, we have performed a novel series of structural relaxations of the protonated hexamer, where we imposed the central proton to be situated exactly at the middle of the central oxygen-oxygen distance $\overline{O_1O_2}$. We have decided to work with $\overline{O_1O_2} = 2.45$ Å and 2.5 to go far beyond typical oxygen-oxygen distances observed in Zundel-type structures ($\overline{OO} = 2.39 - 2.42$ Å)^[248-250]. Indeed, there is the experimental evidence that the centrosymmetric $\text{H}_{13}\text{O}_6^+$ ion can be elongated ($\overline{OO} = 2.57$ Å)^[251], due to the environment. Looking at Figure 6.5, the variation of the cluster asymmetry as a function of the \overline{OO} distance is pronounced in the 2.45 – 2.5 Å region, indicating the trapping of the hydrated proton. We then evaluate the electronic energy of the symmetric-constrained structures obtained at the same level of theory, except for the CCSD/CCSD(T) calculations where we use the MP2 geometries.

We estimate the PT static barrier as the energy difference between the fully relaxed structure and the one with the symmetrized proton at the same \overline{OO} distance. The barriers are reported in Table 6.4 for each level of theory. Values are given in Kelvin, to directly show the minimal required temperature the cluster should have to exhibit spontaneous PT processes inside its core.

$d_{O_1O_2}$ (Å)	2.45	2.50
PBE-DFT	0	96
DF2-DFT	39	483
MP2	85	327
VMC	195 ± 25	562 ± 27
LRDMC	222 ± 56	389 ± 64
CCSD	211	592
CCSD(T)	141	431

Table 6.4 – Static symmetrization barriers (in Kelvin units) of the protonated water hexamer at different $d_{O_1O_2}$ for various computational methods. Results are computed as energy differences between the true minimum energy configuration and the symmetrized one.

First of all, let us point out that the PBE functional does not predict any barrier at $\overline{O_1O_2} = 2.45$ Å since the corresponding cluster equilibrium geometry remains fully symmetric. It thus leads to a severe underestimation of the PT static barrier for oxygen-oxygen distances close to the equilibrium geometry. This observation is consistent with the result obtained by Dagrada and coworkers who found a vanishing value of the PBE barrier for $\overline{O_1O_2} = 2.47$ Å in the Zundel complex^[72]. Therefore, we expect the PBE functional to overestimate the proton diffusion the bulk water, which has been recently confirmed in Ref. 16. The DF2 functional also underestimates the barrier by about 150 K (0.3 kcal/mol) at $\overline{O_1O_2} = 2.45$ Å by comparison with the CCSD(T) result, but the agreement is significantly improved at 2.5 Å. This functional is able to provide a quite accurate description of the PT barrier at large \overline{OO} distances. Unfortunately, the weight of the distorted Zundel configuration is cer-

tainly larger since it corresponds to the physical minimum of the cluster. The DF2 functional can however be used if one wants to obtain a not-so-rough estimation of the proton diffusion in aqueous systems at a moderate computational cost.

The biggest surprise comes from the MP2 result. This theory systematically underestimates the PT static barrier by about 100 K (0.2 kcal/mol) compared to the CCSD(T) predictions performed on the same geometries. In absolute, this energy difference is very small, but can be prohibitive in the perspective on reaching an accurate description of PT physics in water and aqueous solutions.

Last but not least, we notice that the static barriers predicted within the QMC framework (both VMC and LRDMC) are, within the error bar, in very good agreement with the reference CCSD(T) results, at least for $\overline{O_1O_2} = 2.45 \text{ \AA}$. Indeed, the VMC barrier is only 100 K (0.2 kcal/mol) larger than the CCSD(T) estimate when the oxygen-oxygen distance is stretched (2.5 Å). This slight overestimation by the VMC approach is consistent with previous observations for the Zundel cation, where the VMC PT barrier is 140 K at $\overline{O_1O_2} = 2.47 \text{ \AA}$, only 30 K larger than the CCSD(T) estimate^[72]. The agreement is even better between the LRDMC and the CCSD(T) results, but this method is unfortunately too computationally demanding to be incorporated into a MD framework.

At fixed \overline{OO} distance, the predicted barriers are larger in the $\text{H}_{13}\text{O}_6^+$ ion than in the Zundel cation. This is consistent with the idea that one should pay an additional price for the rearrangement of the molecules in the solvation shell during the PT process. Indeed, such a process creates a charge redistribution inside the cluster whose geometry is readjusted for stability.

To conclude, we have highlighted the weaknesses of the DFT approach to properly describe *both* the protonated water hexamer PES and geometries. We have also verified that the QMC approach, especially its VMC variant that we will use in the following, provides an accurate and reliable description of the $\text{H}_{13}\text{O}_6^+$ PES, the equilibrium geometries and static barriers, in agreement with the most advanced electronic structure methods. Therefore, using the tools developed in this thesis, we can confidently explore the finite temperature dynamics of the protonated hexamer.

6.2 Classical hexamer dynamics

In this Section, we investigate the impact of thermal effects on the protonated hexamer dynamics in the classical case. We are aware that we should take into account the quantum nature of the protons to obtain a realistic description of the PT processes occurring inside the cluster. Nevertheless, the analysis of the elementary mechanisms driving the proton diffusion in such systems is not trivial when NQE are taken into account, as they compete with thermal effects. Consequently, the analysis of the protonated hexamer behavior in the (simpler) classical case will provide precious indications about the exact role played by the temperature in the hydrated proton dynamics. This information will be later used to fully understand the results presented in the quantum Section.

6.2.1 Radial distribution functions

Firstly, we focus our attention on the equilibrium, or the static properties of the $\text{H}^+(\text{H}_2\text{O})_6$ ion. The latter are obtained by performing QMC-driven Langevin Dynamics using the Classical Momentum-Position Correlator (CMPC) algorithm introduced in Chapter 3. However, since the protonated hexamer is much larger than the Zundel ion, its WF is more complex and one cannot straightforwardly apply the same strategy employed to carry out CPMC-LD simulations of the Zundel cation. Indeed, the number of electronic WF parameters increases linearly with the system size. In practice, if one uses exactly the same JAGP WF employed to describe the Zundel ion in Chapter 4, 15763 parameters should be optimized between each LD step, against only 2391. Consequently, the part dedicated to the WF optimization, which is the most time-consuming, would be multiplied at least by a factor of 5. We have the further constraint that the generated MD trajectories should be long enough to display several spontaneous PT processes inside the cluster, implying that they should last at least 20-25 ps. Therefore, using the JAGP WF of Ref. 72 is not viable in a reasonable amount of computational time.

To solve that issue, we apply a Geminal embedding scheme^[202] to the Jastrow factor (this technique has already been used for the AGP part of the WF), using the O[6]H[2] contraction described in Chapter 5. In this way, a significant fraction of the water dimer binding energy is recovered, with a reasonable number of electronic parameters. Thanks to this strategy, we have reduced the number of electronic parameters to 6418, representing a significant gain. Furthermore, since the Hilbert space of the electronic parameters is larger than in the Zundel case, the WF optimization, using the SRH method, is less stable and one should reduce the acceleration of the electronic parameters $\delta t_{\text{par}} = 0.3$ (see Eq. (C.9)). We also increased the number of MC samples to $\simeq 4.0 \times 10^5$ to provide an accurate evaluation of the ionic forces and energy derivatives at each MD step. These choices allow reaching an accuracy of 3 mHa (1.9 kcal/mol) in the total energy per VMC energy minimization step. Thus, the statistical error on the Born-Oppenheimer (BO) surface sampling is slightly larger than the Zundel one, but is sufficient not to spoil the quality of the protonated hexamer dynamics.

To quantify the impact of thermal effects on the proton dynamics in the $\text{H}^+(\text{H}_2\text{O})_6$ cluster, we decided to generate 5 trajectories of 25 ps, propagated using the CPMC algorithm at various temperatures ranging from 200 to 400 K. This choice is motivated by the fact that very often PT reactions occur at room temperature, in chemical or biological systems. The Langevin damping γ_{BO} and the time step δt employed to propagate the CPMC-LD dynamics are the same that those employed for the study of the Zundel ion (Chapter 4). We verified that the average simulation temperature is compatible with the target temperature of the considered calculation, within a 1% error. This result ensures that the sampling of the BO surface remains correct during the dynamics.

We represent in Figure 6.6 the normalized RDFs g_{OO} for the two central oxygen atoms O_1 and O_2 for various temperatures ranging from 200 to 400 K. The vertical lines indicate the average distance $\langle d_{\text{O}_1\text{O}_2} \rangle$ for each simulation. If we compare the protonated hexamer g_{OO} with those extracted from MD simulations of the Zundel cation at room temperature (upper

right panel of Figures 4.7 and 4.8), we notice that the shape of the function is almost identical, but the peak position is shifted by about ~ 0.05 Å. For instance, the average $\overline{O_1O_2}$ distance is equal to $\langle d_{O_1O_2} \rangle = 2.447(5)$ Å at room temperature ($T = 300$ K), being ~ 0.054 Å larger than the zero temperature result. Consequently, going back to Figure 6.5, it seems that at room temperature, the $H^+(H_2O)_6$ cluster mainly adopts an Eigen-like configuration, where the hydrated proton is covalently bound to one side water molecule. Therefore, the presence of solvating water molecule drastically reduces the mobility of the hydrated proton, which is much more localized with respect to the Zundel case, where it freely oscillates between the 2 neighboring water molecules.

The evolution of the shape of the g_{OO} functions when one increases the temperature is somewhat expected. Indeed, the low-temperature ($T = 200$ K) distribution is sharper than the high-temperature ($T = 400$ K) one, which displays longer tails in the large \overline{OO} distances region. This systematic broadening can be interpreted as follows. We provide more thermal energy to the system, thus increasing its kinetic energy and the average velocity of the particles. Consequently, the amplitude of the intra- and intermolecular vibrations is enhanced and larger values of the oxygen-oxygen distance are obtained. This is also consistent with the strongly anharmonic shape of the protonated hexamer PES (Figure 6.3) that becomes very flat for large $\overline{O_1O_2}$. Finally, we emphasize that the temperature dependance of the oxygen-oxygen RDFs is *monotonic* and saturates at 350 K, confirming that thermal effects are straightforward to interpret.

From now on, we will analyze the temperature-dependence of the normalized oxygen-proton RDFs g_{OH} that are plotted in Figure 6.7. At variance with the g_{OO} , the protonated water hexamer g_{OH} are completely different from the computed RDFs of the Zundel ion (bottom panels of Figures 4.7 and 4.8). Indeed, they are characterized by the presence of two distinct peaks around $d_{OH} = 1.1$ Å and $d_{OH} = 1.35$ Å, corresponding to a localized proton, covalently bound to its closest neighbor. We thus confirm that the more representative configuration of the protonated hexamer in the classical case is Eigen-type contributing to the long-range tails in the g_{OO} . Looking closer to Figure 6.7, we notice that the height of the two peaks decreases with the temperature, and the distributions are broadened. Indeed, from $T = 350$ K, the second maximum around $d_{OH} = 1.35$ Å, starts to disappear. This suggests that at this temperature, the ambient thermal energy is large enough to trigger spontaneous proton jumps inside the core of the protonated hexamer. Such proton hops involve the presence of a fully Zundel-like reaction intermediate, explaining the little increase of the g_{OH} around $d_{OH} = 1.2$ Å. Furthermore, we notice that the low-temperature ($T = 200$ K) g_{OH} displays a first peak at slightly larger \overline{OH} distance, compared with higher temperature RDFs where the peak position remains unchanged. This can be related to an enhanced proton mobility at higher temperatures.

To complete the analysis of the static properties of the classical protonated hexamer, we now analyze the bidimensional oxygen-oxygen and oxygen-proton probability distributions ρ_{2D} , represented in Figure 6.8. This probability distribution provides a graphical representation of the protonated hexamer PES at finite temperature projected along the reaction coordinate $\overline{O_1O_2}$. To further analyze the impact of thermal effects on the phase space configurations

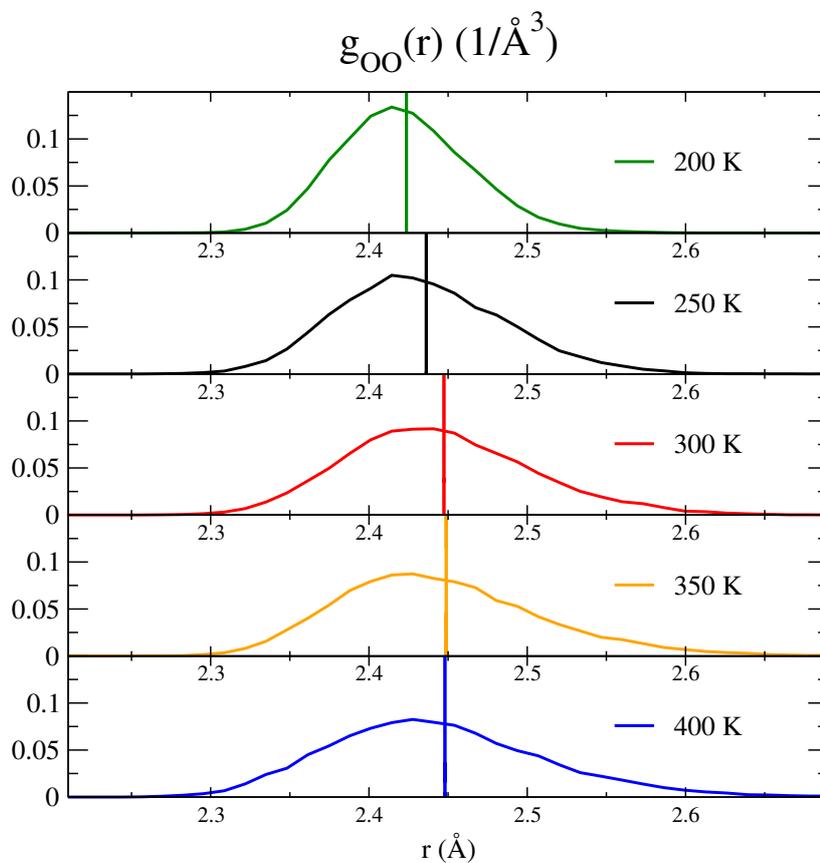


Figure 6.6 – Oxygen-oxygen RDFs obtained by QMC-driven CMPC-LD simulations at various temperatures ranging from 200 K to 400 K. The vertical lines indicate the $\langle d_{O_1O_2} \rangle$ for each simulation.

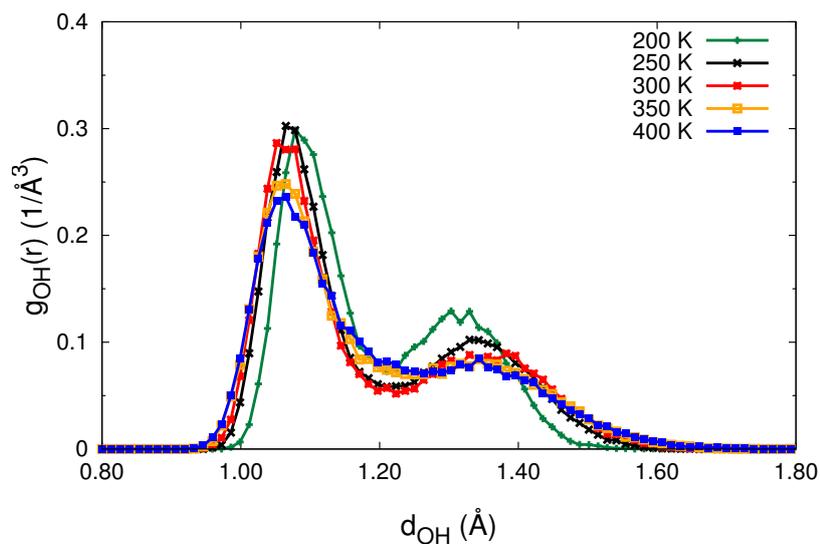


Figure 6.7 – Oxygen-proton RDFs obtained by QMC-driven CMPC-LD simulations at various temperatures ranging from 200 K to 400 K.

that may be visited by the hydrated proton, we also plot the equilibrium geometries obtained by VMC at $T = 0$ K (black circles).

First of all, we notice that the global shape of the distributions ρ_{2D} is very similar to those obtained for the Zundel cation, forming two wings that stretch along the equilibrium geometries for large \overline{OO} distances. This expected result confirms that the presence of the solvation shell does not drastically impact the global core geometry with no proton hopping from the Zundel core to the outer shell. At low temperature ($T = 200$ K), the configurations in the range $d_{O_1O_2} = 2.38 - 2.45$ Å are predominantly visited, while this region becomes larger and larger when one increases the cluster temperature: $d_{O_1O_2} = 2.38 - 2.53$ Å at $T = 300$ K and $d_{O_1O_2} = 2.35 - 2.58$ Å at $T = 400$ K), in agreement with the g_{OO} . However, the striking difference with the Zundel case is the depletion of the density ρ_{2D} in the short \overline{OO} distances region at low and room temperature. This stems from the quasi absence of symmetric or Zundel-like configurations, confirming again that, in the presence of classical particles, the protonated hexamer is described by an Eigen-like motif. However, when looking at the high-temperature distribution ($T = 400$ K), the density depletion, although still present, is less noticeable. The thermal energy is thus large enough to enable PT processes inside the core of the hexamer, confirming that the hydrated proton can classically cross the static barrier. The proton jumps are certainly less easy than in the Zundel complex since they imply a *concerted* rearrangement of the solvating molecules, which is more constraining.

The study of the structural properties of the classical protonated hexamer enabled us to stress the critical role played by the temperature to trigger spontaneous PT processes. Indeed, at low temperature, the hydrated proton is frozen around its Eigen-like minimum energy configuration. When the temperature becomes higher than the ambient one, the covalent and H-bonds fluctuations are large enough to enable proton jumps within the core, implying the hexamer goes through a symmetric Zundel-like transition state. These conclusions are however based on thermal equilibrium distribution functions, and must be confirmed by the analysis of the time-dependent properties of the system.

6.2.2 Proton displacement

In this Section, we study some dynamical properties of the protonated water hexamer, focusing our attention on the localization of the hydrated proton. We thus need to find clear indicators of the presence or the absence of PT processes occurring in the Zundel-like or Eigen-like core of the cluster.

The first intuitive observable that comes in mind is the *proton displacement* d defined as

$$d = d_{1/2} = \tilde{d}_{O_{1/2}H^+} - \frac{d_{O_1O_2}}{2}, \quad (6.1)$$

with $\tilde{d}_{O_{1/2}H^+}$ the distance projected onto the O_1O_2 segment. In the above equation, the $O_{1/2}$ notation suggests that the choice of the reference oxygen atom is arbitrary. Indeed, if the generated MD trajectory is long enough, the hydrated proton dynamics is *ergodic* and its residence time near the left O_1 or right O_2 oxygen atoms is *a priori* the same. Unfortunately, due to the computational cost of such calculations, our CMPC-LD trajectories are too short

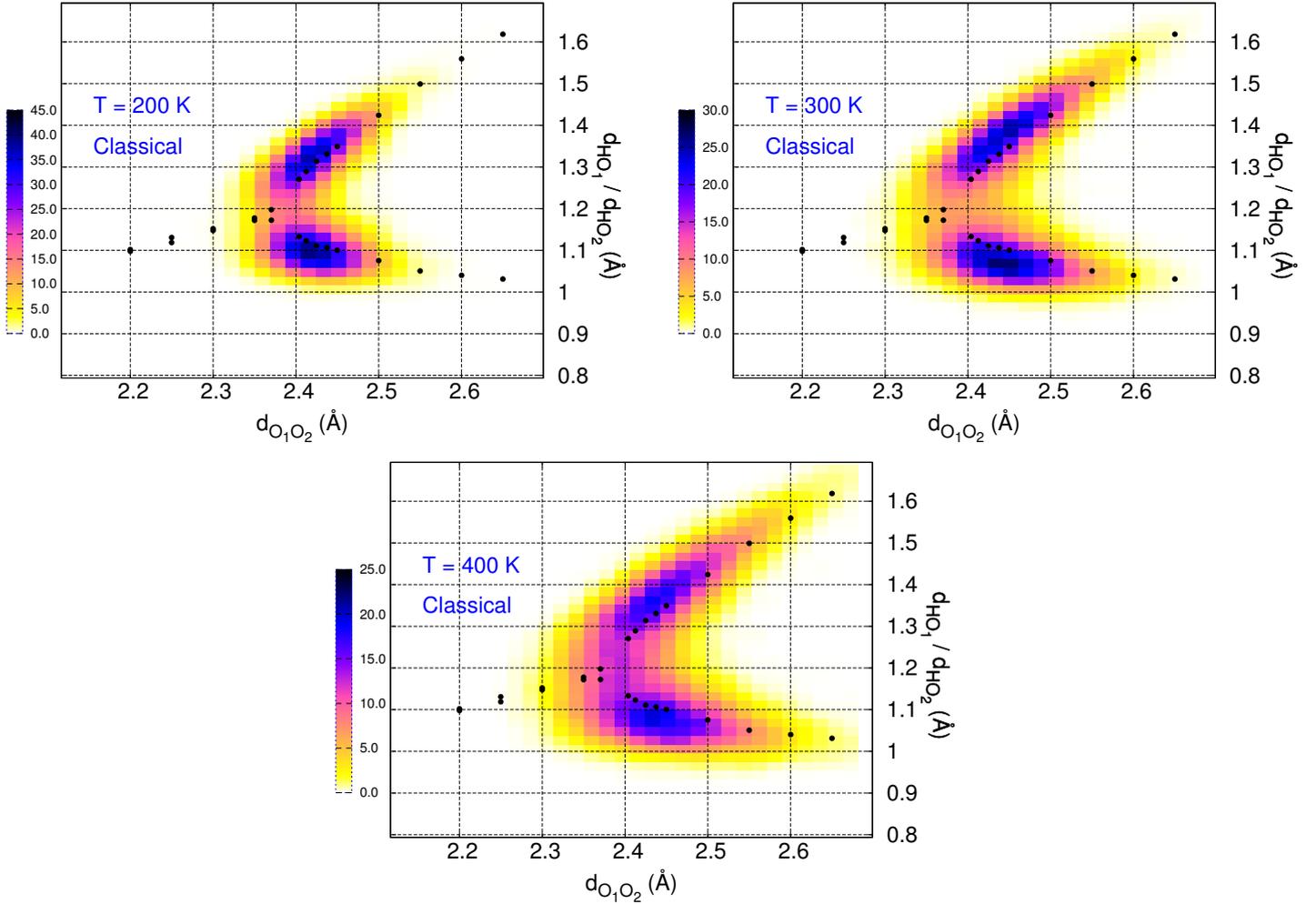


Figure 6.8 – Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven CMPC-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high temperature $T = 400$ K (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.

to ensure the full ergodicity of proton dynamics. Consequently, the obtained distributions with the simple application of the Eq. (6.1) would be asymmetric, which is not in agreement with the expected behavior of the cluster.

To solve that issue, we apply Eq. (6.1) taking each central oxygen atom as reference and we symmetrize the distribution by binning both the d_1 and d_2 distances. As we can see in the left panel of Figure 6.9, the obtained distributions are symmetric but not smooth because of the limited number of sampled classical configurations. The dark green curve, corresponding to a low-temperature ($T = 200$ K) simulation, displays two sharp peaks that confirm the trapping of the hydrated proton around its minimum energy Eigen-like configuration. When one increases the temperature, the position of the two peaks migrates to the larger values of $|d|$, which confirms the larger amplitude of the proton fluctuations. We also notice that one can directly relate the height of the central part ($|d| < 0.05$ Å) of the histograms to the number of times the protonated cluster has adopted its Zundel-like configurations, suggesting that spontaneous PT happened. It is clear that the higher the temperature, the more numerous are the Zundel-like configurations. The greater number of visited Zundel-like configurations when the cluster temperature increases is in agreement with Figure 6.8. The proton hopping is more frequent at higher temperatures.

We present in the right panel of Figure 6.9 another observable that can help to understand the nature of the configurations adopted by the $\text{H}^+(\text{H}_2\text{O})_6$ cluster: the *sharing proton coordinate*, denoted δ . It corresponds to

$$\delta = |d_{\text{O}_1\text{H}^+} - d_{\text{H}^+\text{O}_2}|, \tag{6.2}$$

and has been used in previous studies of aqueous proton defects in condensed phase^[246,252,253]. This observable, strongly correlated to the proton displacement d , gives us precious information about the symmetry of the cluster core. Indeed, when the considered configuration is symmetric or Zundel-like, the proton sharing displacement δ vanishes² because it is equally shared between its two neighboring water molecules. On the contrary, when the hydrated proton forms a covalent bond with its closest neighboring oxygen atom, δ takes finite values (up to $\delta \sim 0.8$ Å) which are the signature of Eigen-like states. Inspecting the right panel of Figure 6.9, we can notice that the larger the temperature, the larger the average sharing proton coordinate δ . This confirms that thermal effects enhance the amplitude of the hydrated proton fluctuations. In the meantime, the high-temperature distributions ($T = 350 - 400$ K) exhibit a plateau in the $\delta \sim 0$ region, corresponding to Zundel-like configurations. This further proves that the hydrated proton is less trapped at these temperatures and can more easily jump from one oxygen atom to another.

To pursue this configurational analysis, we would like to estimate, as quantitatively as possible, the weights of Zundel- and Eigen-like configurations from the distributions of the proton displacement d . To that purpose, we decided to fit the probability distribution $P(d)$ according

²as the proton displacement d does.

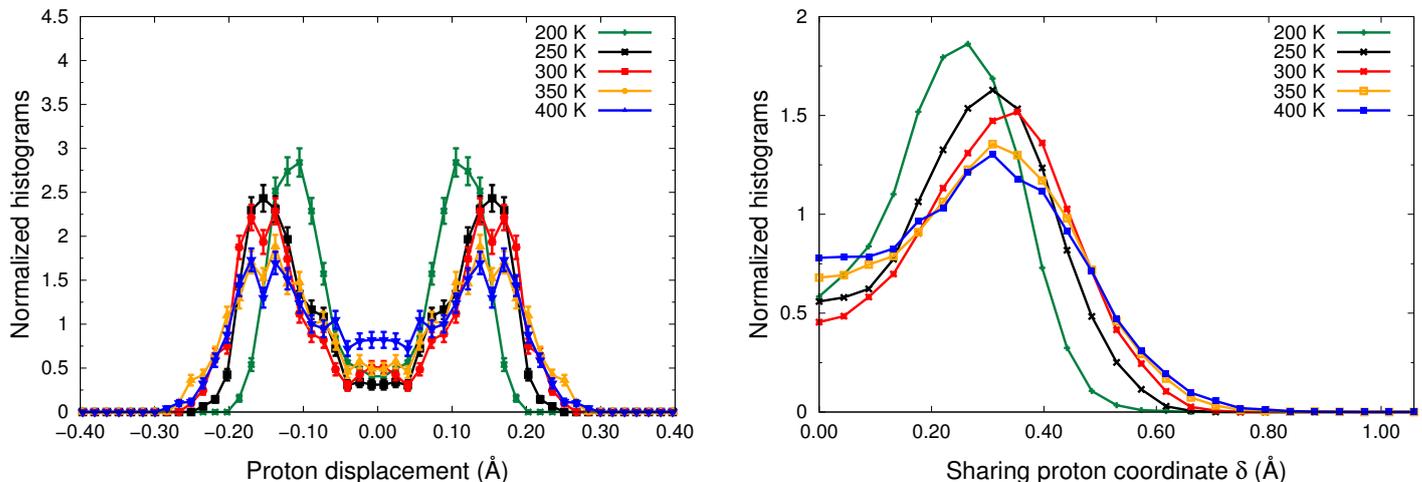


Figure 6.9 – Distributions of the proton displacement d with respect to the midpoint of the oxygen-oxygen distance (left panel) and the proton sharing coordinate δ (right panel), as a function of the temperature, for classical particles.

the following 3-gaussian, or 2-species model:

$$P(d) = \underbrace{\frac{(1 - \lambda_E)}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{d^2}{2\sigma_Z^2}}}_{\text{Zundel configuration}} + \underbrace{\frac{\lambda_E}{2\sqrt{2\pi\sigma_E^2}} \left[e^{-\frac{(d-d_E)^2}{2\sigma_E^2}} + e^{-\frac{(d+d_E)^2}{2\sigma_E^2}} \right]}_{\text{Eigen configurations}}, \quad (6.3)$$

where λ_E , d_E , σ_E and σ_Z are tunable parameters. The fitting function $P(d)$ of Eq. (6.3) is built to fulfill the normalization condition $\int P(d) = 1$ and to reproduce at best the signal observed in Figure 6.9. At least 2 of the 4 fitting parameters have a meaningful physical interpretation: λ_E simply represents the weight of Eigen-like configurations within the 2-species model, whereas d_E quantifies the distortion of the hydrated proton along the oxygen-oxygen distance.

The results of the fitting procedure are reported in Figure 6.10 and Table 6.5 for the 5 studied temperatures, ranging from $T = 200$ K to $T = 400$ K. The fitting parameters are given in Table F.1 of the Appendix F. We first note that, despite the noisy shape of the proton displacement distributions extracted from CMPC-LD simulations, due to the shortness of their trajectory, a good agreement with the fitting function $P(d)$ is found. The relevance of the 2-species model is further validated by the stable behavior of the 4 fitting parameters when one increasing the temperature. In particular, the parameters related to the Eigen-like configuration, namely λ_E , d_E and σ_E tend to increase with the temperature. This further proves that thermal effects tend to enhance the asymmetry of the protonated hexamer core, localizing the hydrated proton on a side water molecule.

It is also clear from Table 6.5, that the number of spontaneous proton jumps and charge rearrangements occurring inside the water hexamer core increases with the cluster temperature. Indeed, the weight of the symmetric Zundel-like transition states is multiplied by a factor

3 between the low- ($T = 200$ K) and the high- ($T = 400$ K) temperature regimes. This implies that the proton diffusion is enhanced by thermal effects, due to a greater proton mobility. This point will be later discussed in the Section 6.4. Let us also notice that, within a classical picture, the hydrated proton remains extremely localized since the Eigen-like structure still dominates the Zundel-like one at ambient conditions.

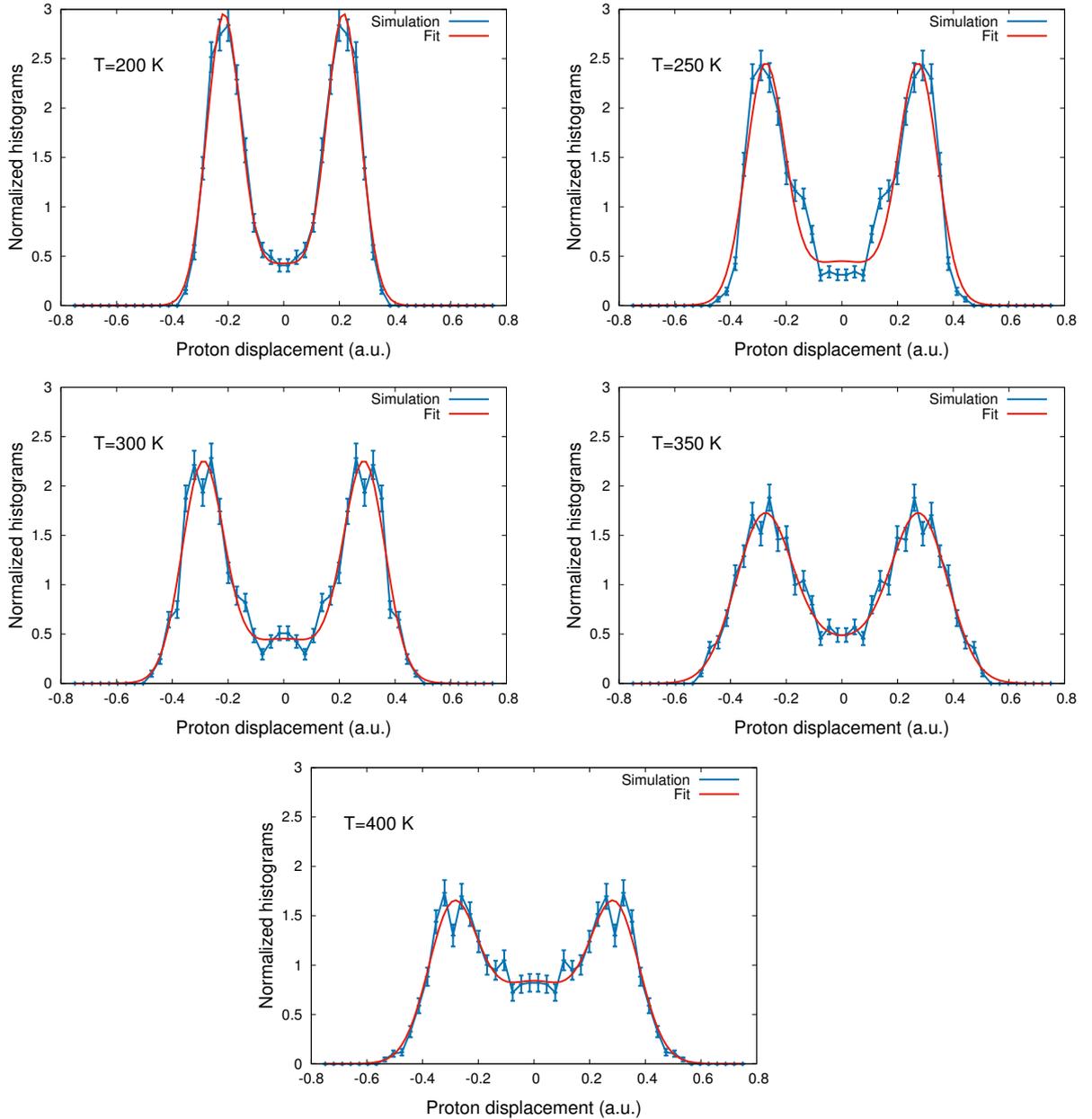


Figure 6.10 – Comparison of the $P(d)$ distributions obtained from QMC-driven CMPC-LD simulations (blue) and fitted using Eq. (6.3) for $T = 200, 250, 300, 350, 400$ K.

To conclude, we have elucidated the impact of thermal effects on the hydrated proton dynamics in the protonated water cluster. Indeed, in the low-temperature regime (up to $T = 200$ K), the proton is chemically inert since it is trapped around its minimum energy configura-

T (K)	% Zundel	% Eigen
200	13	87
250	20	80
300	20	80
350	18	82
400	38	62

Table 6.5 – Species proportions obtained by fitting distributions of the proton displacement with the 3-gaussian model of Eq. (6.3) in the classical particles case.

tion, which is Eigen-like. At room conditions, the proton trapping is found to remain quite important and the predicted majority configuration of the $\text{H}_{13}\text{O}_6^+$ is still Eigen-like, which is in contradiction with recent experimental data^[17,247]. Indeed, these works emphasize the paramount importance of the Zundel complex as transition state for the hydrated proton dynamics in water and aqueous solutions. As we have seen in Chapter 4, NQE dominate over thermal effects, especially at ambient conditions: their essential role must be taken into account, as we will do in the following Section.

6.3 Quantum hexamer dynamics

Here, we follow exactly the same route taken in the previous Section but within a quantum framework. Indeed, NQE are known to be crucial in water and one cannot expect to provide a quantitative description of PT in bulk liquid without caring about them. Within the Feynman Path Integral (PI) framework, we will analyze the impact of such effects on the structural and dynamical properties of the $\text{H}^+(\text{H}_2\text{O})_6$ cluster at finite temperature. In the last Subsection (6.3.3), the competition between NQE and thermal effects will be investigated.

6.3.1 Radial distribution functions

We performed a series of QMC-driven Langevin Dynamics using the Path Integral Ornstein Uhlenbeck (PIOUD) algorithm introduced in Chapter 3 at five various temperatures ranging from 200 to 400 K. The simulation parameters, and more specifically the temperature, are taken to be the same as those employed for the propagation of the classical trajectories, to enable a direct comparison of the observables extracted from the LD. Within the PI approach, one must also specify the number P of beads that will be used to describe the quantum particles, particularly the hydrated proton. This number is deduced from the convergence of the quantum kinetic energy of the Zundel complex (upper left panel of Figures 3.1 and 3.2), which is justified by the similarity between the two considered clusters. Consequently, we choose $P = 64$, for the lowest temperature simulation ($T = 200$ K) while we used the value of $P = 32$ for the other calculations.

We represent in Figure 6.11 the quantum oxygen-oxygen RDFs g_{OO} of the protonated water hexamer at various temperatures. As already observed in the Zundel cation, there is a broadening of the g_{OO} due to the presence of NQE, which is significant at all temperatures taken into account in our study. Let us however note that at $T = 200$ K, the peak of the

quantum g_{OO} is shifted to larger \overline{OO} distances, with respect to the classical case. In the high-temperature limit, the shape of the quantum RDFs g_{OO} is also modified: the peak position is decreased and the Pair Correlation Functions (PCFs) display longer tails, especially at large \overline{OO} distances ($d_{O_1O_2} \sim 2.55 - 2.6 \text{ \AA}$). This suggests that thermal effects start to take over NQE, leading to larger fluctuations of the oxygen-oxygen distance. We thus expect to find a larger weight of classical Eigen-type configurations, following the classical high-temperature limit of quantum systems.

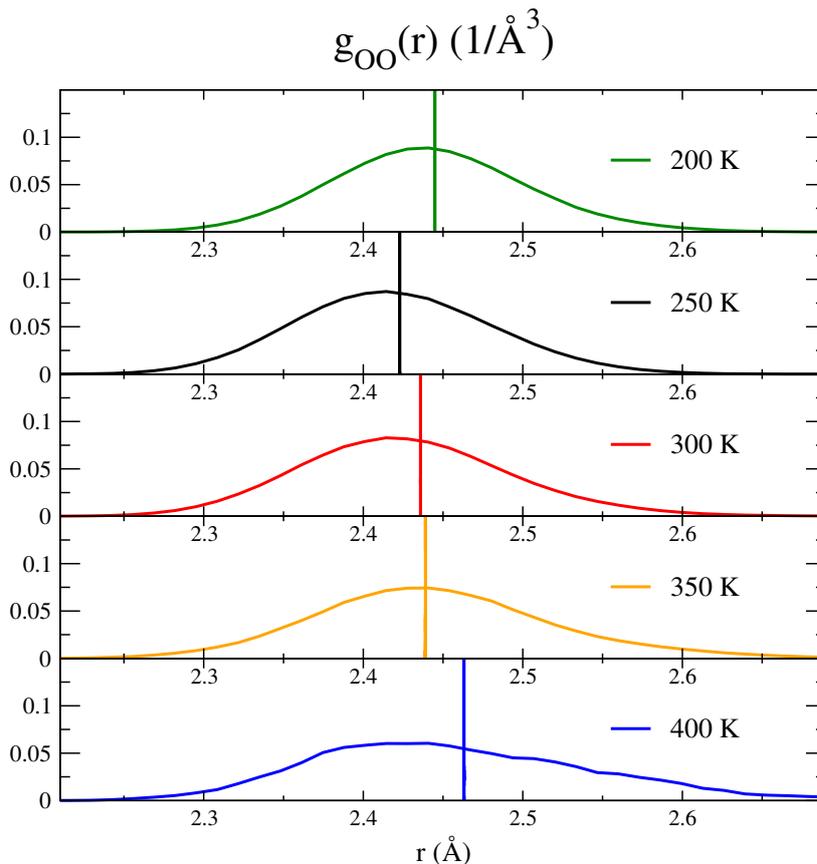


Figure 6.11 – Oxygen-oxygen RDFs obtained by QMC-driven PIOUS-LD simulations at various temperatures ranging from 200 K to 400 K. The vertical lines indicate the $\langle d_{O_1O_2} \rangle$ for each simulation.

Although this gives some useful information on the structural properties of the $H^+(H_2O)_6$ cluster, the analysis of the quantum oxygen-oxygen RDFs is far from being sufficient since NQE are moderate for the oxygen atoms. On the contrary, such quantum effects are spectacular for the proton, and one should inspect the oxygen-proton RDFs g_{OH} to gather further information on the hydrated proton localization. The protonated hexamer g_{OH} are represented in Figure 6.12 for various temperatures. First of all, let us notice that the quantum RDFs g_{OH} (and g_{OO}) are smoother and less noisy than their classical counterparts. To improve the statistics during the computation of these observables, we decided to consider all the beads instead of retaining the configurations of only one arbitrary bead as we did for the Zundel ion³.

³Even though the bead positions are correlated within a necklace, keeping all of them in the quantum

Second, at variance with the oxygen-oxygen RDFs, the protonated hexamer g_{OH} are more similar to those obtained for the Zundel complex. Indeed, the g_{OH} displays here a unique peak. This suggests that the trapping of the quantum hydrated proton is considerably reduced since it can spontaneously jump from one side water molecule to the other, leading to a much more important weight of the Zundel-like configurations during the cluster dynamics. Therefore, it looks that the Zero Point Energy and quantum tunneling effects are essential to help the hydrated proton to go across the static PT barrier, leading to an increased proton mobility. Another consequence of NQE is the delocalization of the nuclei that leads to much faster rearrangements of the solvation shell, making instantaneous proton jumps much easier. We also notice that the low and room temperature ($T = 200 - 300$ K) oxygen-proton RDFs almost superimpose, while the higher temperatures ($T = 350 - 400$ K) g_{OH} are broadened and present larger tails, especially in the large \overline{OH} distance limit. This seems to be linked to the already mentioned modification of the g_{OO} at these temperature, as discussed above. This can be attributed to thermal effects that likely start to dominate over NQE, pushing the hydrated proton to form strong covalent bonds with its neighboring oxygen atoms.

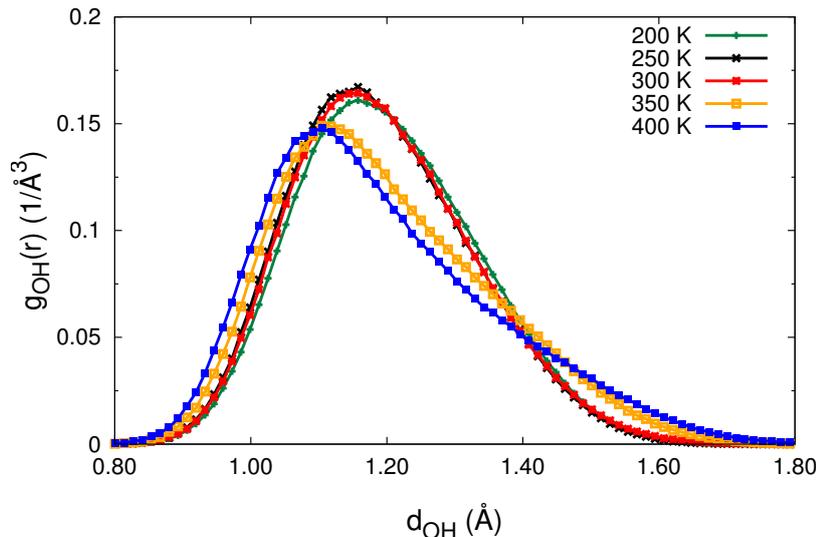


Figure 6.12 – Oxygen-proton RDFs obtained by QMC-driven PIOUD-LD simulations at various temperatures ranging from 200 K to 400 K.

To pursue our analysis of the static properties of the quantum protonated hexamer, we represent the bidimensional oxygen-oxygen and oxygen-proton probability distributions $\rho_{2\text{D}}$ in Figure 6.13. We also plot this quantity in Figure 6.14 for the *centroid*, whose coordinates are defined by

averages improves the statistics.

$$\begin{aligned}
 \mathbf{p}_i^{\text{centro}} &= \frac{1}{P} \sum_{j=1}^P \mathbf{p}_i \\
 \mathbf{q}_i^{\text{centro}} &= \frac{1}{P} \sum_{j=1}^P \mathbf{q}_i,
 \end{aligned} \tag{6.4}$$

for $i = 1 \dots N_{\text{at}}$ ($N_{\text{at}} = 19$ for the $\text{H}^+(\text{H}_2\text{O})_6$ cluster). The centroid analysis is a useful tool since it gives information about the *average* position of the ring polymer, disregarding its spatial extension that can blur the signal. Indeed, if we look at the low and room temperature $\rho_{2\text{D}}$ in the quantum case, we merely observe a characteristic spread distribution, corresponding to a strongly delocalized proton, with no major difference between the two distributions. Nevertheless, the situation is different for the centroid: at $T = 200$ K, we notice a slight depletion of $\rho_{2\text{D}}$ in the *symmetric* region ($\overline{\text{O}_1\text{O}_2} = 2.35 \text{ \AA}$), while the symmetric region is correctly explored at $T = 300$ K. Once again, it proves that, despite the evidence, the hydrated proton remains partially *trapped* at low temperature, explaining the observed shift of the g_{OO} to larger oxygen-oxygen distances. At room temperature, the bidimensional distribution $\rho_{2\text{D}}$ takes larger values around $\overline{\text{OO}} = 2.4 \text{ \AA}$, indicating the release of Zundel-like configurations during the protonated hexamer dynamics. In the classical case, the central proton is still trapped, proving that NQE are *a priori* essential to provide a good proton diffusion at ambient conditions. In the high-temperature limit, the shape of $\rho_{2\text{D}}$ is no longer the same and the cluster starts to behave similarly to the classical case. The centroid $\rho_{2\text{D}}$ at $T = 400$ K is quite similar to the classical $\rho_{2\text{D}}$ at the same temperature, confirming that thermal effects start to take over NQE in the high-temperature limit.

To summarize, we have demonstrated once again the paramount importance of NQE to properly describe the microscopic mechanisms underlying the PT processes in water clusters. Indeed, at variance with the classical case, the Zundel-like configuration is certainly the most representative structure of the $\text{H}_{13}\text{O}_6^+$ ion at finite temperature, where the hydrated proton mobility is strongly enhanced. This is in agreement with Ref. 247, where the experimental evidence of fast nonoscillatory proton-dynamics in a Zundel-type solvation structure has been found at $T_{\text{exp}} = 293$ K. The Eigen-like configuration of the hydrated proton is however still present, especially in the low- and high- temperature regimes. Nevertheless, these considerations remain qualitative, and must be confirmed by a more focused analysis on the proton displacement properties of the protonated water hexamer.

6.3.2 Proton displacement

We have introduced in Subsection 6.2.2 two useful observables to evaluate the hydrated proton localization inside the Zundel- or Eigen-like core of the $\text{H}^+(\text{H}_2\text{O})$ cluster, namely the proton displacement d and the proton sharing coordinate δ . We plot these quantities as a function of the temperature in the left and right panels of Figure 6.15, respectively.

Looking at the histograms of the proton displacement d , we can see that, except at low temperature ($T = 200$ K), the quantum distributions do not have the same symmetry as their

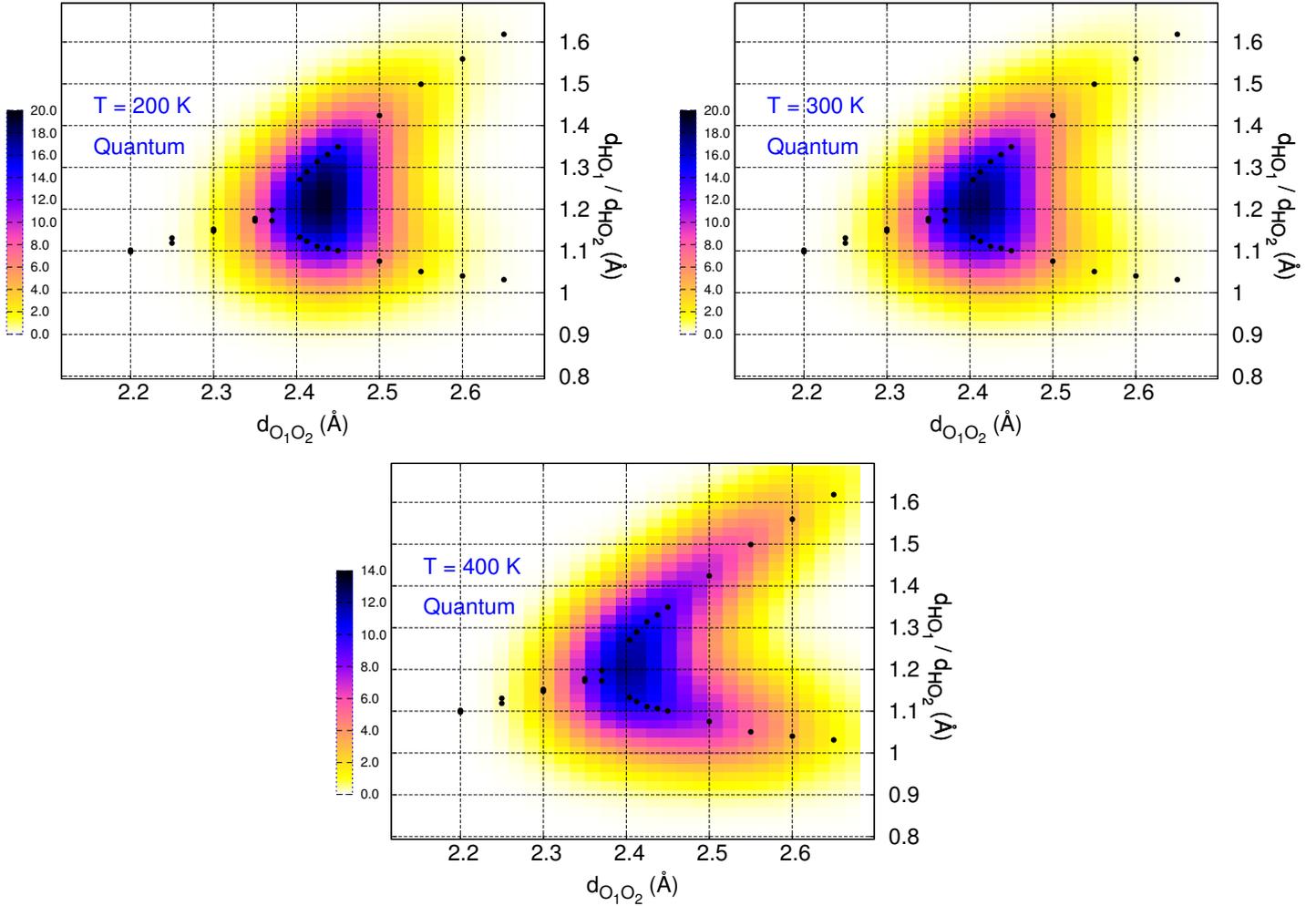


Figure 6.13 – Bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven PIOUD-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high-temperature (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.

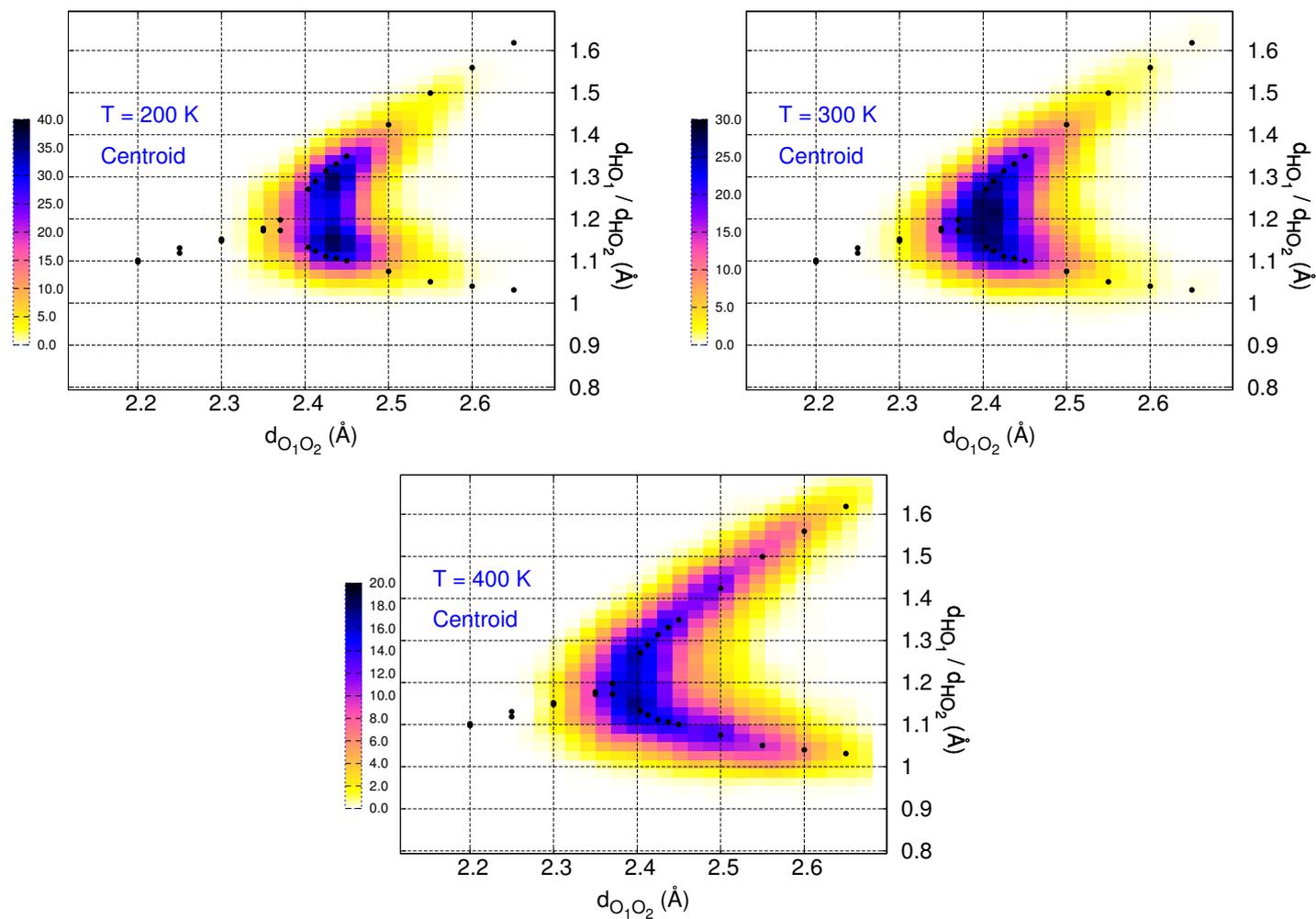


Figure 6.14 – Centroid bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven PIOUD-LD simulations at low temperature $T = 200$ K (upper left panel), room temperature $T = 300$ K (upper right panel) and high temperature $T = 400$ K (bottom panel). The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.

classical counterparts (bottom left panel of Figure 6.9). Indeed, at $T = 200$ K, the proton displacement presents two maxima around $|d| = 0.1$ Å, typical of the partial trapping of the hydrated proton when a centroid analysis or running averages are carried out (bottom left panel). When one increases the simulation temperature close to ambient conditions, the total amount energy, coming from the sum of the ZPE with the thermal energy, is larger than the PT static barrier. This is characterized by the merging of the peaks into a single maximum at $d = 0$ Å, corresponding to a majority of symmetric (or Zundel-like) configurations. This temperature transition between an asymmetric to a symmetric structure of the protonated hexamer core is consistent with the change of the centroid ρ_{2D} discussed above. We also notice that the height of the distribution at $d = 0$ is maximal around ambient conditions, suggesting that the presence of the Zundel-like configuration is favored in the $T = 250 - 300$ K region. Consequently, we expect the hydrated proton mobility to be optimal near room temperature. Indeed, it corresponds to the best compromise between acquiring enough energy to go across the static barrier and, in the meantime, controlling the amplitude of the chemical (covalent or H-) bonds fluctuations that might trap the proton. Indeed, when the temperature is again increased ($T = 350 - 400$ K), there is a significant broadening of the proton displacement histograms to larger values of $|d| \sim 0.3 - 0.4$ Å, which corresponds to the emergence of Eigen-like configurations. The latter ones tend to localize the hydrated proton around its closest neighboring oxygen atom, reducing its mobility.

The configurational transition between a Zundel-like to an Eigen-like dynamics of the protonated hexamer cluster around $T = 350$ K is also characterized by the broadening of the proton sharing coordinate δ distributions (right panel of Figure 6.15). Nevertheless, we are not able to distinguish the low-temperature ($T = 200$ K), where the hydrated proton is partially trapped, from the ambient temperature ($T = 250 - 300$ K) distributions, where the excess proton motion is enhanced. This is simply due to a different processing of the MD trajectories⁴ during the evaluation of the d or δ histograms. Indeed, we performed *running averages* using a $t_{\text{run}} = 200$ fs time window to disregard the proton rattling defined as fluctuations in which the proton returns to the original oxygen atom after a series of PT events within a Zundel-like state^[254]. Moreover, our choice is further justified by recent absorption spectrum experiments which established that within the lifetime of a particular hydration geometry (about ~ 1 ps), the proton explores a multitude of positions on a sub-100-femtosecond time scale^[17]. Consequently, performing such running averages looks appropriate since we aim at isolating the different hydrations geometries of the quantum proton, that strongly depend on the temperature of the $\text{H}^+(\text{H}_2\text{O})_6$ cluster. We also point out that without the use of running averages, we would not have been able to distinguish the partial trapping of the hydrated proton in the low-temperature regime (top left panel of Figure 6.15).

To evaluate the weight of the Zundel and Eigen species during the protonated hexamer dynamics as a function of the temperature, we decided to fit the proton displacement distributions according to the 2-species model $P(d)$ given in Eq. (6.3). This fitting function, efficiently used for the classical description of the cluster, is however not able to faithfully reproduce the quantum distributions because they do not have the same shape, especially in the low-temperature regime. Indeed, we highlight that, using running averages, we detected a partial

⁴This was also the case for classical particles, but this discussion was irrelevant since there was no clear temperature-induced configurational transition of the classical protonated hexamer.

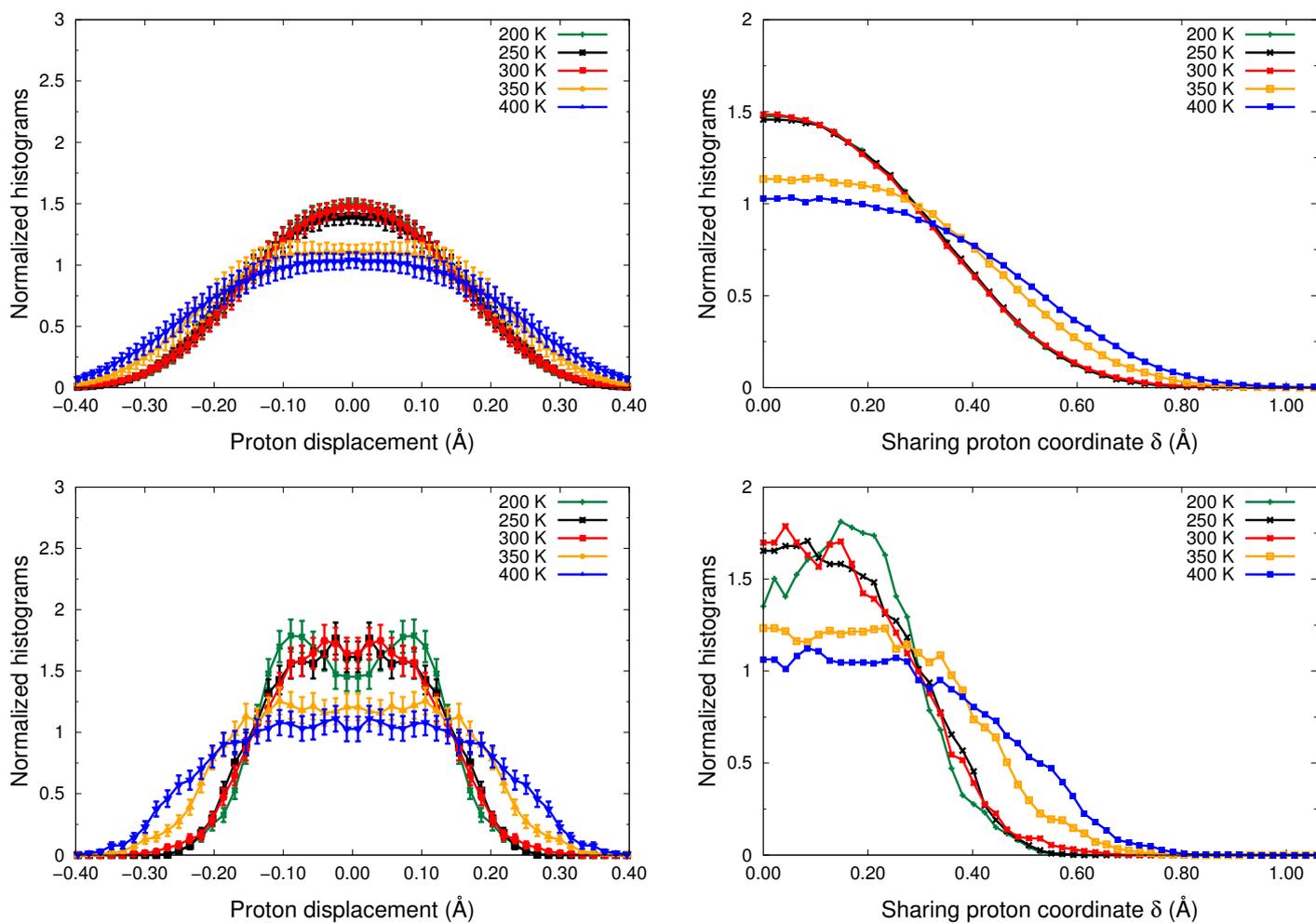


Figure 6.15 – Left panels: quantum distributions of the proton displacement d with respect to the midpoint of the oxygen-oxygen distance with (bottom) or without (top) running averages. Right panels: quantum (top) and centroid (bottom) distributions of the proton sharing coordinate δ .

6.3. Quantum hexamer dynamics

trapping of the hydrated proton at $T = 200$ K. However, looking at the bottom right panel of Figure 6.15, the symmetry of the protonated hexamer core differs from the usual Zundel- and Eigen-like configurations. This suggests the presence of a third species that needs to be identified. To solve the fitting problem, we propose on the basis of the previous remark, a novel *3-species* model, where the fitting function is composed by 5 distinct gaussians:

$$Q(d) = \underbrace{\frac{(1 - \lambda_E - \lambda_{EZ})}{\sqrt{2\pi\sigma_{SZ}^2}} e^{-\frac{d^2}{2\sigma_{SZ}^2}}}_{\text{Short Zundel configurations}} + \underbrace{\frac{\lambda_{EZ}}{2\sqrt{2\pi\sigma_{EZ}^2}} \left[e^{-\frac{(d-d_{EZ})^2}{2\sigma_{EZ}^2}} + e^{-\frac{(d+d_{EZ})^2}{2\sigma_{EZ}^2}} \right]}_{\text{Elongated Zundel configurations}} + \underbrace{\frac{\lambda_E}{2\sqrt{2\pi\sigma_E^2}} \left[e^{-\frac{(d-d_E)^2}{2\sigma_E^2}} + e^{-\frac{(d+d_E)^2}{2\sigma_E^2}} \right]}_{\text{Eigen configurations}}, \quad (6.5)$$

where λ_E , λ_{EZ} , d_E , d_{EZ} , σ_E , σ_{EZ} and σ_{SZ} are tunable parameters. The 5 gaussians of the above equation describe the populations of 3 distinct species, namely: the short Zundel (SZ), the elongated Zundel (EZ) and the Eigen (E). 'Short' and 'Elongated' refer to the value of the corresponding oxygen-oxygen distance that, as already discussed in the previous Subsection, strongly depends on the temperature of the system. On the one hand, the short Zundel corresponds to the usual representation of the Zundel ion, which is by definition fully symmetric: the cluster core belongs to the C_2 symmetry group and the hydrated proton is located at the middle of the inter-oxygen distance. On the other hand, the elongated Zundel is also C_2 -symmetric but the excess proton is no longer equally shared between the two sides molecules and is softly localized on one neighboring oxygen atom. One could criticize this choice that seems arbitrary but it is justified by the difference of behavior of these two species when they interact with the Eigen cation. Indeed, as we will see in the last Section of this Chapter (6.4), the short Zundel exhibits much more numerous and efficient PT processes with the Eigen ion than the elongated Zundel. Indeed, the existence of a short oxygen-oxygen distance is crucial to enable a large hydrated proton mobility via the Grotthuss or Eigen-Zundel-Eigen (EZE) mechanism^[13,163].

The proton displacement distributions extracted from QMC-driven PIOUS simulations and obtained by fitting the $Q(d)$ function given in Eq. (6.5) are compared in Figure 6.16, for various temperatures ranging from 200 to 400 K. The agreement between the simulated and the fitted distributions is excellent, validating the 3-species model described above.

To pursue our species analysis, we report the estimated percentages of the three distinct configurations that the protonated hexamer takes during its dynamics in Table 6.6. The detailed evolution of the fitting parameters is given in Table F.2 that can be found in Appendix F. At low temperature ($T = 200$ K), the protonated hexamer is almost exclusively in its elongated Zundel configuration. The hydrated proton is thus trapped around its minimum energy configuration which is asymmetric. Nuclear Quantum Effects, particularly the ZPE, delocalize the proton which is no longer covalently bonded to one of the neighboring oxygen atoms, at variance with the Eigen configuration. Instead, it shares two *distorted* H-bonds with its oxygen neighbors. Due to the greater impact of thermal effects, the weight of the elongated Zundel diminishes when the cluster temperature increases, indicating the onset of PT processes at near ambient conditions. Interestingly, in the high-temperature regime, the ratio between the overall Zundel-like and Eigen-like species is found to be 1:1, in agreement with the EZE mechanism for the proton transfer in bulk water. However, at variance with the classical case, we do not expect the proton mobility to linearly increases with the cluster

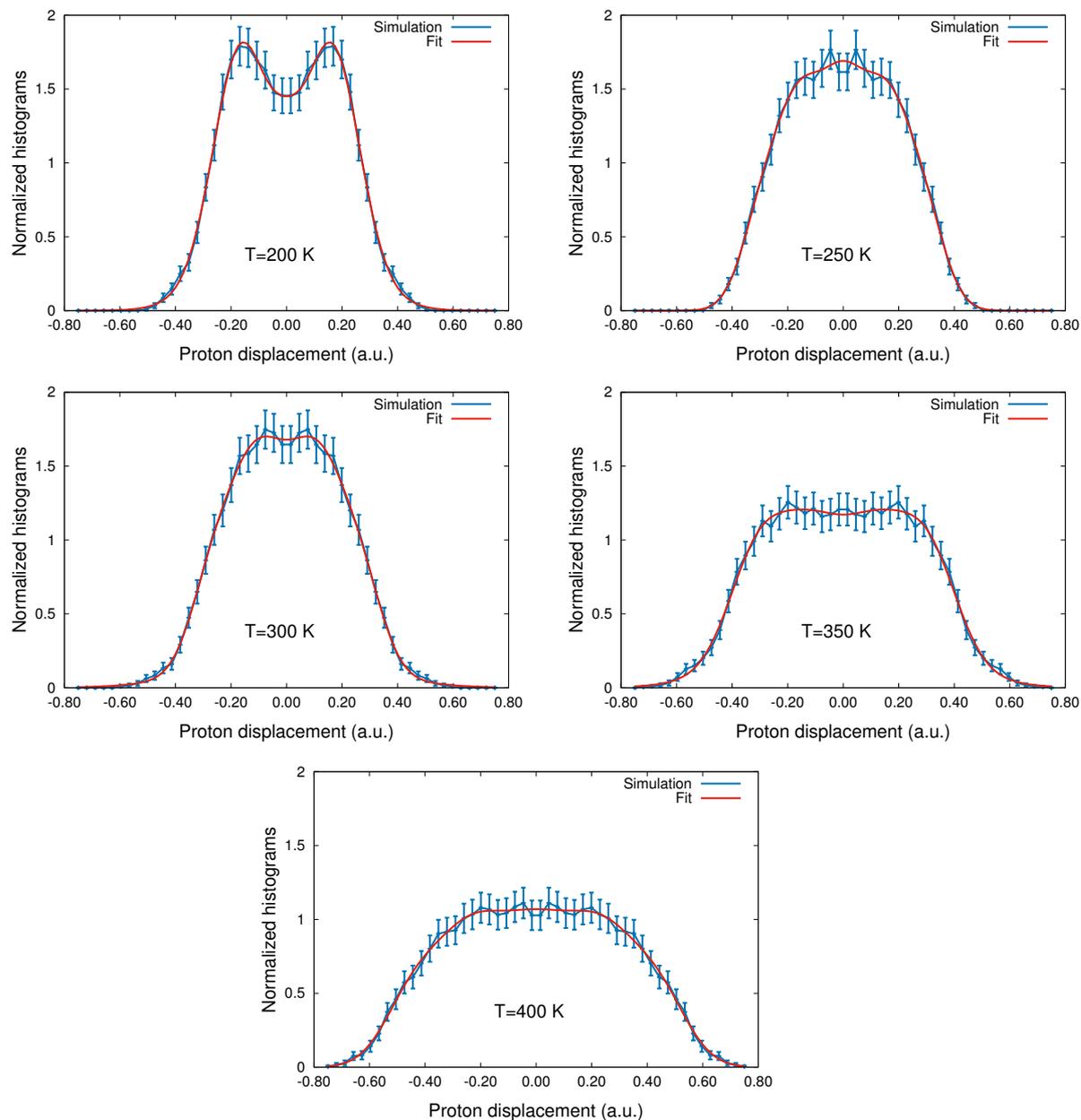


Figure 6.16 – Comparison of the $Q(d)$ distributions obtained from QMC-driven PIOUD-LD simulations (blue) and fitted using Eq. (6.5) for $T = 200, 250, 300, 350, 400$ K.

temperature because of the PT process between the Eigen and the short Zundel species is less efficient than between the short and elongated Zundel. Indeed, these two configurations belong to the same symmetry group and it thus costs less energy to the solvation shell to react during the charge rearrangements. Strickly speaking, we can refer to a 'proton jump' or the PT process occurring between a Zundel- and an Eigen-type species whereas the term 'proton shuttling' or 'special pair dance' should be employed for the short-to-elongated Zundel process and vice versa^[255].

T (K)	% Short Zundel	% Elongated Zundel	% Zundel	% Eigen
200	3	95	98	2
250	14	70	84	16
300	13	71	84	16
350	31	20	51	49
400	31	20	51	49

Table 6.6 – Estimated species proportions obtained by fitting distributions of the proton displacement with the 5-gaussian model in the quantum particles case.

To conclude, we learned from the analysis of the dynamical properties that NQE dramatically change the hydrated proton picture in the protonated water hexamer, and by extension in the bulk liquid. Indeed, NQE naturally increase the hydrated proton mobility which is predicted to be optimal near ambient conditions. Indeed, within the 3-species model, the excess proton mobility seems to be strongly related to the nature of the species present in the cluster. The (short) Zundel complex is identified as a true configurational state of the system, and should no longer be considered as a short-lived transition state, in agreement with recent spectroscopic studies^[17]. The subtle competition between NQE and thermal effects is thus of paramount importance to fully understand the extreme mobility of the hydrated proton around ambient conditions.

6.3.3 Thermal effects versus Nuclear Quantum Effects

In 2017, Wilkins *et al.* established that, at room temperature, NQE accelerate by $\sim 13\%$ the water dynamics compared to a classical description of the ions^[256], performing both classical and quantum MD simulations with the flexible q-TIP4P/F potential^[257]. They also assess NQE do not change the jump amplitude distributions and no significant tunneling is found. The faster jump dynamics of water is quantitatively related to a less structured g_{OO} , when NQE are included. In this Section, we investigate their role in the protonated hexamer, especially their importance when the temperature varies.

To that purpose, we go back to the bidimensional oxygen-oxygen and oxygen-proton distribution functions ρ_{2D} that are represented in Figures 6.8, 6.14 and 6.13 for classical, centroid and quantum particles, respectively. We have suggested in the two previous Sections that thermal effects tend to increase the hydrated proton mobility in a classical description, while the quantum case seems to be more complex. To pursue the analysis, we plot the *differences* between the bidimensional distributions ρ_{2D} calculated at a higher temperature with the obtained signal at a lower temperature, for three cases: $\rho_{2D}(T = 300 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$ (upper panels)

Chapter 6. The protonated water hexamer

to study the low-temperature regime, $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 300 \text{ K})$ (central panels) for the high-temperature limit, and finally $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$ (bottom panels) to have a complete view of the temperature-dependence of NQE.

Let us start with the interpretation of the low-temperature regime. For the classical protonated hexamer, it is clear that thermal effects tend to stretch the average oxygen-oxygen distance, which leads to the onset of PT processes at room temperature via the special pair dance (red wings around $d_{O_1O_2} = 2.5 \text{ \AA}$). The hydrated proton remains however trapped at $T = 200 \text{ K}$, as indicated by the blue central distribution at $d_{O_1O_2} = 2.4 \text{ \AA}$, because of the PT static barrier. The scenario is strikingly different for both the centroid and quantum particles that display positive values of $\rho_{2D}(T = 300 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$ for *shorter* oxygen-oxygen distances ($d_{O_1O_2} < 2.4 \text{ \AA}$). This indicates that NQE shorten the average oxygen-oxygen distance, that consequently make the protonated hexamer core symmetric or Zundel-like. Looking at the quantum particles case (upper right panel), we see no clear shape difference between the red and the blue distributions, confirming there is no deep structural change between the majority low-temperature species, namely the elongated Zundel and the short Zundel which dominates near ambient conditions. We point out that our naming of 'short' and 'elongated' for the two aforementioned species becomes now obvious. In the centroid description, we note that the low-temperature negative signal is situated *inside* the region defined by the zero temperature asymmetric branches (black circles) while the positive region mainly locates *outside* of it. This further tells that, in spite of the presence of NQE, the hydrated proton is yet partially trapped at $T = 200 \text{ K}$ while the combination of NQE and thermal effects provide it enough energy to freely move inside the protonated hexamer core.

In the high-temperature limit, the classical distribution difference also becomes slightly positive for short oxygen-oxygen distances ($d_{O_1O_2} = 2.3 - 2.4 \text{ \AA}$), indicating that the weight of fully symmetric or short-Zundel configurations increase with the temperature, in agreement with Table 6.5. Therefore, the hydrated proton is much more mobile at high temperature because the solvation shell moves fast enough to redistribute the charge within the core of the cluster during the PT process. Moreover, we notice the appearance of the red wings at large oxygen-oxygen distances ($d_{O_1O_2} = 2.5 - 2.7 \text{ \AA}$), which are the signature of the presence of Eigen-like states, with a strongly localized proton even in the quantum case. This is related to the disappearance of the elongated Zundel configurations, indicated by the blue distribution around $d_{O_1O_2} \sim 2.4 \text{ \AA}$, confirming that the Eigen-like state is favored at high temperature. We merely note that the frontier between the (elongated) Zundel-like and the Eigen-like regions is very clear within this representation, suggesting that their structures significantly differ from each other. Therefore, when the overall symmetry of the protonated hexamer is modified, it costs a lot of thermal energy ($T \sim 200 \text{ K}$, the PT static barrier) to the solvation shell to rearrange during the charge transfer.

The above discussion can be summarized looking at the plots at the bottom of Figure 6.17, representing the $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$ difference. In a classical description, the $H^+(H_2O)_6$ cluster is mainly asymmetric or Eigen-like, with a localized hydrated proton than can hop according to a EZE mechanism when the cluster temperature is large enough ($T > 350 \text{ K}$). On the other hand, we can distinguish two regimes in the quantum case. Under

6.3. Quantum hexamer dynamics

low-temperature conditions ($T = 200$ K), the hydrated proton is partially trapped, despite its delocalization due to NQE. When one increases the cluster temperature, the central oxygen-oxygen significantly shortens thanks to the combination of NQE and thermal effects which give to the cluster just the necessary energy to enable proton shuttling within its core. At higher temperature, the thermal energy brought to the system is even larger, making the hydrated proton hop to side water molecules, via the special pair dance or the EZE mechanism. The simultaneous disappearance of the elongated Zundel species and appearance of the Eigen-like configurations is clearly noticeable examining $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$ for the centroid. Finally, let us note that the spatial extension of the wings is the same along the \overline{OO} distribution for both classical, centroid and quantum particles, indicating that there is no significant change in the proton jumps amplitude, in agreement with recent water studies^[256].

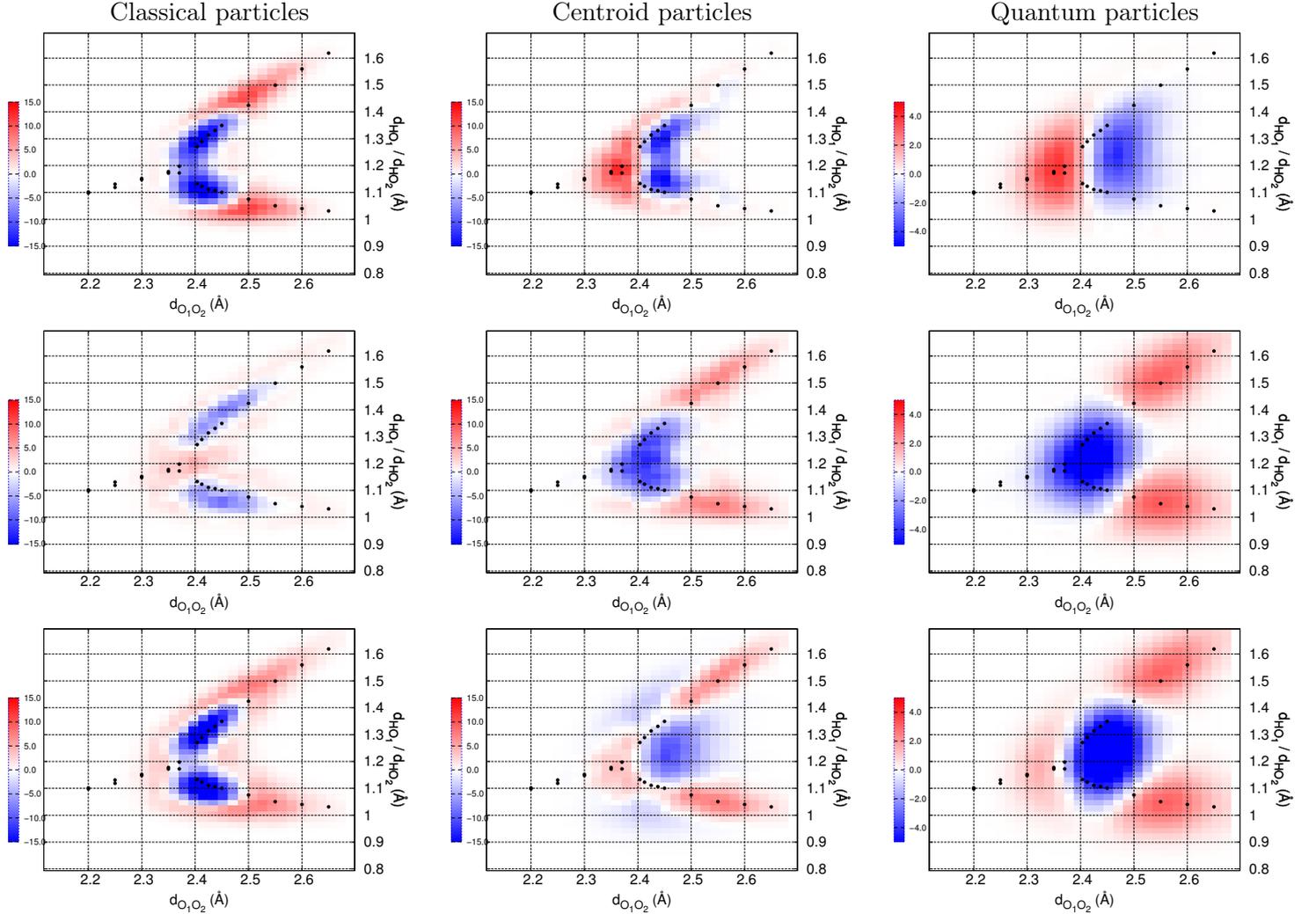


Figure 6.17 – Difference between bidimensional oxygen-oxygen and oxygen-proton distributions obtained by QMC-driven LD simulations for classical (left panels), centroid (central panels) and quantum (right panels) particles. Upper panels: $\rho_{2D}(T = 300 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$. Middle panels: $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 300 \text{ K})$. Bottom panels: $\rho_{2D}(T = 400 \text{ K}) - \rho_{2D}(T = 200 \text{ K})$. The black circles correspond to the zero temperature equilibrium geometries of the protonated hexamer.

To conclude, we confirmed that NQE deeply impact the nature and the dynamics of the hydrated proton in the protonated water hexamer. Indeed, there is a subtle competition between NQE and thermal effects than can be only understood within the 3-species model that we designed. The classical hydrated proton is Eigen-like, whereas it displays a Zundel-like symmetry in the quantum case because of its delocalization due to NQE. The average oxygen-oxygen distance plays a critical role for the proton mobility in the cluster since the elongated Zundel configuration, existing at $d_{O_1O_2} = 2.4 - 2.5 \text{ \AA}$, is the essential species to strongly enhance PT in the protonated hexamer. Indeed, it has a large cooperativity with the short Zundel configurations of the clusters, at variance with the Eigen that is inert. Nevertheless, this species analysis does not provide a quantitative description of the existing proton hopping within the Zundel- or Eigen-like core of the protonated hexamer. Therefore, there is need to find new observables to quantify the PT rate in this system, as we will do in the following.

6.4 Proton transfer in the protonated hexamer

In this Section, we aim at evaluating PT rate in the protonated water hexamer as a function of temperature, for both classical and quantum nuclei, by our QMC-driven LD simulations. The results will be compared together to refine our understanding of the interplay between NQE and thermal effects which governs the hydrated proton mobility at finite temperature.

6.4.1 Classical counting

We start with the simpler case of the classical $H^+(H_2O)_6$ cluster. When the ions, and more particularly the hydrated proton, are described within a classical picture, tracing their time evolution during the dynamics is trivial. Therefore, to evaluate the classical PT rate, we simply consider again the proton displacement d with respect to the midpoint of the $\overline{O_1O_2}$ distance. If this quantity changes its sign, a PT process might have happened because it indicates that the closest oxygen neighbor has changed. Similarly to the static properties, we are thus compelled to perform running averages to disregard the proton rattling between the two side oxygen atoms whose typical timescale is about 180 fs. By taking a time window of 200 fs, we ensure that the time evolution of the proton displacement d is representative of the different configurations (Zundel- or Eigen-like) that are adopted by the $H_{13}O_6^+$ ion during its dynamics. Nevertheless, this quantity is not self-sufficient since it does not provide any information about the geometry and the charge redistribution of the solvation shell when the hydrated proton jumps from one oxygen atom to another.

For this reason, we decided to evaluate, in a simple way, the value of the electrostatic potential on the left (O_1) and right (O_2) oxygen atoms. To that purpose, we attribute effective charges to the oxygen and the hydrogen atoms in the solvation shell, by analogy with the SPC model for water^[117]. Within this approach, the partial charges are: $Z_O = -0.82 e$ and $Z_H = -0.41 e$ since the oxygen has a greater electronegativity than the hydrogen. The same values are employed in this work. The electrostatic potential is evaluated by taking into account the effective interaction between one oxygen atom in the core and its neighboring particles in the solvation shell. Indeed, the core potential can be considered as a constant offset since it does not significantly modify the value of the overall electrostatic potential during the PT process.

6.4. Proton transfer in the protonated hexamer

The electrostatic potential is evaluated according to

$$V_{\text{left/right}} = \sum_{i \in \text{left/right shell}} \frac{Z_{\text{O}_{1/2}} Z_{\text{O}_i}}{|\mathbf{q}_{\text{O}_{1/2}} - \mathbf{q}_{\text{O}_i}|} + \sum_{j \in \text{left/right shell}} \frac{Z_{\text{O}_{1/2}} Z_{\text{H}_j}}{|\mathbf{q}_{\text{O}_{1/2}} - \mathbf{q}_{\text{H}_j}|}, \quad (6.6)$$

where $V_{\text{left/right}}$ is the calculated potential on the left/right part of the protonated hexamer. We have seen in the previous Sections that the global symmetry of the cluster is modified when it goes from a symmetric Zundel-type geometry to an asymmetric Eigen-type configuration. To visualize simply this structural change implying a charge redistribution inside the cluster during the PT process, we thus compute the electrostatic potential difference, which reads as

$$\Delta V = V_{\text{left}} - V_{\text{right}}, \quad (6.7)$$

where we employ the Eq. (6.6) to evaluate each component of the r.h.s. of the above equation. ΔV can be either positive or negative, depending on the location of the hydrated proton and the local arrangement of the solvation shell. Calculations based on the EVB-MD model argued that the collective electric field in the proton transfer direction is the appropriate coordinate to describe the creation and relaxation of these Zundel-like transition states.^[258] The molecular electrostatic potential thus constitutes a good theoretical descriptor of PT processes in water and aqueous systems, and has been recently used to characterize excited state acidity^[259].

We represent in Figure 6.18 the time evolution of the proton displacement d (black solid curve) and the scaled electrostatic potential difference $C\Delta V$ ⁵ (red solid curve) for a CMPC-LD simulation of the $\text{H}^+(\text{H}_2\text{O})_6$ cluster at room temperature ($T = 300$ K). The blue dashed line is a eye guide that defines the limit between positive and negative values of d and $C\Delta V$. The proton displacement d fluctuates around $d = 0.4$ a.u. and $d = -0.4$ a.u., with persistence periods of 4 – 5 ps. This confirms that, within a classical description, the protonated hexamer core is mainly in its Eigen-like configuration, where the hydrated proton is strongly localized. We notice that, at room temperature, the PT processes are not so frequent (~ 5 for a 20 ps trajectory), confirming the proton trapping predicted by the study of the static properties. We can count the number of proton hops occurring along the dynamics if the following criterion is satisfied: the sign of *both* the proton displacement d and the scaled electrostatic potential $C\Delta V$ changes during the proton jump. This method enables us to take into account those PT processes that imply a full charge rearrangement in the cluster and disregard shorter time scales processes, such as the proton rattling.

In Table 6.7, we report the number of PT that are directly counted from the QMC-driven CMPC-LD trajectories and we define the PT rate as $k_{\text{PT}} = N_{\text{PT}}/t_{\text{traj}}$. We can evaluate the characteristic time constant for the proton transfer $\tau_{\text{PT}} = 1/k_{\text{PT}}$ as a function of the temperature. Disregarding the low-temperature ($T = 200$ K) case, we confirm that thermal effects naturally increase the proton mobility since the proton jumps are more and more numerous at higher temperatures. This monotonic effect can be related to the Arrhenius law where the PT rate is given by $k_{\text{PT}} = A \exp(-\frac{E_a}{RT})$ where E_a is the activation energy, which is

⁵with C an applied constant to make the d and $C\Delta V$ signals almost superimpose.

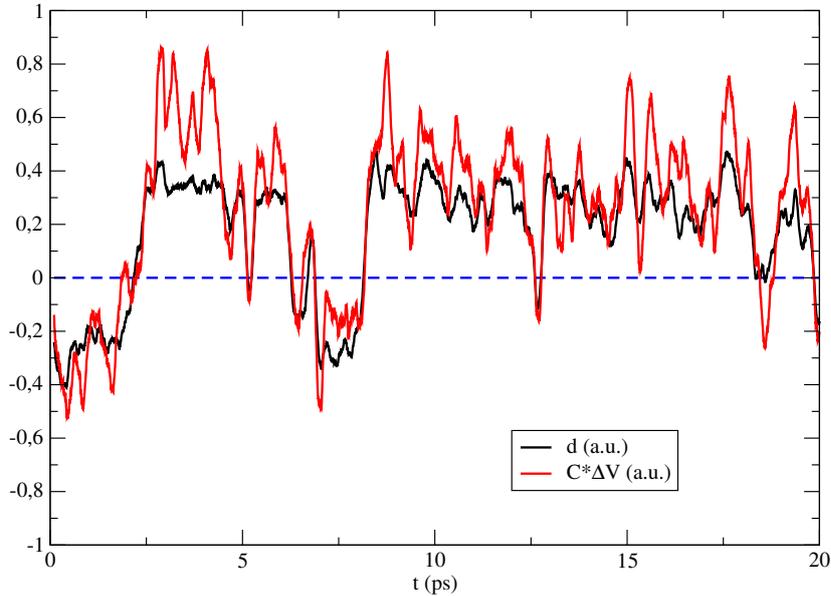


Figure 6.18 – Time evolution of the proton displacement and the electrostatic potential difference in the classical protonated water hexamer at room temperature ($T = 300$ K).

related to the PT static barrier in our case⁶. The $T = 200$ K result is special situation since the proton is normally trapped by the static barrier because it has not enough kinetic energy to spontaneously escape from its neighboring oxygen atom. However, as we have seen in Figure 6.3 and Table 6.4, the height of the PT barrier strongly depends on the oxygen-oxygen distance. At low temperature, the average distance $\langle O_1O_2 \rangle$ is close to the equilibrium one, enabling PT despite the lack of thermal (kinetic) energy of the hydrated proton. By the way, this low-temperature analysis is irrelevant without considering NQE which largely take over thermal effects, as we noticed for the Zundel complex (Chapter 4). Finally, the typical time constant for the proton relaxation is found to be around $\tau_{PT} = 1.9$ ps, being in agreement with the Car-Parrinello MD result of Tuckerman and his colleagues who found, disregarding the proton rattling, $\tau_{jump} = 1.68$ ps in liquid water, with the BLYP functional^[254].

T (K)	t_{traj} (ps)	N_{PT}	τ_{PT} (ps)
200	20.5	14	1.5
250	24.2	6	4.0
300	22.7	12	1.9
350	26.5	16	1.7
400	20.5	24	0.9

Table 6.7 – Estimated number of proton jumps and its corresponding time constant τ_{jump} as a function of the temperature, for the classical protonated hexamer

To put it in a nutshell, within a classical description, we confirm the Arrhenius behavior of the proton diffusion when one increases the cluster temperature. Nevertheless, although the obtained PT time constants have the correct order of magnitude, they are certainly not

⁶and A is a statistical prefactor.

physical since NQE deeply change the temperature-dependence of the PT physics in water clusters, and more generally in the bulk liquid.

6.4.2 Quantum counting: the instanton theory

We repeat here the same counting of the PT processes that occur inside the protonated hexamer during the dynamics, by analyzing the different QMC-driven PIOUD-LD trajectories at different temperatures, as we did in the last Subsection for classical particles. Nevertheless, within the PI framework, the ions and thus the hydrated proton, are strongly delocalized. Indeed, the calculated giration radius of the ring polymer^[260] is $R_G \sim 0.15 \text{ \AA}$ for the $T = 300 \text{ K}$ simulation, which makes impossible the analysis of the proton displacement d of a single bead⁷, as we did in the classical case. We could consider taking the *centroid* proton displacement to overcome this issue, but the result would be approximate.

A much more accurate solution is to apply the *instanton theory* to describe the dynamics of the hydrated proton, within the PI formalism. The instanton describes the dynamics of a chemical reaction at low temperature when tunnelling effects become dominant, and provides a generalization of the transition-state theory for anharmonic quantum systems^[261], such as the protonated water hexamer (Figure 6.3). Instanton theory has been very recently applied in a QMC framework^[262,263], where it has been demonstrated that QMC simulations efficiently recover the scaling of ground-state tunneling rates due to the existence of an instanton path, which always connects the reactant state with the product. Within the PI approach, the geometry of the beads at the transition state on the ring-polymer potential surface describes a finite-difference approximation to the instanton path. This is the minimal action trajectory, periodic in the quantum imaginary time $\tau = \beta\hbar$ ^[264], which links both reactants and products minima.

In our case, the instanton theory can thus be applied to estimate the quantum PT rate k_{PT}^{Q} by identifying the occurrence of the instantons^[265]. A quantitative way to find them during the dynamics is to compute the central proton ring-polymer fraction F , defined as $F = 100 \frac{N_{\text{beads}}}{P}$ ⁸ (N_{beads} is the number of beads making part of the instanton), that is situated on the left or on the right with respect to the midpoint of the central oxygen-oxygen distance. The instanton then corresponds to a configuration where $F = 50$ because the PT barrier separates two symmetric minima in this cluster. The time evolution of the ring-polymer fraction that is situated on the left part of oxygen-oxygen distance is represented by the black solid curve in Figure 6.19, for a simulation at ambient conditions ($T = 300 \text{ K}$). We also report on the same plot the scaled electrostatic potential difference $C\Delta + 50$ (red solid curve) for beads at the centroid, where we applied an offset to superimpose its signal with the instanton. We thus consider that a PT process occurs in the protonated hexamer when the ring-polymer fraction gets larger or lower than 50%, indicating that the proton crossed the static PT barrier, either by quantum tunneling or by a classical jump. We notice that the correlation between the instanton and the electrostatic potential is still present, although less neat than between the classical proton displacement and the classical electrostatic potential difference. Indeed, the quantum beats

⁷arbitrarily chosen since within PI, the equilibrium properties are trace-invariant.

⁸P is the total number of beads, defining the entire instanton path.

of the hydrated proton are not necessarily accompanied with a charge rearrangement of the solvation shell. The quantum tunnelling can be evinced by a delay between the instanton cross and the sign flip of the electrostatic potential. Thus, NQE significantly increase the PT rate k_{PT}^Q because they enable PT via this tunneling mechanism, at variance with the classical case.

Similarly to the classical case, we report on Table 6.8 the number of PT processes occurring inside the $\text{H}^+(\text{H}_2\text{O})_6$ core within the instanton theory at various temperatures ranging from 200 to 400 K. We also give the corresponding values of the PT time constant $\tau_{\text{PT}}^Q = t_{\text{traj}}/N_{\text{PT}}^Q$, which is inversely proportionnal to the quantum PT rate k_{PT}^Q . We first stress that, for trajectories of comparable length, the PT processes are more numerous in the quantum protonated hexamer than in the classical one, thanks to NQE which provide to the hydrated proton the necessary energy to easily go across the PT barrier. Moreover, the latter is probably smaller in the quantum case since the charge defect is naturally delocalized and more rapidly redistributed between the four solvating water molecules. Finally, we remark that the quantum PT rate k_{PT}^Q is not monotonic when one increases the temperature. Indeed, it is maximal near ambient conditions, with a typical timescale of $\tau_{\text{PT}}^Q = 1$ ps, which is consistent with the lifetime of the Zundel species^[17]. Such a non monotonic behavior manifestation of NQE has already been observed in DNA base pairs, where the H-bonds strength are counterintuitively less influenced by NQE at low temperature that around ambient conditions^[266].

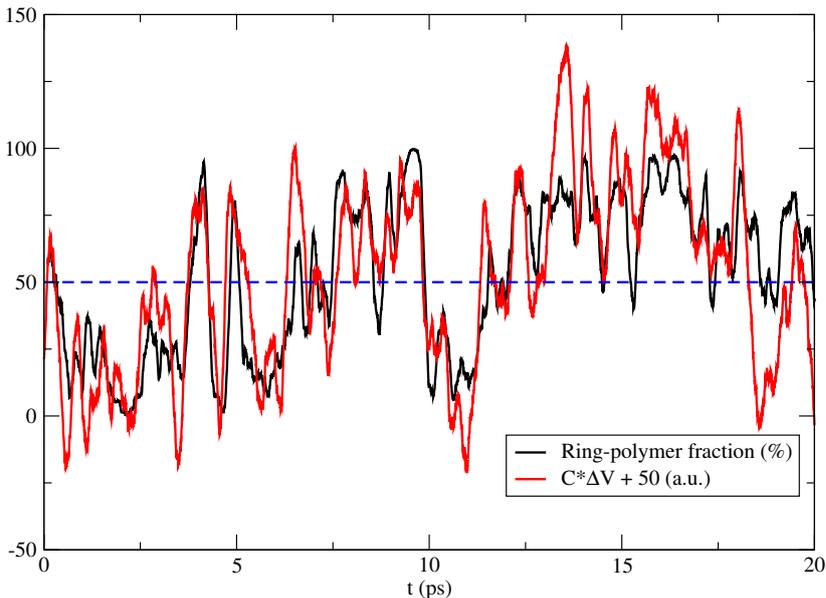


Figure 6.19 – Time evolution of the proportion of the ring-polymer located on the left of $\frac{d_{\text{O}_1\text{O}_2}}{2}$ and the electrostatic potential difference in the quantum protonated water hexamer at room temperature ($T = 300$ K).

Last but not least, we compare in Figure 6.20 the evolution for the classical k_{PT} and the quantum k_{PT}^Q PT rate of the protonated water hexamer as a function of the temperature. As already discussed, the classical PT rate increases with the temperature, in agreement with the Arrhenius law. Indeed, thermal effects increase the amplitude of the H-bond fluctuations and accentuate the special oxygen-oxygen pair dance, making the EZE mechanism more efficient

6.4. Proton transfer in the protonated hexamer

T (K)	t_{traj} (ps)	N_{PT}^Q	τ_{PT}^Q (ps)
200	23.1	12	1.9
250	23.9	16	1.5
300	21.9	21	1.0
350	18.5	14	1.3
400	23.0	18	1.3

Table 6.8 – Estimated number of proton jumps and its corresponding time constant τ_{jump} as a function of the temperature, for the quantum protonated hexamer

thus giving to the hydrated proton a greater mobility. On the other hand, the non monotonic temperature behavior of the quantum PT rate is surprising. Indeed, the proton mobility is maximal around ambient conditions, making the $T = 250\text{--}300$ K temperature interval a *sweet spot*. This optimal PT rate stems from the subtle interplay between NQE and thermal effects. Indeed, NQE allow the hydrated proton to cross the barrier, with controlled fluctuations and to shuttle from one neighboring oxygen atom to another in the (short) Zundel-like core. When the cluster temperature is too high, the amplitude of the H-bond fluctuations is so important that the proton is sometimes trapped into a largely distorted Eigen-like configuration, that is less cooperative than its Zundel counterpart for the proton diffusion. This behavior is strongly correlated to the average oxygen-oxygen distance that is ~ 0.02 Å smaller around ambient conditions than in the high-temperature limit. This spectacular result needs to be verified in the bulk liquid water, but is of paramount importance for the understanding of the fundamental mechanisms that governs the proton motion in water and aqueous systems.

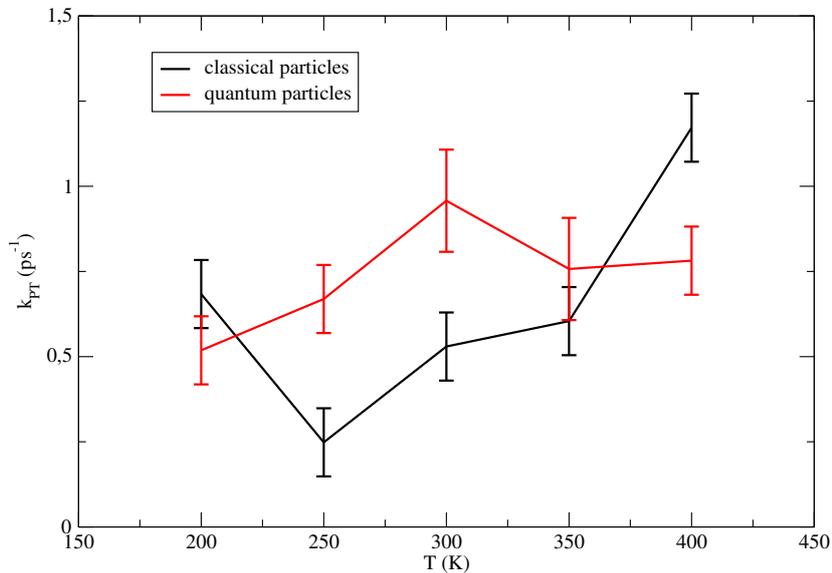


Figure 6.20 – Evolution of the classical k_{PT} and the quantum k_{PT}^Q PT rate of the protonated hexamer as a function of the temperature.

General conclusions and Perspectives

IN THIS THESIS, we have shed light on the microscopic mechanisms underlying the proton transfer (PT) process in water by taking the protonated water hexamer as a model. Thanks to the use of a novel Quantum Monte Carlo-driven Path Integral Langevin Dynamics framework, we have been able to investigate, at the highest level of accuracy, the impact of thermal and Nuclear Quantum Effects (NQE) on the hydrated proton localization in the protonated water dimer (Zundel cation) and hexamer. More specifically, we have demonstrated that NQE are significant for temperatures up to $T = 900$ K in the Zundel cation H_5O_2^+ , while they are still noticeable at $T = 400$ K in the protonated water hexamer $\text{H}_{13}\text{O}_6^+$. Moreover, NQE largely prevail over thermal effects for temperatures below the ambient conditions, allowing a faster diffusion of the proton by tunneling and Zero Point Energy (ZPE). We found that the PT characteristics are *non monotonic* with the temperature in the $\text{H}^+(\text{H}_2\text{O})_6$ complex, with an optimal PT rate around room temperature. This striking result raises many fundamental questions about the critical role played by the temperature and NQE on all chemical reactions where PT plays a crucial role, such as acid/base reactions occurring in complex chemical or biological systems.

Detecting such a subtle impact of the temperature on the hydrated proton mobility would not have been possible without the very high accuracy of the Quantum Monte Carlo (QMC) approach. This highly correlated method, based on a variational ansatz that has been introduced in Chapter 1, is a direct competitor of the most advanced quantum chemistry techniques such as the Møller-Plesset (MP2) and the Coupled Cluster (CCSD/CCSD(T)) approaches. Indeed, we have estimated the Variational Monte Carlo (VMC) accuracy be close to CCSD and the CCSD(T) methods for the evaluation of the Potential Energy Surface (PES) and the geometry of the Zundel cation (Chapter 4). Despite its stochastic nature that can complicate at the first sight its use for routine Molecular Dynamics (MD) simulations, the QMC approach exhibits the main advantage of a milder scaling with the system size, with respect to its quantum chemistry competitors. Moreover, the embarrassingly parallel formulation of its algorithms gives an additional appeal to this method, as its efficiency will keep increasing with the development of supercomputers. We therefore expect this electronic structure technique to be much more widely used in the next decades. At this stage of development, QMC methods remain however quite computationally demanding, even for the small neutral or charged water clusters studied in this thesis.

In this thesis, we have brought QMC to a wider level of applicability. Indeed, we have explored the possibility to incorporate the noisy QMC forces into a MD framework, via the Langevin Dynamics (LD) approach. Usually, the extension to the quantum nuclei is achieved by combining this MD with the Feynman Path Integral (PI) formalism. As a fundamental methodological development, we have paved the way to perform unbiased, accurate and efficient (PI)LD simulations of any physical or chemical system with either deterministic or stochastic (noisy) forces. First, the simulation temperature is controlled by an explicit correction of the total noise incorporated into the LD to fulfill the fluctuation-dissipation theorem.

Second, three original integration schemes, namely CMPC, PIMPC and PIOUD, based on the use of joint momentum-position coordinates have been devised to propagate efficiently the ions, using a large time interval between each nuclear iteration. To be more precise, the PIOUD algorithm provides an exact integration of the ring-polymer harmonic modes whose thermalization is ensured within one single operation, at variance with previous published algorithms. The remarkable stability of the PIOUD algorithm thus enables us to work with a large time step, leading to an efficient exploration of the phase space despite the shortness of our QMC-MD trajectories. We also introduced the bead-grouping approximation, a useful trick to make the computational cost of the quantum simulations almost equivalent to their classical counterparts. The aforementioned technical developments have been first validated on an analytical PES parametrized from CCSD(T) calculations and later with noisy QMC forces.

In Chapter 4, we provided a detailed study of the zero and finite temperature properties of the protonated water dimer H_5O_2^+ . This complex displays a centrosymmetric minimum with a proton evenly shared between its two neighboring oxygen atoms. When thermal fluctuations are important, the Zundel cation can be distorted as the oxygen-oxygen distance increases, leading to a partial localization of the excess proton close to an oxygen atom. This picture is particularly true within a classical description of the nuclei, as the Zero Point Energy (ZPE) and NQE tend to delocalize the proton by quantum tunneling. At ambient conditions, the PT physics in the Zundel ion is completely governed by NQE and one has to increase the cluster temperature up to $T = 900$ K to notice significant thermal effects. Although useful to validate our novel methodology, the Zundel ion does not constitute a realistic system to investigate PT in water and aqueous systems since solvation effects are missing. The natural continuation of this preliminary study is thus to incorporate 4 solvating water molecules around the Zundel core, defining the protonated water hexamer studied in Chapter 6.

Before studying such a large systems, simplifying the QMC wave function (WF) is needed. Indeed, based on the Jastrow-Slater (JSD) or Jastrow Antisymmetrized Geminal Power (JAGP) ansatz introduced in Chapter 1, the quality of the QMC WF is of paramount importance since it should be able to retain the major, if not the whole part, of the total correlation energy, at the lowest possible computational cost. Therefore, the JAGP WF, already used in Chapters 3 and 4 is revisited in Chapter 5, where contraction schemes to simplify the WF basis sets are applied. Testing the QMC WF on a simple system, namely the water dimer, allowed us to build a WF with a reasonable number of variational parameters and properly describe the bonding and non bonding interactions of the $(\text{H}_2\text{O})_2$ complex. We also noticed the potential significance of the spin-spin correlations on the overall binding energy of the bonding water dimer, suggesting that such interactions should also be taken into account to obtain a fully satisfactory description of the water dimer, and thus of the bulk liquid water.

Using the JAGP WF designed in Chapter 5, we studied the protonated water hexamer following the same route as in Chapter 4. We emphasize that the presence of solvating molecules tends to make the protonated hexamer PES more anharmonic than in the Zundel case (top panel of Figure 6.21). On the other hand, the nature of the minimum energy configuration is radically changed as the minimum of the $\text{H}_{13}\text{O}_6^+$ complex is *asymmetric*, with an excess

General conclusions and Perspectives

proton that is located closer to an oxygen atom than the other (bottom panel of Figure 6.21). Consequently, the PT static barriers, corresponding to a classical activation temperature of hundredths Kelvin degrees higher than in the Zundel ion, are at the origin of the observed partial trapping of the hydrated proton in the low-temperature limit. To describe this situation, we propose a 3-species model, involving two different Zundel-like complexes, namely the short (ambient) and the elongated (cold) Zundel. The former one, essential for an efficient proton diffusion in water, is favored around ambient conditions, due to the subtle combination of both thermal and quantum effects. When the cluster temperature is too large, the Eigen complex, with a fully localized proton, starts to take over and the proton only diffuses according to a Eigen-Zundel-Eigen (EZE) mechanism, with no more quantum tunneling.

The methodological advances presented in this thesis pave the way for simulations of larger and more realistic system exhibiting PT processes. Indeed, we restricted our study to finite size clusters in gas phase, whose behavior certainly differs from water in condensed phase. The natural extension of this thesis is to perform a QMC-driven PIOUS simulation of 64 or 128 water molecules in a cubic cell, with periodic boundary conditions. The radial distribution functions and the proton or water self-diffusion constants should be computed and compared to experimental data. Simulating larger systems, such as protein environments is, at this stage of QMC development, out of reach. One can however develop mixed approaches coupling the QMC technique with a cheaper computational method such as the DFT to generate MD trajectories with a limited number of configurations (for instance, 1 out of 10) evaluated at the QMC level. We also point out that the tools designed in this thesis do not allow the computation of dynamical quantities, such as the IR spectrums of the gas phase clusters. Therefore, an effort should be done in this direction, which will lead to a more direct comparison with the experimental data. This further development would be the final bridge between theory and experiment, allowing on the one hand a microscopic interpretation of the experimental results and on the other hand, more stringent tests and validations for new theories.

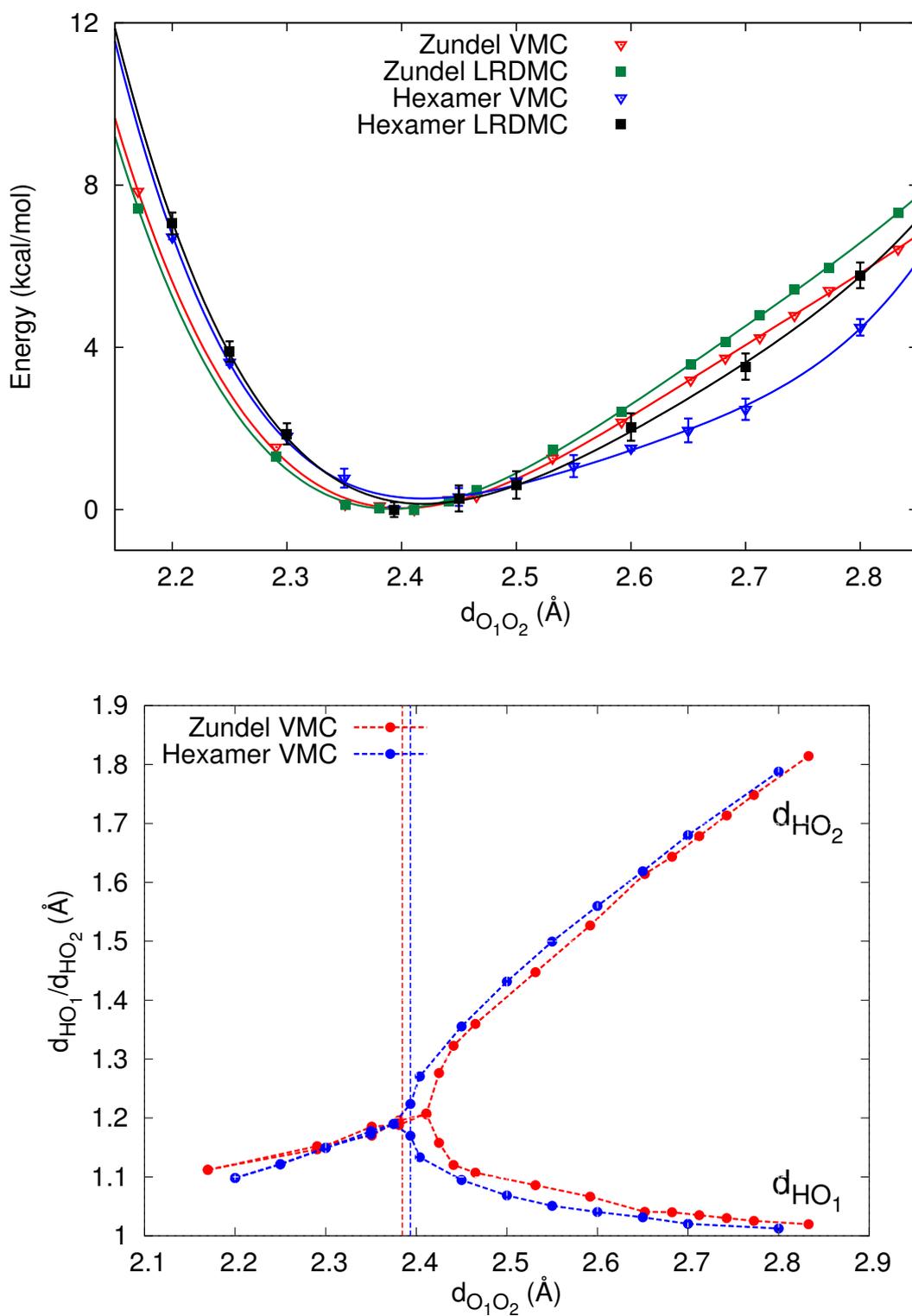


Figure 6.21 – Comparison of the protonated water dimer and hexamer Potential Energy Surface (left) and geometry (right) as a function of the (central) oxygen-oxygen distance $d_{O_1O_2}$. Vertical dashed lines indicate the corresponding equilibrium $d_{O_1O_2}$.

Appendix

Derivation of the MP2 energy

In this Appendix, we derive the explicit expression of the second order energy given in Eq. (1.19), starting from Eq. (1.18). Let us first take the WF of \hat{H}_0 to be normalized, *i.e.* $\langle i|i\rangle = 1$ and then choose the normalization of $|\Psi_i\rangle$ such that $\langle i|\Psi_i\rangle = 1$ (this choice is always possible in practice unless $|i\rangle$ and $|\Psi_i\rangle$ are orthogonal). Consequently, multiplying Eq. (1.18) by $\langle i|$, one obtains

$$\langle i|\Psi_i\rangle = \langle i|i\rangle + \lambda\langle i|\Psi_i^{(1)}\rangle + \lambda^2\langle i|\Psi_i^{(2)}\rangle = 1 \quad (\text{A.1})$$

The above equation holds for all values of λ ; therefore the coefficients of λ^n ; $n = 0, 1, 2$ on both sides must be equal. Hence, $\langle i|\Psi_i^{(n)}\rangle = 0$; $n = 0, 1, 2$. By substituting Eq. (1.18) into Eq. (1.16) and equating coefficients of λ^n ; $n = 0, 1, 2$, we find:

$$\begin{aligned} n = 0 : \hat{H}_0|i\rangle &= E_i^{(0)}|i\rangle \\ n = 1 : \hat{H}_0|\Psi_i^{(1)}\rangle + \hat{V}|i\rangle &= E_i^{(0)}|\Psi_i^{(1)}\rangle + E_i^{(1)}|i\rangle \\ n = 2 : \hat{H}_0|\Psi_i^{(2)}\rangle + \hat{V}|\Psi_i^{(1)}\rangle &= E_i^{(0)}|\Psi_i^{(2)}\rangle + E_i^{(1)}|\Psi_i^{(1)}\rangle + E_i^{(2)}|i\rangle. \end{aligned} \quad (\text{A.2})$$

We then multiply each line of Eq. (A.2) by $\langle i|$ and using the orthogonality relation $\langle i|\Psi_i^{(n)}\rangle = 0$; $n = 0, 1, 2$, we obtain the following expressions for the energies until the second order:

$$\begin{aligned} E_i^{(0)} &= \langle i|\hat{H}_0|i\rangle \\ E_i^{(1)} &= \langle i|\hat{V}|i\rangle \\ E_i^{(2)} &= \langle i|\hat{V}|\Psi_i^{(1)}\rangle. \end{aligned} \quad (\text{A.3})$$

All that remains is to solve the set of Eqs. (A.2) for $|\Psi_i^{(n)}\rangle$; $n = 0, 1, 2$ and then determine the n th-order energy using (A.3). We recall the equation which determines the first order WF $|\Psi_i^{(1)}\rangle$ extracted from Eq. (A.2) a few lines above:

$$\hat{H}_0|\Psi_i^{(1)}\rangle + \hat{V}|i\rangle = E_i^{(0)}|\Psi_i^{(1)}\rangle + E_i^{(1)}|i\rangle. \quad (\text{A.4})$$

The previous equation can be rewritten as

$$(E_i^{(0)} - \hat{H}_0)|\Psi_i^{(1)}\rangle = (\hat{V} - E_i^{(1)})|i\rangle = (\hat{V} - \langle i|\hat{V}|i\rangle)|i\rangle. \quad (\text{A.5})$$

One way to solve this integro-differential equation is to expand the first order WF $|\Psi_i^{(1)}\rangle$ in terms of eigenfunctions of \hat{H}_0 which are taken to be complete

$$|\Psi_i^{(1)}\rangle = \sum_n c_n^{(1)}|n\rangle. \quad (\text{A.6})$$

The eigenfunctions of \hat{H}_0 are orthonormal by definition, thus multiplying Eq. (A.6) by $\langle n|$, one obtains

$$c_n^{(1)} = \langle n|\Psi_i^{(1)}\rangle. \quad (\text{A.7})$$

Using the fact that $\langle i|\Psi_i^{(1)}\rangle = 0$; $n = 0, 1, 2$, it is clear that $c_i^{(1)} = 0$, so we can write

$$|\Psi_i^{(1)}\rangle = \sum_{n \neq i} |n\rangle \langle n|\Psi_i^{(1)}\rangle. \quad (\text{A.8})$$

Now, let us multiply Eq. (A.5) by $\langle n|$ and still using the orthogonality of the zeroth order WF, we have

$$(E_i^{(0)} - E_n^{(0)})\langle n|\Psi_i^{(1)}\rangle = \langle n|\hat{V}|i\rangle. \quad (\text{A.9})$$

Using the expression (A.8) in Eq. (A.3) for the second order energy, one obtains

$$E_i^{(2)} = \langle i|\hat{V}|\Psi_i^{(1)}\rangle = \sum_{n \neq i} \langle i|\hat{V}|n\rangle \langle n|\Psi_i^{(1)}\rangle \quad (\text{A.10})$$

and hence, combining with Eq. (A.9), we finally obtain the desired explicit expression for the second order energy in the framework of the perturbation theory

$$E_i^{(2)} = \sum_{n \neq i} \frac{\langle i|\hat{V}|n\rangle \langle n|\hat{V}|i\rangle}{E_i^{(0)} - E_n^{(0)}} = \sum_{n \neq i} \frac{|\langle i|\hat{V}|n\rangle|^2}{E_i^{(0)} - E_n^{(0)}}. \quad (\text{A.11})$$

Higher order terms can be obtained following the same procedure than described above, but pushing the Taylor expansion of Eq. (1.18) to the order we want to stop.

Calculation of excitation amplitudes within CC approach

In this Appendix, we exploit the CC WF function ansatz to obtain an explicit expression of the CC WF and the amplitudes of first-, second- or higher-order excitations. Let us start with a simple Taylor expansion of the $e^{\hat{T}}$

$$\begin{aligned} e^{\hat{T}} &= 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \dots \\ &= 1 + \hat{C}_1 + \hat{C}_2 + \dots \end{aligned} \quad (\text{B.1})$$

We have rearranged in the above equation the operators in terms of excitation levels: $\hat{C}_1 = \hat{T}_1$ generates the single excitations whereas $\hat{C}_2 = \hat{T}_2 + \frac{\hat{T}_1^2}{2}$ generates double excitations and so on. Hence, double excitations can be generated in two ways: by a simultaneous excitation of two electrons (\hat{T}_2) or by exciting simply two independent single electrons ($\frac{\hat{T}_1^2}{2}$).

The whole CC WF function can thus be written as

$$|\Psi_{\text{CC}}\rangle = |\Phi_0\rangle + \sum_i^{N_{\text{orb}}} \sum_a^{N_{\text{vir}}} c_i^a |\Phi_i^a\rangle + \sum_{i<j}^{N_{\text{orb}}} \sum_{a<b}^{N_{\text{vir}}} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle + \dots \quad (\text{B.2})$$

with coefficients related to the cluster amplitudes by

$$\begin{aligned} c_i^a &= t_i^a, \\ c_{ij}^{ab} &= t_{ij}^{ab} + t_i^a \star t_j^b, \\ c_{ijk}^{abc} &= t_{ijk}^{abc} + t_i^a \star t_{jk}^{bc} + t_i^a \star t_j^b \star t_k^c \end{aligned} \quad (\text{B.3})$$

and so on. In the Eq. (B.3), \star implies an antisymmetric product making the resulting coefficient properly antisymmetric with respect to any exchange of two occupied orbitals or two virtual orbitals (i.e. $t_i^a \star t_j^b = t_i^a t_j^b - t_j^a t_i^b$).

Consequently, it is clear from Eq. (B.3) that the coefficients of triple excitations c_{ijk}^{abc} are fully determined by only products of single- and double-excitation amplitudes t_i^a and t_{jk}^{bc} , and similarly for the coefficients of all higher-level excitations. Presumably, higher-order excitation amplitudes are smaller than the double-excitation amplitudes t_{ij}^{ab} and can be reasonably neglected, as it is done in CCSD or CCSD(T) approaches.

The CC WF contains all the excited determinants and one can easily choose at which order the WF truncation must be done. According to Slater's rule, third- and higher-order excitations of the HF WF $|\Phi_0\rangle$ do not directly contribute to the correlation between electrons, which justifies the very frequent choice to limit the expansion until the second-order. In this thesis, we also apply this choice by performing CCSD or CCSD(T) calculations.

Wave function optimization methods

In this Appendix, we detail the two main techniques we used in this thesis to optimize the QMC WF of the water clusters we studied, namely the Stochastic Reconfiguration (SR)^[85] and the Stochastic Reconfiguration with Hessian Accelerator (SRH)^[86].

C.1 Stochastic Reconfiguration

The SR technique is a method that takes benefit from the direct knowledge of the trial QMC WF Ψ_T to converge as fast as possible. Considering that $\Psi_T(\boldsymbol{\lambda}^0)$ depends on a set of p variational parameters $\{\lambda^k\}_{k=1,\dots,p}$ and by applying an elementary perturbation $\delta\lambda_k$ to each parameter, the variational parameters then read as

$$\lambda_k = \lambda_k^0 + \delta\lambda_k. \quad (\text{C.1})$$

Thus, the corresponding perturbed WF can be written

$$\Psi'_T(\boldsymbol{\lambda}) = \Psi_T(\boldsymbol{\lambda}_0) + \sum_{k=1}^p \delta\lambda_k \frac{\partial}{\partial \lambda_k} \Psi_T(\boldsymbol{\lambda}_0). \quad (\text{C.2})$$

Let us introduce local operators defined on each electronic configuration $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ as logarithmic derivatives with respect to the variational parameters

$$\hat{O}^k = \frac{\partial}{\partial \lambda_k} \ln \Psi_T(\mathbf{R}), \quad (\text{C.3})$$

where we set $\hat{O}^0 = 1$ for convenience. Using Eq. (C.3) in Eq. (C.2), one obtains a more compact form for Ψ'_T

$$\Psi'_T = \sum_{k=0}^p \delta\lambda_k \hat{O}^k \Psi_T. \quad (\text{C.4})$$

For a generic trial WF Ψ_T , it is possible to obtain iteratively an updated WF closer to the ground state of the system by simply applying the shifted Hamiltonian operator $(\Lambda - \hat{H})$ that gives

$$|\Psi_T^P\rangle = (\Lambda - \hat{H})|\Psi_T\rangle, \quad (\text{C.5})$$

with Λ large enough (in practice, we work in the limit $\Lambda \rightarrow +\infty$) and $|\Psi_T^P\rangle$ the projected WF.

C.2. Stochastic Reconfiguration with Hessian accelerator

To make the perturbed WF Ψ'_T as close as possible to the projected WF Ψ_T^P in the subspace generated by the vectors $\hat{O}_k|\Psi_T\rangle$, we impose the following SR conditions

$$\langle \Psi_T | \hat{O}^k | \Psi'_T \rangle = \langle \Psi_T | \hat{O}^k | \Psi_T^P \rangle \quad \text{for } k = 0, \dots, p. \quad (\text{C.6})$$

Eq. (C.6) can be rewritten in a linear system

$$\sum_{l=1}^p \delta\lambda_l S_{lk} = \frac{1}{2} f_k, \quad (\text{C.7})$$

where $S_{lk} = \langle (\hat{O}^l - \langle \hat{O}^l \rangle) (\hat{O}^k - \langle \hat{O}^k \rangle) \rangle$ defines the *covariance matrix* $\mathbf{S} \equiv \{S_{lk}\}$. $f_k = 2(\langle \hat{O}^k \rangle \langle \hat{H} \rangle - \langle \hat{O}^k \hat{H} \rangle)$ is the generalized QMC force introduced in Eq. (1.70) of the manuscript. By simple inversion of the covariance matrix \mathbf{S} , the linear system in Eq. (C.7) can be rewritten

$$\delta\boldsymbol{\lambda} = \frac{1}{2} \mathbf{S}^{-1} \mathbf{f}, \quad (\text{C.8})$$

where the covariance matrix \mathbf{S} and the forces acting on each parameter \mathbf{f} are stochastically evaluated within a QMC framework. Once the solution of the linear problem described in Eq. (C.8) is found, one can update the variational parameters according to

$$\boldsymbol{\lambda}' = \boldsymbol{\lambda} + \delta t_{\text{par}} \delta\boldsymbol{\lambda} = \boldsymbol{\lambda} + \frac{\delta t_{\text{par}}}{2} \mathbf{S}^{-1} \mathbf{f}, \quad (\text{C.9})$$

where δt_{par} is a parameter modulating the acceleration of electronic parameters for a given SR step. Its value should be set up small enough to ensure a minimization of the electronic energy expectation value. Indeed, in that case, it can be easily shown the energy variation ΔE between two SR steps is negative by taking the limit

$$\Delta E = -\delta t_{\text{par}} \sum_{k=1}^p f_k^2 + \mathcal{O}(\delta t_{\text{par}}^2). \quad (\text{C.10})$$

Consequently, the SR method converges to the lowest energy when the electronic forces f_k disappear. To obtain a stable minimization procedure for WF containing a rather large number of electronic parameters ($\sim 2 - 8.10^3$ in this thesis), the acceleration of electronic parameters δt_{par} has to be dramatically reduced, making the WF optimization procedure computationally cumbersome and inefficient. To solve this issue, one can use an improvement of the SR method: the Stochastic Reconfiguration with Hessian accelerator (SRH).

C.2 Stochastic Reconfiguration with Hessian accelerator

As mentioned above, SRH technique is used to fasten the convergence of the WF optimization by using the information contained in the second order derivatives of the electronic energy.

Using this minimization procedure, the variation of the energy for a single change of the

Appendix C. Wave function optimization methods

variational parameters is very similar to the formula obtained within SR approach with an additional second order term

$$\Delta E = - \sum_{k=1}^p \delta \lambda_k f_k + \frac{1}{2} \sum_{k,k'}^p \delta \lambda_k \delta \lambda_{k'} B_{kk'} \quad (\text{C.11})$$

where $B_{kk'} = \frac{\partial^2 E}{\partial \lambda_k \partial \lambda_{k'}}$ defines the Hessian matrix of the energy with respect to the electronic parameters $\mathbf{B} \equiv \{B_{kk'}\}$. By imposing $\frac{\partial E}{\partial \lambda_k}$ for each $k = 1, \dots, p$, we approximate the value of the minimum energy E substituting all the electronic parameters $\boldsymbol{\lambda}$ by their relative variation $\delta \boldsymbol{\lambda}$. Within this approximation, the linear system equivalent to Eq. (C.8) for SRH is thus

$$\delta \boldsymbol{\lambda} = \mathbf{B}^{-1} \mathbf{f}. \quad (\text{C.12})$$

In other words, the knowledge of the Hessian matrix \mathbf{B} helps to update the electronic parameters $\boldsymbol{\lambda}$ in the "right" direction, that is to say, closer to their equilibrium value which minimizes the energy of the QMC trial WF. The SRH minimization procedure is much more stable than SR, especially during the optimization of the variational parameters defining the exponents of the atomic or molecular orbitals of the system. This allows us to use larger values of the parameter driving the acceleration of the electronic parameters, δt_{par} . However, the Hessian matrix \mathbf{B} is more delicate to compute without bias and has required some technical developments, such as Algorithmic Differentiation (AD)^[107].

Noise correction and quantum integration scheme

In this Appendix, we provide the useful formulæ to understand the algorithmic developments introduced in the Chapter 3. In particular, we provide explicit expressions for the noise correlators employed in the CMPC algorithm. We also explain how to correct the intrinsic QMC noise to keep the dynamics free of biases. Finally, we detail the mathematical derivation of the PIOUS algorithm to perform PILD simulations, with quantum particles.

D.1 Noise correlators in the CMPC algorithm

We report the variance-covariance properties which define the time integrated noises $\tilde{\boldsymbol{\eta}}$ and $\tilde{\tilde{\boldsymbol{\eta}}}$ for momenta and positions, respectively. Their variance-covariance can be computed by imposing the FDT (2.32) to be fulfilled, under the hypothesis that the $\boldsymbol{\alpha}$ matrix is \mathbf{q} -time independent (as we have done in Chapter 3, Subsection 3.1.1), and by exploiting that $[\boldsymbol{\alpha}, \boldsymbol{\gamma}] = 0$. One finds:

$$\begin{aligned}
 \tilde{\boldsymbol{\alpha}}_{11} &= \langle \tilde{\boldsymbol{\eta}}^T \tilde{\boldsymbol{\eta}} \rangle = k_B T \gamma^2 \coth\left(\gamma \frac{\delta t}{2}\right), \\
 \tilde{\boldsymbol{\alpha}}_{22} &= \langle \tilde{\tilde{\boldsymbol{\eta}}}^T \tilde{\tilde{\boldsymbol{\eta}}} \rangle = k_B T (2\boldsymbol{\Theta} - \boldsymbol{\Gamma}^2) \boldsymbol{\Theta}^{-2}, \\
 \tilde{\boldsymbol{\alpha}}_{12} &= \langle \tilde{\boldsymbol{\eta}}^T \tilde{\tilde{\boldsymbol{\eta}}} \rangle = \tilde{\boldsymbol{\alpha}}_{21} = \langle \tilde{\tilde{\boldsymbol{\eta}}}^T \tilde{\boldsymbol{\eta}} \rangle = k_B T \boldsymbol{\gamma} \boldsymbol{\Gamma} \boldsymbol{\Theta}^{-1},
 \end{aligned} \tag{D.1}$$

while the mean of $\tilde{\boldsymbol{\eta}}$ and $\tilde{\tilde{\boldsymbol{\eta}}}$ is zero.

D.2 Quantum Monte Carlo noise correction in the CMPC algorithm

We provide the noise correction which needs to be applied if one wants to include the QMC noise into the Langevin dynamics without inducing a bias on the final target temperature. Analogously to the equation (3.31), the integrated noise can be written

$$\begin{aligned}
 \tilde{\eta}_i &= \tilde{\eta}_i^{\text{ext}} + \tilde{\eta}_i^{\text{QMC}}, \\
 \tilde{\tilde{\eta}}_i &= \tilde{\tilde{\eta}}_i^{\text{ext}} + \tilde{\tilde{\eta}}_i^{\text{QMC}}.
 \end{aligned} \tag{D.2}$$

Therefore, still assuming that the external noise is independent from the QMC noise, we obtain very similar relations to Eq. (3.33):

$$\begin{aligned}
\langle\langle (\tilde{\boldsymbol{\eta}}^{\text{ext}})^{\text{T}} \tilde{\boldsymbol{\eta}}^{\text{ext}} \rangle\rangle &= \tilde{\boldsymbol{\alpha}}_{11} - \boldsymbol{\alpha}^{\text{QMC}}, \\
\langle\langle (\tilde{\tilde{\boldsymbol{\eta}}}^{\text{ext}})^{\text{T}} \tilde{\tilde{\boldsymbol{\eta}}}^{\text{ext}} \rangle\rangle &= \tilde{\boldsymbol{\alpha}}_{22} - \boldsymbol{\alpha}^{\text{QMC}}, \\
\langle\langle (\tilde{\boldsymbol{\eta}}^{\text{ext}})^{\text{T}} \tilde{\tilde{\boldsymbol{\eta}}}^{\text{ext}} \rangle\rangle &= \tilde{\boldsymbol{\alpha}}_{12} - \boldsymbol{\alpha}^{\text{QMC}},
\end{aligned} \tag{D.3}$$

where the explicit expressions of the $\tilde{\boldsymbol{\alpha}}$ matrix components have been given in the main text in Eq. (D.1). By simply computing the square root of this 2×2 matrix which is positive definite by construction, we are able to evaluate explicitly the exact integrated external noise to propagate the dynamics.

D.3 Quantum integration scheme

We detail in this section the mathematical derivation of the PIOUD integrator described in Chapter 3, Subsection 3.1.3 and the explicit formulas of the $\mathbf{\Lambda}$, $\mathbf{\Gamma}$ and $\mathbf{\Theta}$ matrices. For the following algebra, it is useful to evaluate the inverse of the matrix $\hat{\boldsymbol{\gamma}}$

$$\hat{\boldsymbol{\gamma}}^{-1} = \begin{pmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{K}^{-1} & \mathbf{K}^{-1}\boldsymbol{\gamma} \end{pmatrix}, \tag{D.4}$$

where we replaced $\boldsymbol{\gamma}^{\text{harm}}$ of Eq. (3.9) by $\boldsymbol{\gamma}$, for the sake of readability. In order to solve the differential system in Eq. (3.4) for a generic time step δt in the quantum case, we need to exponentiate the matrix $\hat{\boldsymbol{\gamma}}$. Using the fundamental assumption $[\mathbf{K}, \boldsymbol{\gamma}] = 0$ previously justified in the manuscript, we consider each common eigenvector of \mathbf{K} and $\boldsymbol{\gamma}$ that can correspond to a joint momentum and coordinate mode. In this two-fold basis, the matrix $\hat{\boldsymbol{\gamma}}$ is a simple 2×2 block matrix where as K and γ are just numbers. The block matrix $\hat{\boldsymbol{\gamma}}$ can be more conveniently rewritten in terms of the Pauli matrices $\boldsymbol{\sigma}_x, \boldsymbol{\sigma}_y, \boldsymbol{\sigma}_z$:

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} &= \frac{\gamma}{2}\mathbf{I} + \mathbf{x} \\
\mathbf{x} &= \begin{pmatrix} \frac{\gamma}{2} & K \\ -1 & -\frac{\gamma}{2} \end{pmatrix} \\
&= \frac{K-1}{2}\boldsymbol{\sigma}_x + i\frac{K+1}{2}\boldsymbol{\sigma}_y + \frac{\gamma}{2}\boldsymbol{\sigma}_z.
\end{aligned} \tag{D.5}$$

Then, the exponentiation can be straightforwardly obtained, by using standard Pauli matrices algebra

$$e^{\hat{\boldsymbol{\gamma}}\delta t} = e^{\frac{\gamma\delta t}{2}} \begin{cases} \cosh(|x|\delta t)\mathbf{I} + \frac{\mathbf{x}}{|x|}\sinh(|x|\delta t) & \text{for } \gamma^2 \geq 4K \\ \cos(|x|\delta t)\mathbf{I} + \frac{\mathbf{x}}{|x|}\sin(|x|\delta t) & \text{for } \gamma^2 < 4K \end{cases} \tag{D.6}$$

where $|x| = \sqrt{|\gamma^2/4 - K|}$. Recombining Eq. (D.4) and Eq. (D.6) one obtains:

$$\hat{\boldsymbol{\gamma}}^{-1}(\mathbf{I} - e^{-\hat{\boldsymbol{\gamma}}\delta t}) = \begin{pmatrix} e^{-\frac{\gamma\delta t}{2}} \frac{\sinh(|x|\delta t)}{|x|} & -(1 - e^{-\frac{\gamma\delta t}{2}} \cosh(|x|\delta t)) + \gamma e^{-\frac{\gamma\delta t}{2}} \frac{\sinh(|x|\delta t)}{2|x|} \\ 1 - e^{-\frac{\gamma\delta t}{2}} \frac{\cosh(|x|\delta t)}{K} - \gamma e^{-\frac{\gamma\delta t}{2}} \frac{\sinh(|x|\delta t)}{2K|x|} & \gamma(1 - e^{-\frac{\gamma\delta t}{2}} \frac{\cosh(|x|\delta t)}{K}) + (1 - \frac{\gamma^2}{2K})e^{-\frac{\gamma\delta t}{2}} \frac{\sinh(|x|\delta t)}{|x|} \end{pmatrix} \tag{D.7}$$

Appendix D. Noise correction and quantum integration scheme

for $\gamma^2 > 4K$, while for $\gamma^2 \leq 4K$:

$$\hat{\gamma}^{-1}(\mathbf{I} - e^{-\hat{\gamma}\delta t}) = \begin{pmatrix} e^{-\frac{\gamma\delta t}{2}} \frac{\sin(|x|\delta t)}{|x|} & -(1 - e^{-\frac{\gamma\delta t}{2}} \cos(|x|\delta t)) + \frac{\gamma e^{-\frac{\gamma\delta t}{2}} \sin(|x|\delta t)}{2|x|} \\ \frac{1 - e^{-\frac{\gamma\delta t}{2}} \cos(|x|\delta t)}{K} - \frac{\gamma e^{-\frac{\gamma\delta t}{2}} \sin(|x|\delta t)}{2K|x|} & \frac{\gamma(1 - e^{-\frac{\gamma\delta t}{2}} \cos(|x|\delta t))}{K} + (1 - \frac{\gamma^2}{2K}) \frac{e^{-\frac{\gamma\delta t}{2}} \sin(|x|\delta t)}{|x|} \end{pmatrix}. \quad (\text{D.8})$$

Finally, for the case $\gamma^2 > 4K$, we obtain:

$$\begin{aligned} \mathbf{\Gamma} &= e^{-\frac{\gamma}{2}\delta t} \frac{\sinh(|x|\delta t)}{|x|} \\ \mathbf{\Theta} &= \frac{1 - e^{-\frac{\gamma\delta t}{2}} \cosh(|x|\delta t)}{K} - \frac{\gamma e^{-\frac{\gamma\delta t}{2}} \sinh(|x|\delta t)}{2K|x|} \\ \mathbf{\Lambda} &= e^{-\frac{\gamma\delta t}{2}} \begin{pmatrix} \cosh(|x|\delta t) - \frac{\gamma}{2} \frac{\sinh(|x|\delta t)}{|x|} & -K \frac{\sinh(|x|\delta t)}{|x|} \\ \frac{\sinh(|x|\delta t)}{|x|} & \cosh(|x|\delta t) + \frac{\gamma \sinh(|x|\delta t)}{2|x|} \end{pmatrix}, \end{aligned} \quad (\text{D.9})$$

while for $\gamma^2 \leq 4K$, we have:

$$\begin{aligned} \mathbf{\Gamma} &= e^{-\frac{\gamma}{2}\delta t} \frac{\sin(|x|\delta t)}{|x|} \\ \mathbf{\Theta} &= \frac{1 - e^{-\frac{\gamma\delta t}{2}} \cos(|x|\delta t)}{K} - \frac{\gamma e^{-\frac{\gamma\delta t}{2}} \sin(|x|\delta t)}{2K|x|} \\ \mathbf{\Lambda} &= e^{-\frac{\gamma\delta t}{2}} \begin{pmatrix} \cos(|x|\delta t) - \frac{\gamma}{2} \frac{\sin(|x|\delta t)}{|x|} & -K \frac{\sin(|x|\delta t)}{|x|} \\ \frac{\sin(|x|\delta t)}{|x|} & \cos(|x|\delta t) + \frac{\gamma \sin(|x|\delta t)}{2|x|} \end{pmatrix}. \end{aligned} \quad (\text{D.10})$$

To evaluate the time-integrated noise covariance matrix $\tilde{\boldsymbol{\alpha}}$ in the quantum case, defined by Eq. (3.16), namely $\tilde{\boldsymbol{\alpha}} \equiv \langle \mathbf{E}_{\text{int}}^T \mathbf{E}_{\text{int}} \rangle$, it is useful to introduce the following integral:

$$\Gamma_z = \int_{-\delta t}^0 dt e^{zt} = \frac{1 - e^{-z\delta t}}{z}. \quad (\text{D.11})$$

By using Eq. (D.6), it is straightforward to carry out the integrals to obtain

$$\begin{aligned} \tilde{\boldsymbol{\alpha}} &= \frac{\Gamma_{\gamma+2|x|} + \Gamma_{\gamma-2|x|} + 2\Gamma_\gamma}{4} \hat{\boldsymbol{\alpha}} \\ &+ \frac{\mathbf{x}\hat{\boldsymbol{\alpha}}\mathbf{x}^\dagger}{|x|^2} \frac{\Gamma_{\gamma+2|x|} + \Gamma_{\gamma-2|x|} - 2\Gamma_\gamma}{4} \\ &+ \frac{\hat{\boldsymbol{\alpha}}\mathbf{x}^\dagger + \mathbf{x}\hat{\boldsymbol{\alpha}}}{|x|} \frac{\Gamma_{\gamma+2|x|} - \Gamma_{\gamma-2|x|}}{4} \end{aligned} \quad (\text{D.12})$$

D.3. Quantum integration scheme

for $\gamma^2 \geq 4K$, whereas for $\gamma^2 \leq 4K$ we obtain a very similar formula where $|x|$ is replaced by $i|x|$:

$$\begin{aligned}
\tilde{\alpha} &= \frac{\Gamma_{\gamma+2i|x|} + \Gamma_{\gamma-2i|x|} + 2\Gamma_{\gamma}}{4} \hat{\alpha} \\
&- \frac{\mathbf{x}\hat{\alpha}\mathbf{x}^\dagger \Gamma_{\gamma+2i|x|} + \Gamma_{\gamma-2i|x|} - 2\Gamma_{\gamma}}{|x|^2} \\
&- i \frac{\hat{\alpha}\mathbf{x}^\dagger + \mathbf{x}\hat{\alpha}}{|x|} \frac{\Gamma_{\gamma+2i|x|} - \Gamma_{\gamma-2i|x|}}{4}.
\end{aligned} \tag{D.13}$$

$\hat{\alpha}$ has been defined in Eq. (3.16). In order to conclude the analytic derivation of the quantum integration scheme, we also give the explicit expression of the two matrix products appearing in the Eq. (D.12) and Eq. (D.13):

$$\begin{aligned}
\mathbf{x}\hat{\alpha}\mathbf{x}^\dagger &= \begin{pmatrix} \frac{\gamma^2\alpha}{4} & -\frac{\gamma\alpha}{2} \\ -\frac{\gamma\alpha}{2} & \alpha \end{pmatrix}, \\
\hat{\alpha}\mathbf{x}^\dagger + \mathbf{x}\hat{\alpha} &= \begin{pmatrix} \gamma & -\alpha \\ -\alpha & 0 \end{pmatrix}.
\end{aligned} \tag{D.14}$$

Protonated hexamer geometries

In this Appendix, we provide the obtained geometries for the protonated hexamer at the VMC level for various oxygen-oxygen distances $d_{\text{O}_1\text{O}_2}$ (see Figure 6.2). Results are given in atomic units.

E.1 Minimum geometry

O₁ 0.0507882315187001762 0.000709729258697364292 0.285378811112213804
H₁ -0.588908309572161781 -1.50916427426181143 1.14553358618697887
H₂ -0.594542041413769762 1.50841288801772544 1.13644136895695169
H⁺ 2.24518427383086694 0.00669502090316738581 0.0200749585936824959
O₃ -1.65980132332560193 4.26597184961100861 2.47433767465341825
H -3.01280788761494556 5.25378600003923157 1.79611713827810826
H -1.53311715628076772 4.73830248178763469 4.21400571902787124
O₄ -1.65103670216086962 -4.27103150563844292 2.47327243060715496
H -3.01959943643830719 -5.23873631958746167 1.79456467107910389
H -1.52586019854693311 -4.75056777770878380 4.20775945789407313
O₂ 4.54182071095368922 0.00114072825318341556 -0.251859090320561640
H₃ 5.20104667377932106 1.50615576559710007 -1.09483876564196247
H₄ 5.20693218110539213 -1.48927734882025198 -1.10970177150247218
O₅ 6.32015005426037391 -4.28087841192594709 -2.43540083233940585
H 6.24018546662413698 -4.79741171253304444 -4.16604952053338096
H 7.67145094281254813 -5.23519780246613120 -1.69871359284725520
O₆ 6.31384117894622943 4.29130875981402671 -2.42770929861173901
H 7.70020569151150269 5.22262043109420127 -1.73976512929067662
H 6.20794585000182231 4.78378826156620907 -4.16197743530833808

E.2 Other geometries

$d_{\text{O}_1\text{O}_2} = 2.2 \text{ \AA}$

O₁ 0.000000000000000000 0.000000000000000000 0.000000000000000000
H₁ -0.743586267341124318 -1.48775917683513570 0.805360324563675789
H₂ -0.734532372670881450 1.49766666964075545 0.794865339656707248
H⁺ 2.07703452489563567 -0.00782795162408958696 0.00254865012463290942
O₃ -1.78090480055100198 4.23306705176871034 2.30522925399525302
H -3.16370670078231386 5.25062598844431072 1.73940304026091130

H -1.47868190021890644 4.70190938058996544 4.02155876930897005
O₄ -1.77590858265049611 -4.22111737753139238 2.31351303736293223
H -3.08254178027425274 -5.30682478279973857 1.70696553489168079
H -1.53399434683936420 -4.67220210939131597 4.05007081541536262
O₂ 4.1574099999999961 0.00000000000000000 0.00000000000000000
H₃ 4.90998605451964298 1.49225096384235001 -0.781804092125152672
H₄ 4.91923494404092754 -1.48827754867483963 -0.780761814665269283
O₅ 6.05659840543499062 -4.20373264521914525 -2.25935023805129687
H 5.82071624009628952 -4.69135833238366651 -3.98156235020776306
H 7.36818104980110355 -5.27499584156926904 -1.61872564663338214
O₆ 6.02609721101799600 4.21365583963911661 -2.24399518443713442
H 7.39757384723865652 5.23453508977235327 -1.66029744465594686
H 5.78598346522806395 4.66863418980124223 -3.97912920400652004

$d_{O_1O_2} = 2.3 \text{ \AA}$

O₁ 0.00000000000000000000000000000000 0.00000000000000000000000000000000
H₁ -0.672231861277311049 -1.48943177887905009 0.865941443068235950
H₂ -0.671965088430154012 1.49840679620850992 0.854385371302619046
H⁺ 2.21095942438598980 -0.00583325830061789691 -0.011460583322823939
O₃ -1.66599922669396006 4.22977903652838005 2.36659617936467992
H -3.06315210669202020 5.22086015507950041 1.78974191469109001
H -1.45739527989850992 4.65213336428948043 4.11038567801618004
O₄ -1.64223778290977007 -4.22630965541418036 2.37857885680366010
H -3.02836007492444992 -5.22007667073768022 1.76163285829786997
H -1.48564871407700005 -4.64586700491449012 4.12791486339451019
O₂ 4.34637000000000029 0.00000000000000000 0.00000000000000000
H₃ 5.04286999999999974 1.49404799571396008 -0.833589179192546048
H₄ 5.04434000000000005 -1.48653180168575005 -0.851189702118658964
O₅ 6.16173000000000037 -4.19605933331070968 -2.35451065023932005
H 6.00002999999999975 -4.69066854269435041 -4.08045575777784997
H 7.47946000000000044 -5.23585562792101999 -1.67700285709634000
O₆ 6.13734000000000002 4.21663188798901967 -2.32391321244980009
H 7.52583000000000002 5.19509452834735974 -1.70869148193543996
H 5.96028000000000002 4.66095132700209991 -4.06439644607784967

$d_{O_1O_2} = 2.45 \text{ \AA}$

O₁ 0.00000000000000000 0.00000000000000000 0.00000000000000000
H₁ -0.649457870422531 -1.51146344938398 0.867787711652323
H₂ -0.660679410743838 1.51140859782633 0.861427111666774
H⁺ 2.06874378590135 -0.001330665292651613 -0.01970565869139078
O₃ -1.70483674995470 4.15060127420642 2.32625655857781

Appendix E. Protonated hexamer geometries

H -3.06258085513488 5.19412570007702 1.74680438752827
H -1.41713939716088 4.64156957966311 4.04615145212262
O₄ -1.68934538969044 -4.15613617640511 2.34230769674381
H -3.01333586569087 -5.22777368640082 1.73340024612704
H -1.44537686678492 -4.61155584226275 4.07305794546894
O₂ 4.62983000000000 0.00000000000000 0.00000000000000
H₃ 5.34677865916538 1.48153237558591 -0.823025465736171
H₄ 5.35654404724709 -1.46989314295902 -0.825051323370663
O₅ 6.45657539876186 -4.27891433461566 -2.34761469008447
H 6.33785949003679 -4.73171126037798 -4.09306736602237
H 7.79348171662438 -5.28280624617687 -1.67042450612108
O₆ 6.42418694520445 4.27987934901450 -2.31200193818929
H 7.84050536187695 5.22541461297016 -1.71377537671137
H 6.26654800927304 4.69491721093731 -4.06173936702921

$$d_{O_1O_2} = 2.5 \text{ \AA}$$

O₁ 0.00000000000000 0.00000000000000 0.00000000000000
H₁ -0.641260237936665 -1.51625786457856 0.883096970942756
H₂ -0.638857029826591 1.52299050583927 0.870034370192830
H⁺ 2.01891648967164 -0.001847164175086084 -0.02209527303551334
O₃ -1.69432397251871 4.11789829981312 2.31211154516694
H -3.04276676427100 5.16315879403155 1.71512443618780
H -1.38857712555825 4.62420535634594 4.01951550969670
O₄ -1.69504176009187 -4.11177582288520 2.32756676382287
H -2.99533858703151 -5.19629594287010 1.69764975238922
H -1.42149457338607 -4.60060227817583 4.04365809746370
O₂ 4.72432000000000 0.00000000000000 0.00000000000000
H₃ 5.45038458079236 1.47555549113125 -0.809558616819048
H₄ 5.45258594765357 -1.46304849451531 -0.831884240828586
O₅ 6.52962153333428 -4.33415288580668 -2.34931764280799
H 6.46823037378107 -4.74395550102890 -4.11193518066912
H 7.89882377593056 -5.30313252484014 -1.68501039376993
O₆ 6.53656634943420 4.31654271578474 -2.31844333505248
H 7.96805288968467 5.24602998234402 -1.72509731159944
H 6.38750607873918 4.73308804209213 -4.07256862525548

$$d_{O_1O_2} = 2.65 \text{ \AA}$$

O₁ 0.00000000000000 0.00000000000000 0.00000000000000
H₁ -0.621758275069062 -1.52695489268690 0.900167471796610
H₂ -0.626740828989495 1.54190046473686 0.864412633985831
H⁺ 1.94907961776403 0.001091654668240812 -0.02752992948223971
O₃ -1.68130339455875 4.08858096846665 2.28155182199791

H -3.03111941645082 5.14042006554552 1.69783501428546
H -1.35803365626228 4.59179761863472 3.99259094858927
O₄ -1.70092072249403 -4.05762416427756 2.32171432981653
H -2.96888543111223 -5.16749026629747 1.66379259604977
H -1.40704936660107 -4.57932016918384 4.02895923050992
O₂ 5.00778000000000 0.000000000000000 0.000000000000000
H₃ 5.68787643744101 1.48257180050759 -0.831649926466367
H₄ 5.69372246002546 -1.45020056158041 -0.872941144756767
O₅ 6.65155229716932 -4.38576346609515 -2.38216462717774
H 6.57892754205484 -4.82100983771418 -4.13248451371152
H 8.04215984421512 -5.33451314014360 -1.71724733186054
O₆ 6.62332329609797 4.41124152301511 -2.34134717242466
H 8.07169230052394 5.31759990724924 -1.75612379923995
H 6.53243475516271 4.75763503292777 -4.10818763501353

Protonated hexamer fitting parameters

In this Appendix, we report the obtained parameters by fitting the proton displacement probability distribution of the protonated hexamer, for both classical and quantum particles. Results are given in atomic units.

F.1 Classical fit: the 2-species model

T (K)	λ_E	$d_E(a.u.)$	$\sigma_E^2(a.u.)$	$\sigma_Z^2(a.u.)$	χ^2
200	0.873	0.216	0.003673	0.0146	0.00561229
250	0.801	0.275	0.00472	0.0315	0.0285128
300	0.80	0.291	0.00555	0.0314	0.0185303
350	0.819	0.281	0.0102	0.029	0.0089176
400	0.623	0.294	0.00762	0.0326	0.0119708

Table F.1 – Fitting parameters obtained by fitting distributions of the proton displacement with the 3-gaussian model of Eq. (6.3) in the classical particles case.

F.2 Quantum fit: the 3-species model

T (K)	λ_E	$d_E(a.u.)$	$\sigma_E^2(a.u.)$	λ_{EZ}	$d_{EZ}(a.u.)$	$\sigma_{EZ}^2(a.u.)$	$\sigma_{SZ}^2(a.u.)$	χ^2
200	0.023	0.307	0.0297	0.946	0.153	0.0112	0.00162	0.000499472
250	0.165	0.307	0.0047	0.691	0.153	0.0102	0.00611	0.0010972
300	0.160	0.307	0.0114	0.708	0.153	0.0126	0.00920	0.000745683
350	0.492	0.310	0.0115	0.204	0.149	0.0046	0.00963	0.00112696
400	0.495	0.368	0.0184	0.203	0.189	0.0079	0.01612	0.00100822

Table F.2 – Fitting parameters obtained by fitting distributions of the proton displacement with the 5-gaussian model of Eq. (6.5) in the quantum particles case.

Bibliography

- [1] Nilsson, A. & Pettersson, L. G. M. The structural origin of anomalous properties of liquid water. *Nat Commun*, 6:R1669, **2015**. Cited on page 1
- [2] Russo, J., Akahane, K. & Tanaka, H. Water-like anomalies as a function of tetrahedrality. *Proc Natl Acad Sci USA*, 115:E3333–E3341, **2018**. Cited on page 1
- [3] Cipcigan, F., Sokhan, V., Martyna, G. & Crain, J. Structure and hydrogen bonding at the limits of liquid water stability. *Sci Rep*, 8:5355, **2018**. Cited on page 1
- [4] Gaiduk, A. P., Gustafson, J., Gygi, F. & Galli, G. First-principles simulations of liquid water using a dielectric-dependent hybrid functional. *J. Phys. Chem. Lett.*, 9:3068–3073, **2018**. Cited on page 1
- [5] Hernandez, J.-A. & Caracas, R. Proton dynamics and the phase diagram of dense water ice. *The Journal of Chemical Physics*, 148:214501, **2018**. Cited on page 1
- [6] Rozsa, V., Pan, D., Giberti, F. & Galli, G. Ab initio spectroscopy and ionic conductivity of water under earth mantle conditions. *Proc Natl Acad Sci USA*, 115:6952–6957, **2018**. Cited on page 1
- [7] Hynes, J. T. The protean proton in water. *Nature*, 397:565–567, **1999**. Cited on page 1
- [8] von Grotthuss, F. C. J. D. T. *Ann. Chim.*, 58:54–73, **1806**. Cited on pages 1 & 123
- [9] Eigen, M. Proton transfer, acid-base catalysis, and enzymatic hydrolysis. part i: Elementary processes. *Angew. Chem. Int. Ed. Engl.*, 3:1–19, **1964**. Cited on pages 1 & 123
- [10] Zundel, G. & Metzger, H. Energiebänder der tunnelnden überschuß-protonen in flüssigen säuren. eine ir-spektroskopische untersuchung der natur der gruppierungen. *Zeitschrift für Physikalische Chemie*, 58:225–245, **1968**. Cited on pages 1 & 123
- [11] Luz, Z. & Meiboom, S. Rate and mechanism of proton exchange in aqueous solutions of phenol-sodium phenolate buffer. *J. Am. Chem. Soc.*, 86:4766–4768, **1964**. Cited on page 1
- [12] Schmitt, U. W. & Voth, G. A. The computer simulation of proton transport in water. *J. Chem. Phys.*, 111(20):9361–9381, **1999**. URL <http://scitation.aip.org/content/aip/journal/jcp/111/20/10.1063/1.480032>. Cited on page 1
- [13] Marx, D., Tuckerman, M. E., Hutter, J. & Parrinello, M. *Nature*, 397:601–604, **1999**. Cited on pages 1, 57, 73, & 152
- [14] Marx, D., Chandra, A. & Tuckerman, M. E. Aqueous basic solutions: Hydroxide solvation, structural diffusion, and comparison to the hydrated proton. *Chem. Rev.*, 110(4):2174–2216, **2010**. URL <http://pubs.acs.org/doi/abs/10.1021/cr900233f>. PMID: 20170203. Cited on page 1
- [15] Hassanali, A., Giberti, F., Cuny, J., Kühne, T. D. & Parrinello, M. Proton transfer through the water gossamer. *Proc. Natl. Acad. Sci. U.S.A.*, 110(34):13723–13728, **2013**. URL <http://www.pnas.org/content/110/34/13723.abstract>. Cited on page 1

-
- [16] Chen, M., Zheng, L., Santra, B., Ko, H.-Y., DiStasio Jr, R. A., Klein, M. L., Car, R. & Wu, X. Hydroxide diffuses slower than hydronium in water because its solvated structure inhibits correlated proton transfer. *Nature Chem*, 10:413–419, **2018**. Cited on pages 1 & 134
- [17] Dahms, F., Fingerhut, B. P., Nibbering, E. T. J., Pines, E. & Elsaesser, T. Large-amplitude transfer motion of hydrated excess protons mapped by ultrafast 2d ir spectroscopy. *Science*, 357:491–495, **2017**. Cited on pages 1, 87, 144, 150, 154, & 161
- [18] Williamson, A. J., Hood, R. Q. & Grossman, J. C. Linear-scaling quantum monte carlo calculations. *Phys. Rev. Lett.*, 87:33, **2001**. Cited on page 1
- [19] Luo, Y., Zen, A. & Sorella, S. Ab initio molecular dynamics with noisy forces: Validating the quantum monte carlo approach with benchmark calculations of molecular vibrational properties. *J. Chem. Phys.*, 141:194112, **2014**. Cited on pages 2, 70, & 81
- [20] Luo, Y. & Sorella, S. Ab initio molecular dynamics with quantum monte carlo. *Front. Mater.*, 2, **2015**. Cited on page 2
- [21] CEPERLEY, D. & ALDER, B. Quantum monte carlo. *Science*, 231:555–560, **1986**. Cited on page 2
- [22] Jiang, J.-C., Wang, Y.-S., Chang, H.-C., Lin, S. H., Lee, Y. T., Niedner-Schatteburg, G. & Chang, H.-C. Infrared spectra of h. *J. Am. Chem. Soc.*, 122:1398–1410, **2000**. Cited on pages 2 & 124
- [23] Mizuse, K. & Fujii, A. Infrared photodissociation spectroscopy of $\text{h}+(\text{h}_2\text{o})_6\hat{\text{A}}\cdot\text{mm}$ ($m = \text{ne, ar, kr, xe, h}_2, \text{n}_2, \text{and ch}_4$): messenger-dependent balance between h_3o^+ and h_5o_2^+ core isomers. *Phys. Chem. Chem. Phys.*, 13:7129, **2011**. Cited on pages 2 & 124
- [24] Coppens, P. Electron density from x-ray diffraction. *Annu. Rev. Phys. Chem.*, 43:663–692, **1992**. Cited on page 9
- [25] Hohenberg, P. & Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B*, 136:864–871, **1964**. Cited on page 9
- [26] Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, **1965**. Cited on page 9
- [27] Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Dal Corso, A., de Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A. P., Smogunov, A., Umari, P. & Wentzcovitch, R. M. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21(39):395502 (19pp), **2009**. URL <http://www.quantum-espresso.org>. Cited on page 10
- [28] Werner, H.-J., Knowles, P. J., Knizia, G., Manby, F. R. & Schütz, M. Molpro: a general purpose quantum chemistry program package. *WIREs Comput Mol Sci*, 2:242–253, **2012**. Cited on page 10
- [29] Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, **1980**. Cited on pages 11, 22, & 30
- [30] Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, **1981**. Cited on page 11
-

Bibliography

- [31] Harris, J. Simplified method for calculating the energy of weakly interacting fragments. *Phys. Rev. B*, 31:1770–1779, **1985**. Cited on page 11
- [32] Lee, C., Chen, H. & Fitzgerald, G. Structures of the water hexamer using density functional methods. *The Journal of Chemical Physics*, 101:4472–4473, **1994**. Cited on page 11
- [33] Lee, C., Chen, H. & Fitzgerald, G. Chemical bonding in water clusters. *The Journal of Chemical Physics*, 102:1266–1269, **1995**. Cited on page 11
- [34] Kanai, Y. & Grossman, J. C. Role of exchange in density-functional theory for weakly interacting systems: Quantum monte carlo analysis of electron density and interaction energy. *Phys. Rev. A*, 80:1200, **2009**. Cited on page 12
- [35] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, **1988**. Cited on page 12
- [36] Lee, C., Yang, W. & Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37:785–789, **1988**. Cited on page 12
- [37] Grossman, J. C., Schwegler, E., Draeger, E. W., Gygi, F. & Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. *The Journal of Chemical Physics*, 120:300–311, **2004**. Cited on page 12
- [38] McGrath, M. J., Siepmann, J. I., Kuo, I.-F. W., Mundy, C. J., VandeVondele, J., Hutter, J., Mohamed, F. & Krack, M. Isobaric-isothermal monte carlo simulations from first principles: Application to liquid water at ambient conditions. *ChemPhysChem*, 6:1894–1901, **2005**. Cited on page 12
- [39] Schmidt, J., VandeVondele, J., Kuo, I.-F. W., Sebastiani, D., Siepmann, J. I., Hutter, J. & Mundy, C. J. Isobaric–isothermal molecular dynamics simulations utilizing density functional theory: An assessment of the structure and density of water at near-ambient conditions. *J. Phys. Chem. B*, 113:11959–11964, **2009**. Cited on page 12
- [40] Wang, Y., Huang, X., Shepler, B. C., Braams, B. J. & Bowman, J. M. Flexible,. *The Journal of Chemical Physics*, 134:094509, **2011**. Cited on pages 12 & 21
- [41] Santra, B., Michaelides, A. & Scheffler, M. Coupled cluster benchmarks of water monomers and dimers extracted from density-functional theory liquid water: The importance of monomer deformations. *The Journal of Chemical Physics*, 131:124509, **2009**. Cited on pages 12 & 21
- [42] Todorova, T., Seitsonen, A. P., Hutter, J., Kuo, I.-F. W. & Mundy, C. J. Molecular dynamics simulation of liquid water: Hybrid density functionals. *J. Phys. Chem. B*, 110:3685–3691, **2006**. Cited on page 13
- [43] Xu, X. & Goddard, W. A. Bonding properties of the water dimer: A comparative study of density functional theories. *J. Phys. Chem. A*, 108:2305–2313, **2004**. Cited on pages 13 & 115
- [44] Bryantsev, V. S., Diallo, M. S., van Duin, A. C. T. & Goddard, W. A. Evaluation of b3lyp, x3lyp, and m06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters. *J. Chem. Theory Comput.*, 5:1016–1026, **2009**. Cited on page 13
- [45] Shi, R., Huang, X., Su, Y., Lu, H.-G., Li, S.-D., Tang, L. & Zhao, J. Which density functional should be used to describe protonated water clusters? *J. Phys. Chem. A*, 121:3117–3127, **2017**. Cited on page 13

-
- [46] Grimme, S. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.*, 27:1787–1799, **2006**. Cited on page 13
- [47] Ma, Z., Zhang, Y. & Tuckerman, M. E. *The Journal of Chemical Physics*, 137:044506, **2012**. Cited on page 14
- [48] Dion, M., Rydberg, H., Schröder, E., Langreth, D. C. & Lundqvist, B. I. Van der waals density functional for general geometries. *Phys. Rev. Lett.*, 92:175, **2004**. Cited on page 14
- [49] Lee, K., Murray, E. D., Kong, L., Lundqvist, B. I. & Langreth, D. C. Higher-accuracy van der waals density functional. *Phys. Rev. B*, 82, **2010**. Cited on page 14
- [50] Kelkkanen, A. K., Lundqvist, B. I. & Nørskov, J. K. Density functional for van der waals forces accounts for hydrogen bond in benchmark set of water hexamers. *The Journal of Chemical Physics*, 131:046102, **2009**. Cited on page 14
- [51] Møgelhøj, A., Kelkkanen, A. K., Wikfeldt, K. T., Schiøtz, J., Mortensen, J. J., Pettersson, L. G. M., Lundqvist, B. I., Jacobsen, K. W., Nilsson, A. & Nørskov, J. K. Ab initio van der waals interactions in simulations of water alter structure from mainly tetrahedral to high-density-like. *J. Phys. Chem. B*, 115:14149–14160, **2011**. Cited on page 14
- [52] Zhang, C., Wu, J., Galli, G. & Gygi, F. Structural and vibrational properties of liquid water from van der waals density functionals. *J. Chem. Theory Comput.*, 7:3054–3061, **2011**. Cited on page 14
- [53] Gillan, M. J., Alfè, D. & Michaelides, A. Perspective: How good is dft for water? *J. Chem. Phys.*, 144:130901, **2016**. Cited on page 14
- [54] Sun, J., Remsing, R. C., Zhang, Y., Sun, Z., Ruzsinszky, A., Peng, H., Yang, Z., Paul, A., Waghmare, U., Wu, X., Klein, M. L. & Perdew, J. P. Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nature Chem*, 8:831–836, **2016**. Cited on page 14
- [55] Chen, M., Ko, H.-Y., Remsing, R. C., Calegari Andrade, M. F., Santra, B., Sun, Z., Selloni, A., Car, R., Klein, M. L., Perdew, J. P. & Wu, X. Ab initio theory and modeling of water. *Proceedings of the National Academy of Sciences*, 114(41):10846–10851, **2017**. URL <http://www.pnas.org/content/114/41/10846>. Cited on page 15
- [56] Gillan, M. J., Alfè, D., Bygrave, P. J., Taylor, C. R. & Manby, F. R. Energy benchmarks for water clusters and ice structures from an embedded many-body expansion. *The Journal of Chemical Physics*, 139:114101, **2013**. Cited on page 17
- [57] Miliordos, E. & Xantheas, S. S. An accurate and efficient computational protocol for obtaining the complete basis set limits of the binding energies of water clusters at the mp2 and ccSD(T) levels of theory: Application to (h). *The Journal of Chemical Physics*, 142:234303, **2015**. Cited on page 17
- [58] Parkkinen, P., Riikonen, S. & Halonen, L. Global minima of protonated water clusters (h). *J. Phys. Chem. A*, 116:10826–10835, **2012**. Cited on page 18
- [59] Del Ben, M., Schönherr, M., Hutter, J. & VandeVondele, J. Bulk liquid water at ambient temperature and pressure from mp2 theory. *J. Phys. Chem. Lett.*, 4:3753–3759, **2013**. Cited on page 18
-

Bibliography

- [60] Grimme, S. Improved second-order mller–plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *The Journal of Chemical Physics*, 118:9095–9102, **2003**. Cited on page 18
- [61] Willow, S. Y., Salim, M. A., Kim, K. S. & Hirata, S. Ab initio molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction. *Sci. Rep.*, 5:14358, **2015**. Cited on pages 18 & 19
- [62] Willow, S. Y., Zeng, X. C., Xantheas, S. S., Kim, K. S. & Hirata, S. Why is mp2-water “cooler” and “denser” than dft-water? *J. Phys. Chem. Lett.*, 7:680–684, **2016**. Cited on page 18
- [63] Bartlett, R. J. & MusiaÅ, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.*, 79:291–352, **2007**. Cited on page 19
- [64] Bates, D. M. & Tschumper, G. S. Ccsd(t) complete basis set limit relative energies for low-lying water hexamer structures. *J. Phys. Chem. A*, 113:3555–3559, **2009**. Cited on page 21
- [65] Huang, X., Braams, B. J. & Bowman, J. M. Ab initio potential energy and dipole moment surfaces for h5o2+. *J. Chem. Phys.*, 122(4):044308, **2005**. URL <http://scitation.aip.org/content/aip/journal/jcp/122/4/10.1063/1.1834500>. Cited on pages 21, 51, 71, 87, 88, 93, 94, 98, & 104
- [66] Wang, Y., Shepler, B. C., Braams, B. J. & Bowman, J. M. Full-dimensional, ab initio potential energy and dipole moment surfaces for water. *J. Chem. Phys.*, 131:054511, **2009**. Cited on page 21
- [67] Wang, Y. & Bowman, J. M. *The Journal of Chemical Physics*, 134:154510, **2011**. Cited on page 21
- [68] Babin, V., Leforestier, C. & Paesani, F. Development of a “first principles” water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient. *J. Chem. Theory Comput.*, 9:5395–5403, **2013**. Cited on pages 21 & 70
- [69] Babin, V., Medders, G. R. & Paesani, F. Development of a “first principles” water potential with flexible monomers. ii: Trimer potential energy surface, third virial coefficient, and small clusters. *J. Chem. Theory Comput.*, 10:1599–1607, **2014**. Cited on pages 21 & 70
- [70] 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.*, 10:2906–2910, **2014**. Cited on page 21
- [71] McMillan, W. L. Ground state of liquid. *Phys. Rev.*, 138:A442–A451, **1965**. Cited on page 22
- [72] Dagrada, M., Casula, M., Saitta, A. M., Sorella, S. & Mauri, F. Quantum monte carlo study of the protonated water dimer. *J. Chem. Theory Comput.*, 10:1980–1993, **2014**. Cited on pages vi, vii, 22, 29, 34, 79, 88, 89, 93, 94, 104, 107, 113, 134, 135, & 136
- [73] Booth, G. H., Thom, A. J. W. & Alavi, A. Fermion monte carlo without fixed nodes: A game of life, death, and annihilation in slater determinant space. *J. Chem. Phys.*, 131:054106, **2009**. Cited on page 23
- [74] Zhang, S. & Krakauer, H. Quantum monte carlo method using phase-free random walks with slater determinants. *Phys. Rev. Lett.*, 90:1, **2003**. Cited on page 23
- [75] Casula, M. & Sorella, S. Geminal wave functions with jastrow correlation: A first application to atoms. *The Journal of Chemical Physics*, 119:6500–6511, **2003**. Cited on page 23

-
- [76] Dupuy, N., Bouaouli, S., Mauri, F., Sorella, S. & Casula, M. Vertical and adiabatic excitations in anthracene from quantum monte carlo: Constrained energy minimization for structural and electronic excited-state properties in the jagg ansatz. *The Journal of Chemical Physics*, 142:214109, **2015**. Cited on page 23
- [77] Casula, M., Attaccalite, C. & Sorella, S. Correlated geminal wave function for molecules: an efficient resonating valence bond approach. *J. Chem. Phys.*, 121:7110, **2004**. Cited on pages 23 & 79
- [78] Sorella, S., Casula, M. & Rocca, D. Weak binding between two aromatic rings: Feeling the van der waals attraction by quantum monte carlo methods. *The Journal of Chemical Physics*, 127:014105, **2007**. Cited on page 25
- [79] Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Comm. Pure Appl. Math.*, 10:151–177, **1957**. Cited on page 25
- [80] Burkatzki, M., Filippi, C. & Dolg, M. Energy-consistent pseudopotentials for quantum monte carlo calculations. *J. Chem. Phys.*, 126(23):234105, **2007**. URL <http://scitation.aip.org/content/aip/journal/jcp/126/23/10.1063/1.2741534>. Cited on pages 25, 28, 32, & 90
- [81] De Michelis, C., Masserini, G. & Reatto, L. Jastrow wave function for condensed phases of bose particles: Hard-sphere system. *Phys. Rev. A*, 18:296–306, **1978**. Cited on page 26
- [82] Ceperley, D., Chester, G. V. & Kalos, M. H. Monte carlo simulation of a many-fermion study. *Phys. Rev. B*, 16:3081–3099, **1977**. Cited on page 26
- [83] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092, **1953**. Cited on page 27
- [84] Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, **1970**. Cited on page 27
- [85] Sorella, S. Green function monte carlo with stochastic reconfiguration. *Phys. Rev. Lett.*, 80:4558–4561, **1998**. Cited on pages 29 & 173
- [86] Sorella, S. Wave function optimization in the variational monte carlo method. *Phys. Rev. B*, 71, **2005**. Cited on pages 29 & 173
- [87] Gurtubay, I. G. & Needs, R. J. Dissociation energy of the water dimer from quantum monte carlo calculations. *The Journal of Chemical Physics*, 127:124306, **2007**. Cited on pages 29 & 32
- [88] Sterpone, F., Spanu, L., Ferraro, L., Sorella, S. & Guidoni, L. Dissecting the hydrogen bond: A quantum monte carlo approach. *J. Chem. Theory Comput.*, 4(9):1428–1434, **2008**. URL <http://pubs.acs.org/doi/abs/10.1021/ct800121e>. Cited on pages 29, 34, 115, & 118
- [89] Zen, A., Luo, Y., Sorella, S. & Guidoni, L. Molecular properties by quantum monte carlo: An investigation on the role of the wave function ansatz and the basis set in the water molecule. *J. Chem. Theory Comput.*, 9(10):4332–4350, **2013**. URL <http://pubs.acs.org/doi/abs/10.1021/ct400382m>. Cited on pages 29 & 115
- [90] Kalos, M. H., Levesque, D. & Verlet, L. Helium at zero temperature with hard-sphere and other forces. *Phys. Rev. A*, 9:2178–2195, **1974**. Cited on page 29
- [91] Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum monte carlo simulations of solids. *Rev. Mod. Phys.*, 73:33–83, **2001**. Cited on pages 30 & 31
-

Bibliography

- [92] Klopper, W. Highly accurate coupled-cluster singlet and triplet pair energies from explicitly correlated calculations in comparison with extrapolation techniques. *Molecular Physics*, 99:481–507, **2001**. Cited on page 31
- [93] Caffarel, M., Applencourt, T., Giner, E. & Scemama, A. Communication: Toward an improved control of the fixed-node error in quantum monte carlo: The case of the water molecule. *The Journal of Chemical Physics*, 144:151103, **2016**. Cited on page 31
- [94] Trotter, H. F. On the product of semi-groups of operators. *Proc. Amer. Math. Soc.*, 10:545–545, **1959**. Cited on pages 31, 66, & 69
- [95] Suzuki, M. Decomposition formulas of exponential operators and lie exponentials with some applications to quantum mechanics and statistical physics. *Journal of Mathematical Physics*, 26:601–612, **1985**. Cited on page 31
- [96] Toulouse, J., Assaraf, R. & Umrigar, C. J. Chapter fifteen - introduction to the variational and diffusion monte carlo methods. Dans: *Electron Correlation in Molecules ab initio Beyond Gaussian Quantum Chemistry*, vol. 73 de *Advances in Quantum Chemistry*, édité par Hoggan, P. E. & Ozdogan, T., pages 285 – 314. Academic Press, **2016**. URL <http://www.sciencedirect.com/science/article/pii/S0065327615000386>. Cited on page 31
- [97] Gregory, J. K. & Clary, D. C. A comparison of conventional and rigid body diffusion monte carlo techniques. application to water dimer. *Chemical Physics Letters*, 228(6):547 – 554, **1994**. URL <http://www.sciencedirect.com/science/article/pii/0009261494009872>. Cited on page 32
- [98] Gregory, J. K. & Clary, D. C. Calculations of the tunneling splittings in water dimer and trimer using diffusion monte carlo. *The Journal of Chemical Physics*, 102:7817–7829, **1995**. Cited on page 32
- [99] Xu, J. & Jordan, K. D. Application of the diffusion monte carlo method to the binding of excess electrons to water clusters. *J. Phys. Chem. A*, 114:1364–1366, **2010**. Cited on page 32
- [100] Gillan, M. J., Manby, F. R., Towler, M. D. & Alfé, D. Assessing the accuracy of quantum monte carlo and density functional theory for energetics of small water clusters. *J. Chem. Phys.*, 136(24):244105, **2012**. URL <http://scitation.aip.org/content/aip/journal/jcp/136/24/10.1063/1.4730035>. Cited on pages 32 & 63
- [101] Ceperley, D. M. The statistical error of green’s function monte carlo. *J Stat Phys*, 43:815–826, **1986**. Cited on page 32
- [102] Ma, A., Drummond, N. D., Towler, M. D. & Needs, R. J. All-electron quantum monte carlo calculations for the noble gas atoms he to xe. *Phys. Rev. E*, 71:283, **2005**. Cited on page 32
- [103] Mitáš, L., Shirley, E. L. & Ceperley, D. M. Nonlocal pseudopotentials and diffusion monte carlo. *The Journal of Chemical Physics*, 95:3467–3475, **1991**. Cited on page 32
- [104] Casula, M., Filippi, C. & Sorella, S. Diffusion monte carlo method with lattice regularization. *Phys. Rev. Lett.*, 95:1431, **2005**. Cited on page 33
- [105] Casula, M., Moroni, S., Sorella, S. & Filippi, C. Size-consistent variational approaches to nonlocal pseudopotentials: Standard and lattice regularized diffusion monte carlo methods revisited. *The Journal of Chemical Physics*, 132:154113, **2010**. Cited on page 33
- [106] Filippi, C. & Umrigar, C. J. Correlated sampling in quantum monte carlo: A route to forces. *Phys. Rev. B*, 61:R16291–R16294, **2000**. Cited on page 35

-
- [107] Sorella, S. & Capriotti, L. Algorithmic differentiation and the calculation of forces by quantum monte carlo. *J. Chem. Phys.*, 133:234111, **2010**. Cited on pages 35 & 175
- [108] Marx, D. Proton transfer 200 years after von grotthuss: Insights from ab initio simulations. *ChemPhysChem*, 7(9):1848–1870, **2006**. URL <http://dx.doi.org/10.1002/cphc.200600128>. Cited on pages 37 & 123
- [109] Tuckerman, M. E., Marx, D. & Parrinello, M. The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, 417:925–929, **2002**. Cited on pages 37 & 57
- [110] Tuckerman, M. E., Marx, D., Klein, M. L. & Parrinello, M. On the quantum nature of the shared proton in hydrogen bonds. *Science*, 275(5301):817–820, **1997**. Cited on pages 38, 57, & 87
- [111] Sit, P. H.-L. & Marzari, N. Static and dynamical properties of heavy water at ambient conditions from first-principles molecular dynamics. *J. Chem. Phys.*, 122:204510, **2005**. Cited on pages 39 & 99
- [112] Box, G. E. P. & Muller, M. E. A note on the generation of random normal deviates. *Ann. Math. Statist.*, 29:610–611, **1958**. Cited on page 40
- [113] Verlet, L. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, **1967**. Cited on page 41
- [114] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76:637–649, **1982**. Cited on page 41
- [115] Jones, J. E. On the determination of molecular fields. ii. from the equation of state of a gas. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 106:463–477, **1924**. Cited on page 42
- [116] Buckingham, R. A. The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 168:264–283, **1938**. Cited on page 42
- [117] Berendsen, H. J. C., Postma, J., van Gunsteren, W. & Hermans, J. *Interaction Models for Water in Relation to Protein Hydration*. Springer, Dordrecht, Netherlands, **1981**. Cited on pages 42 & 157
- [118] Jorgensen, W. L. & Madura, J. D. Quantum and statistical mechanical studies of liquids. 25. solvation and conformation of methanol in water. *J. Am. Chem. Soc.*, 105:1407–1413, **1983**. Cited on page 42
- [119] 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. *The Journal of Chemical Physics*, 112:8910–8922, **2000**. Cited on page 42
- [120] Mark, P. & Nilsson, L. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *J. Phys. Chem. A*, 105:9954–9960, **2001**. Cited on page 42
- [121] Skyner, R. E., McDonagh, J. L., Groom, C. R., van Mourik, T. & Mitchell, J. B. O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.*, 17:6174–6191, **2015**. Cited on page 43
-

Bibliography

- [122] Asta, A. J., Levesque, M., Vuilleumier, R. & Rotenberg, B. Transient hydrodynamic finite-size effects in simulations under periodic boundary conditions. *Phys. Rev. E*, 95, **2017**. Cited on page 43
- [123] Yeh, L. I., Okumura, M., Myers, J. D., Price, J. M. & Lee, Y. T. Vibrational spectroscopy of the hydrated hydronium cluster ions h. *The Journal of Chemical Physics*, 91:7319–7330, **1989**. Cited on page 43
- [124] Page, A. J., Isomoto, T., Knaup, J. M., Irle, S. & Morokuma, K. Effects of molecular dynamics thermostats on descriptions of chemical nonequilibrium. *J. Chem. Theory Comput.*, 8:4019–4028, **2012**. Cited on page 43
- [125] Kuo, I.-F. W., Mundy, C. J., McGrath, M. J., Siepmann, J. I., VandeVondele, J., Sprik, M., Hutter, J., Chen, B., Klein, M. L., Mohamed, F., Krack, M. & Parrinello, M. Liquid water from first principles: Investigation of different sampling approaches. *J. Phys. Chem. B*, 108:12990–12998, **2004**. Cited on page 44
- [126] Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, 72:2384–2393, **1980**. Cited on page 44
- [127] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81:3684–3690, **1984**. Cited on page 44
- [128] Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics*, 81:511–519, **1984**. Cited on page 44
- [129] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, **1985**. URL <http://link.aps.org/doi/10.1103/PhysRevA.31.1695>. Cited on page 44
- [130] Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126:014101, **2007**. Cited on page 44
- [131] Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics*, 130:074101, **2009**. Cited on page 44
- [132] Mudi, A. & Chakravarty, C. Effect of the berendsen thermostat on the dynamical properties of water. *Molecular Physics*, 102:681–685, **2004**. Cited on page 44
- [133] Basconi, J. E. & Shirts, M. R. Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations. *J. Chem. Theory Comput.*, 9:2887–2899, **2013**. Cited on page 44
- [134] Kubo, R. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255, **1966**. URL <http://stacks.iop.org/0034-4885/29/i=1/a=306>. Cited on page 46
- [135] Oshovsky, G., Reinhoudt, D. & Verboom, W. Supramolecular chemistry in water. *Angew. Chem. Int. Ed.*, 46:2366–2393, **2007**. Cited on page 46
- [136] Depner, M., Schürmann, B. & Auriemma, F. Investigation of a poly(oxyethylene) chain by a molecular dynamics simulation in an aqueous solution and by langevin dynamics simulations. *Molecular Physics*, 74:715–733, **1991**. Cited on page 46

-
- [137] Widmalm, G. & Pastor, R. W. Comparison of langevin and molecular dynamics simulations. equilibrium and dynamics of ethylene glycol in water. *Faraday Trans.*, 88:1747, **1992**. Cited on page 46
- [138] Cheung, M. S., Garcia, A. E. & Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences*, 99:685–690, **2002**. Cited on page 46
- [139] Damjanović, A., Wu, X., García-Moreno E., B. & Brooks, B. R. Backbone relaxation coupled to the ionization of internal groups in proteins: A self-guided langevin dynamics study. *Biophysical Journal*, 95:4091–4101, **2008**. Cited on page 46
- [140] Skeel, R. D. & Izaguirre, J. A. An impulse integrator for langevin dynamics. *Molecular Physics*, 100:3885–3891, **2002**. Cited on pages 46 & 65
- [141] Ricci, A. & Ciccotti, G. Algorithms for brownian dynamics. *Molecular Physics*, 101:1927–1931, **2003**. Cited on pages 46 & 65
- [142] Vanden-Eijnden, E. & Ciccotti, G. Second-order integrators for langevin equations with holonomic constraints. *Chemical Physics Letters*, 429:310–316, **2006**. Cited on pages 46, 51, & 65
- [143] Bussi, G. & Parrinello, M. Accurate sampling using langevin dynamics. *Phys. Rev. E*, 75, **2007**. Cited on pages 46, 47, 59, & 65
- [144] Melchionna, S. Design of quasisymplectic propagators for langevin dynamics. *J. Chem. Phys.*, 127:044108, **2007**. Cited on pages 46 & 65
- [145] Grønbech-Jensen, N. & Farago, O. A simple and effective verlet-type algorithm for simulating langevin dynamics. *Molecular Physics*, 111:983–991, **2013**. Cited on pages 46 & 65
- [146] Risken, H. *The Fokker-Planck Equation*. Springer, Berlin Heidelberg, Germany, **1984**. Cited on page 47
- [147] Fixman, M. Construction of langevin forces in the simulation of hydrodynamic interaction. *Macromolecules*, 19:1204–1207, **1986**. Cited on page 49
- [148] Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. Langevin stabilization of molecular dynamics. *J. Chem. Phys.*, 114:2090, **2001**. Cited on page 49
- [149] Attaccalite, C. & Sorella, S. Stable liquid hydrogen at high pressure by a novel. *Phys. Rev. Lett.*, 100, **2008**. Cited on pages 49, 63, 64, 70, & 81
- [150] Mazzola, G., Yunoki, S. & Sorella, S. Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation. *Nat Comms*, 5, **2014**. Cited on pages 49, 50, 63, & 64
- [151] Zen, A., Luo, Y., Mazzola, G., Guidoni, L. & Sorella, S. Ab initio molecular dynamics simulation of liquid water by quantum monte carlo. *J. Chem. Phys.*, 142:144111, **2015**. Cited on pages 49, 50, & 64
- [152] Fox, R. F., Gatland, I. R., Roy, R. & Vemuri, G. Fast, accurate algorithm for numerical simulation of exponentially correlated colored noise. *Phys. Rev. A*, 38:5938–5940, **1988**. URL <http://link.aps.org/doi/10.1103/PhysRevA.38.5938>. Cited on page 49
- [153] Posch, H. A., Hoover, W. G. & Vesely, F. J. Canonical dynamics of the nosé oscillator: Stability, order, and chaos. *Phys. Rev. A*, 33:4253–4265, **1986**. Cited on page 51
-

Bibliography

- [154] Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé–hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97:2635–2643, **1992**. Cited on pages 51 & 58
- [155] Mer, V. K. I. & Baker, W. N. The freezing point of mixtures of h₂o and d₂o. :the latent heat of fusion of d₂o. *J. Am. Chem. Soc.*, 56:2641–2643, **1934**. Cited on page 53
- [156] Irikura, K. K. Experimental vibrational zero-point energies: Diatomic molecules. *Journal of Physical and Chemical Reference Data*, 36:389–397, **2007**. Cited on page 53
- [157] Grimm, A. R., Bacskay, G. B. & Haymet, A. D. J. Quantum chemical studies of the solvation of the hydroxide ion. *Mol. Phys.*, 86:369–384, **1995**. Cited on page 53
- [158] Xantheas, S. S. Theoretical study of hydroxide ion-water clusters. *J. Am. Chem. Soc.*, 117:10373–10380, **1995**. Cited on page 53
- [159] Lill, M. A. & Helms, V. Reaction rates for proton transfer over small barriers and connection to transition state theory. *The Journal of Chemical Physics*, 115:7985–7992, **2001**. Cited on page 53
- [160] Benoit, M., Marx, D. & Parrinello, M. Tunnelling and zero-point motion in high-pressure ice. *Nature*, 392:258–261, **1998**. Cited on page 54
- [161] Pugliano, N. & Saykally, R. Measurement of quantum tunneling between chiral isomers of the cyclic water trimer. *Science*, 257:1937–1940, **1992**. Cited on page 54
- [162] Liedl, K. R., Sekušak, S., Kroemer, R. T. & Rode, B. M. New insights into the dynamics of concerted proton tunneling in cyclic water and hydrogen fluoride clusters. *J. Phys. Chem. A*, 101:4707–4716, **1997**. Cited on page 54
- [163] Agmon, N. The grotthuss mechanism. *Chem. Phys. Lett.*, 244(5â6):456–462, **1995**. URL <http://www.sciencedirect.com/science/article/pii/000926149500905J>. Cited on pages 54 & 152
- [164] Masgrau, L. Atomic description of an enzyme reaction dominated by proton tunneling. *Science*, 312:237–241, **2006**. Cited on page 54
- [165] Brovarets', O. O. & Hovorun, D. M. Proton tunneling in the aât watson-crick dna base pair: myth or reality? *Journal of Biomolecular Structure and Dynamics*, 33:2716–2720, **2015**. Cited on page 54
- [166] Romero, E. E. & Hernandez, F. E. Solvent effect on the intermolecular proton transfer of the watson and crick guanine–cytosine and adenine–thymine base pairs: a polarizable continuum model study. *Phys. Chem. Chem. Phys.*, 20:1198–1209, **2018**. Cited on page 54
- [167] Morrone, J. A. & Car, R. Nuclear quantum effects in water. *Phys. Rev. Lett.*, 101, **2008**. Cited on pages 54 & 64
- [168] Sprik, M., Klein, M. L. & Chandler, D. Staging: A sampling technique for the monte carlo evaluation of path integrals. *Phys. Rev. B*, 31:4234–4244, **1985**. Cited on page 56
- [169] Herman, M. F., Bruskin, E. J. & Berne, B. J. On path integral monte carlo simulations. 76:5150–5155, **1982**. Cited on page 57

-
- [170] Wallqvist, A. & Berne, B. Path-integral simulation of pure water. *Chemical Physics Letters*, 117(3):214 – 219, **1985**. URL <http://www.sciencedirect.com/science/article/pii/0009261485802062>. Cited on page 57
- [171] Suzuki, K., Shiga, M. & Tachikawa, M. Temperature and isotope effects on water cluster ions with path integral molecular dynamics based on the fourth order trotter expansion. *The Journal of Chemical Physics*, 129:144310, **2008**. Cited on page 57
- [172] Suzuki, K., Tachikawa, M. & Shiga, M. Temperature dependence on the structure of zundel cation and its isotopomers. *J. Chem. Phys.*, 138:184307, **2013**. Cited on pages 57, 73, & 105
- [173] Takayanagi, T., Yoshikawa, T., Motegi, H. & Shiga, M. Path-integral molecular dynamics simulations for water anion clusters H_2O_5^- and D_2O_5^- . Cited on page 57
- [174] 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.*, 116:7529–7550, **2016**. Cited on page 57
- [175] Markland, T. E. & Ceriotti, M. Nuclear quantum effects enter the mainstream. *Nat. rev. chem.*, 2:0109, **2018**. Cited on page 57
- [176] Parrinello, M. & Rahman, A. Study of an f center in molten kcl. *J. Chem. Phys.*, 80:860, **1984**. Cited on page 58
- [177] Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostating of path integral molecular dynamics. *J. Chem. Phys.*, 133:124104, **2010**. Cited on pages 58, 59, 60, 68, & 71
- [178] Tassone, F., Mauri, F. & Car, R. Acceleration schemes for *ab initio* molecular-dynamics simulations and electronic-structure calculations. *Phys. Rev. B*, 50:10561–10573, **1994**. URL <http://link.aps.org/doi/10.1103/PhysRevB.50.10561>. Cited on pages 59 & 68
- [179] Liu, J., Li, D. & Liu, X. A simple and accurate algorithm for path integral molecular dynamics with the langevin thermostat. *J. Chem. Phys.*, 145:024103, **2016**. Cited on page 60
- [180] Ceriotti, M., Bussi, G. & Parrinello, M. Colored-noise thermostats à la carte. *J. Chem. Theory Comput.*, 6:1170–1180, **2010**. Cited on pages 60 & 80
- [181] Ceriotti, M., Manolopoulos, D. E. & Parrinello, M. Accelerating the convergence of path integral dynamics with a generalized langevin equation. *J. Chem. Phys.*, 134:084104, **2011**. Cited on pages 60 & 80
- [182] Bronstein, Y., Depondt, P., Finocchi, F. & Saitta, A. M. Quantum-driven phase transition in ice described via an efficient langevin approach. *Phys. Rev. B*, 89:214101, **2014**. URL <http://link.aps.org/doi/10.1103/PhysRevB.89.214101>. Cited on page 60
- [183] 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*, 93:024104, **2016**. URL <http://link.aps.org/doi/10.1103/PhysRevB.93.024104>. Cited on page 60
- [184] Dammak, H., Chalopin, Y., Laroche, M., Hayoun, M. & Greffet, J.-J. Quantum thermal bath for molecular dynamics simulation. *Phys. Rev. Lett.*, 103:190601, **2009**. Cited on page 60
- [185] Briec, F., Dammak, H. & Hayoun, M. Quantum thermal bath for path integral molecular dynamics simulation. *J. Chem. Theory Comput.*, 12:1351–1359, **2016**. Cited on page 60
-

Bibliography

- [186] Benedek, N. A., Snook, I. K., Towler, M. D. & Needs, R. J. Quantum monte carlo calculations of the dissociation energy of the water dimer. *The Journal of Chemical Physics*, 125(10):104302, **2006**. URL <http://scitation.aip.org/content/aip/journal/jcp/125/10/10.1063/1.2338032>. Cited on page 63
- [187] Pierleoni, C., Ceperley, D. M. & Holzmann, M. Coupled electron-ion monte carlo calculations of dense metallic hydrogen. *Phys. Rev. Lett.*, 93, **2004**. Cited on page 63
- [188] Pierleoni, C. & Ceperley, D. *The Coupled Electron-Ion Monte Carlo Method*, pages 641–683. Springer Berlin Heidelberg, Berlin, Heidelberg, **2006**. URL http://dx.doi.org/10.1007/3-540-35273-2_18. Cited on pages 63 & 80
- [189] Vega, C., Conde, M. M., McBride, C., Abascal, J. L. F., Noya, E. G., Ramirez, R. & Sesé, L. M. Heat capacity of water: A signature of nuclear quantum effects. *J. Chem. Phys.*, 132:046101, **2010**. Cited on page 64
- [190] Pamuk, B., Soler, J. M., Ramírez, R., Herrero, C. P., Stephens, P. W., Allen, P. B. & Fernández-Serra, M.-V. Anomalous nuclear quantum effects in ice. *Phys. Rev. Lett.*, 108, **2012**. Cited on page 64
- [191] Kong, L., Wu, X. & Car, R. Roles of quantum nuclei and inhomogeneous screening in the x-ray absorption spectra of water and ice. *Phys. Rev. B*, 86, **2012**. Cited on page 64
- [192] Giberti, F., Hassanali, A. A., Ceriotti, M. & Parrinello, M. The role of quantum effects on structural and electronic fluctuations in neat and charged water. *J. Phys. Chem. B*, 118:13226–13235, **2014**. Cited on page 64
- [193] Lobaugh, J. & Voth, G. A. The quantum dynamics of an excess proton in water. *The Journal of Chemical Physics*, 104:2056–2069, **1996**. Cited on page 64
- [194] Markovitch, O., Chen, H., Izvekov, S., Paesani, F., Voth, G. A. & Agmon, N. Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B*, 112:9456–9466, **2008**. Cited on page 64
- [195] Mouhat, F., Sorella, S., Vuilleumier, R., Saitta, A. M. & Casula, M. Fully quantum description of the zundel ion: Combining variational quantum monte carlo with path integral langevin dynamics. *J. Chem. Theory Comput.*, 13:2400–2417, **2017**. Cited on pages 64 & 71
- [196] Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics*, 97(3), **1992**. Cited on page 69
- [197] Sexton, J. C. & Weingarten, D. H. Hamiltonian evolution for the hybrid monte carlo algorithm. *Nuclear Physics B*, 380:667–677, **1992**. Cited on page 69
- [198] Góra, U., Cencek, W., Podeszwa, R., van der Avoird, A. & Szalewicz, K. Predictions for water clusters from a first-principles two- and three-body force field. *The Journal of Chemical Physics*, 140:194101, **2014**. Cited on page 70
- [199] Babin, V., Medders, G. R. & Paesani, F. Toward a universal water model: First principles simulations from the dimer to the liquid phase. *J. Phys. Chem. Lett.*, 3:3765–3769, **2012**. Cited on page 70
- [200] Ceperley, D. M. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.*, 67:279–355, **1995**. Cited on page 72

-
- [201] Spura, T., Elgabarty, H. & Kühne, T. D. “on-the-fly” coupled cluster path-integral molecular dynamics: impact of nuclear quantum effects on the protonated water dimer. *Phys. Chem. Chem. Phys.*, 17:14355–14359, **2015**. Cited on pages 73, 99, & 105
- [202] Sorella, S., Devaux, N., Dagrada, M., Mazzola, G. & Casula, M. Geminal embedding scheme for optimal atomic basis set construction in correlated calculations. *J. Chem. Phys.*, 143:244112, **2015**. Cited on pages 79, 113, & 136
- [203] Sorella, S. Generalized lanczos algorithm for variational quantum monte carlo. *Phys. Rev. B*, 64, **2001**. Cited on page 79
- [204] Umrigar, C. J., Toulouse, J., Filippi, C., Sorella, S. & Hennig, R. G. Alleviation of the fermion-sign problem by optimization of many-body wave functions. *Phys. Rev. Lett.*, 98, **2007**. Cited on page 79
- [205] Cheng, X., Herr, J. D. & Steele, R. P. Accelerating. *J. Chem. Theory Comput.*, 12:1627–1638, **2016**. Cited on page 80
- [206] Poltavsky, I. & Tkatchenko, A. Modeling quantum nuclei with perturbed path integral molecular dynamics. *Chem. Sci.*, 7:1368–1372, **2016**. Cited on page 80
- [207] Markland, T. E. & Manolopoulos, D. E. A refined ring polymer contraction scheme for systems with electrostatic interactions. *Chemical Physics Letters*, 464:256–261, **2008**. Cited on page 80
- [208] Markland, T. E. & Manolopoulos, D. E. An efficient ring polymer contraction scheme for imaginary time path integral simulations. *J. Chem. Phys.*, 129:024105, **2008**. Cited on page 80
- [209] Fanourgakis, G. S., Markland, T. E. & Manolopoulos, D. E. A fast path integral method for polarizable force fields. *J. Chem. Phys.*, 131:094102, **2009**. Cited on page 80
- [210] Ceriotti, M., Bussi, G. & Parrinello, M. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Phys. Rev. Lett.*, 102, **2009**. Cited on page 80
- [211] Ceriotti, M., Bussi, G. & Parrinello, M. Nuclear quantum effects in solids using a colored-noise thermostat. *Phys. Rev. Lett.*, 103, **2009**. Cited on page 80
- [212] Pitsevich, G., Malevich, A., Kozlovskaya, E., Mahnach, E., Doroshenko, I., Pogorelov, V., Pettersson, L. G. M., Sablinskas, V. & Balevicius, V. Mp4 study of the anharmonic coupling of the shared proton stretching vibration of the protonated water dimer in equilibrium and transition states. *J. Phys. Chem. A*, 121:2151–2165, **2017**. Cited on page 87
- [213] Yu, Q. & Bowman, J. M. How the zundel h5o2+ potential can be used to predict the proton stretch and bend frequencies of larger protonated water clusters. *J. Phys. Chem. Lett.*, 7:5259–5265, **2016**. Cited on page 87
- [214] Yu, Q. & Bowman, J. M. High-level quantum calculations of the ir spectra of the eigen, zundel, and ring isomers of h. *J. Am. Chem. Soc.*, 139:10984–10987, **2017**. Cited on page 87
- [215] Zundel, G. & Metzger, H. *Z. Phys. Chem.*, 58:225–245, **1968**. Cited on page 87
- [216] Headrick, J. M., Diken, E. G., Walters, R. S., Hammer, N. I., Christie, R. A., Cui, J., Myshakin, E. M., Duncan, M. A., Johnson, M. A. & Jordan, K. D. Spectral signatures of hydrated proton vibrations in water clusters. *Science*, 308(5729):1765–1769, **2005**. Cited on pages 87 & 124
-

Bibliography

- [217] Wei, D. & Salahub, D. R. Hydrated proton clusters and solvent effects on the proton transfer barrier: A density functional study. *J. Chem. Phys.*, 101(9):7633–7642, **1994**. URL <http://scitation.aip.org/content/aip/journal/jcp/101/9/10.1063/1.468256>. Cited on page 87
- [218] Vuilleumier, R. & Borgis, D. An extended empirical valence bond model for describing proton transfer in $\text{h}^+(\text{h}_2\text{o})_n$ clusters and liquid water. *Chem. Phys. Lett.*, 284:71–77, **1998**. Cited on page 87
- [219] Vuilleumier, R. & Borgis, D. Quantum dynamics of an excess proton in water using an extended empirical valence-bond hamiltonian. *J. Phys. Chem. B*, 102:4261–4264, **1998**. Cited on page 87
- [220] Vuilleumier, R. & Borgis, D. Wavefunction quantization of the proton motion in a h_5o_2^+ dimer solvated in liquid water. *Journal of molecular structure*, 552:117–136, **2000**. Cited on page 87
- [221] Valeev, E. F. & Schaefer, H. F. The protonated water dimer: Brueckner methods remove the spurious c_1 symmetry minimum. *J. Chem. Phys.*, 108(17):7197–7201, **1998**. URL <http://scitation.aip.org/content/aip/journal/jcp/108/17/10.1063/1.476137>. Cited on page 87
- [222] Auer, A. A., Helgaker, T. & Klopper, W. Accurate molecular geometries of the protonated water dimer. *Phys. Chem. Chem. Phys.*, 2:2235–2238, **2000**. URL <http://dx.doi.org/10.1039/A909189K>. Cited on pages 87 & 88
- [223] Wales, D. J. Rearrangements and tunneling splittings of protonated water dimer. *J. Chem. Phys.*, 110(21):10403–10409, **1999**. URL <http://scitation.aip.org/content/aip/journal/jcp/110/21/10.1063/1.478972>. Cited on pages 87, 88, 93, & 94
- [224] Xie, Y., Remington, R. B. & Schaefer, H. F. The protonated water dimer: Extensive theoretical studies of h_5o_2^+ . *J. Chem. Phys.*, 101(6):4878–4884, **1994**. URL <http://scitation.aip.org/content/aip/journal/jcp/101/6/10.1063/1.467409>. Cited on pages 87 & 88
- [225] Park, M., Shin, I., Singh, N. J. & Kim, K. S. Eigen and zundel forms of small protonated water clusters: Structures and infrared spectra. *J. Phys. Chem. A*, 111:10692–10702, **2007**. Cited on page 87
- [226] Schmitt, U. W. & Voth, G. A. Multistate empirical valence bond model for proton transport in water. *J. Phys. Chem. B*, 102:5547–5551, **1998**. Cited on page 87
- [227] Day, T. J. F., Soudackov, A. V., Åsuma, M., Schmitt, U. W. & Voth, G. A. A second generation multistate empirical valence bond model for proton transport in aqueous systems. *The Journal of Chemical Physics*, 117:5839–5849, **2002**. Cited on page 87
- [228] Kale, S. & Herzfeld, J. Proton defect solvation and dynamics in aqueous acid and base. *Angew. Chem. Int. Ed.*, 51:11029–11032, **2012**. Cited on page 87
- [229] Vendrell, O., Gatti, F., Lauvergnat, D. & Meyer, H.-D. Full-dimensional (15-dimensional) quantum-dynamical simulation of the protonated water dimer. i. hamiltonian setup and analysis of the ground vibrational state. *The Journal of Chemical Physics*, 127:184302, **2007**. Cited on page 87
- [230] 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.*, 144:054111, **2016**. Cited on page 87
- [231] Morokuma, K. & Pedersen, L. Molecular orbital studies of hydrogen bonds. an ab initio calculation for dimeric h_2o . 48:3275–3282, **1968**. Cited on page 111

-
- [232] Dyke, T. R., Mack, K. M. & Muentner, J. S. The structure of water dimer from molecular beam electric resonance spectroscopy. *The Journal of Chemical Physics*, 66:498–510, **1977**. Cited on page 111
- [233] Odutola, J. A. & Dyke, T. R. Partially deuterated water dimers: Microwave spectra and structure. *The Journal of Chemical Physics*, 72:5062–5070, **1980**. Cited on page 111
- [234] Curtiss, L. A., Frurip, D. J. & Blander, M. Studies of molecular association in h. *The Journal of Chemical Physics*, 71:2703–2711, **1979**. Cited on page 111
- [235] Mukhopadhyay, A., Cole, W. T. & Saykally, R. J. The water dimer i: Experimental characterization. *Chemical Physics Letters*, 633:13–26, **2015**. Cited on page 111
- [236] Mukhopadhyay, A., Xantheas, S. S. & Saykally, R. J. The water dimer ii: Theoretical investigations. *Chemical Physics Letters*, 700:163–175, **2018**. Cited on page 111
- [237] Lin, I.-C., Seitsonen, A. P., Coutinho-Neto, M. D., Tavernelli, I. & Rothlisberger, U. Importance of van der waals interactions in liquid water. *J. Phys. Chem. B*, 113:1127–1131, **2009**. Cited on page 112
- [238] Lane, J. R. Ccsdtq optimized geometry of water dimer. *J. Chem. Theory Comput.*, 9:316–323, **2012**. Cited on pages 119 & 120
- [239] Cukierman, S. Et tu, grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1757(8):876 – 885, **2006**. URL <http://www.sciencedirect.com/science/article/pii/S0005272805002768>. Proton Transfer Reactions in Biological Systems. Cited on page 123
- [240] Vuilleumier, R. & Borgis, D. Transport and spectroscopy of the hydrated proton: A molecular dynamics study. *The Journal of Chemical Physics*, 111:4251–4266, **1999**. Cited on page 123
- [241] Stoyanov, E. S., Stoyanova, I. V. & Reed, C. A. The structure of the hydrogen ion (h+aq) in water. *J. Am. Chem. Soc.*, 132:1484–1485, **2010**. Cited on page 123
- [242] Knight, C. & Voth, G. A. The curious case of the hydrated proton. *Acc. Chem. Res.*, 45:101–109, **2012**. Cited on pages 123, 124, & 129
- [243] Wang, Y.-S., Tsai, C.-H., Lee, Y. T., Chang, H.-C., Jiang, J. C., Asvany, O., Schlemmer, S. & Gerlich, D. Investigations of protonated and deprotonated water clusters using a low-temperature 22-pole ion trap. *J. Phys. Chem. A*, 107:4217–4225, **2003**. Cited on page 124
- [244] Heine, N., Fagiani, M. R., Rossi, M., Wende, T., Berden, G., Blum, V. & Asmis, K. R. Isomer-selective detection of hydrogen-bond vibrations in the protonated water hexamer. *J. Am. Chem. Soc.*, 135:8266–8273, **2013**. Cited on page 124
- [245] Wei, D. & Salahub, D. R. Hydrated proton clusters and solvent effects on the proton transfer barrier: A density functional study. *The Journal of Chemical Physics*, 101:7633–7642, **1994**. Cited on page 124
- [246] Napoli, J. A., Marsalek, O. & Markland, T. E. Decoding the spectroscopic features and time scales of aqueous proton defects. *The Journal of Chemical Physics*, 148:222833, **2018**. Cited on pages 124 & 141
- [247] Decka, D., Schwaab, G. & Havenith, M. A thz/ftir fingerprint of the solvated proton: evidence for eigen structure and zundel dynamics. *Phys. Chem. Chem. Phys.*, 17:11898–11907, **2015**. Cited on pages 129, 144, & 147
-

Bibliography

- [248] Bell, R. A., Christoph, G. G., Fronczek, F. R. & Marsh, R. E. The cation H_3O^+ : A short, symmetric hydrogen bond. *Science*, 190:151–152, **1975**. Cited on pages 131 & 134
- [249] Jones, D. J., Rozière, J., Penfold, J. & Tomkinson, J. Incoherent inelastic neutron scattering studies of proton conducting materials trivalent metal acid sulphate hydrates: Part i. the vibrational spectrum of H_2O^+ . *Journal of Molecular Structure*, 195:283 – 291, **1989**. URL <http://www.sciencedirect.com/science/article/pii/0022286089801759>. Cited on page 134
- [250] Stoyanov, E. S. & Reed, C. A. Ir spectrum of the H_2O^+ cation in the context of proton disolvates $\text{H}^+\text{H}_2\text{O}$. *J. Phys. Chem. A*, 110:12992–13002, **2006**. Cited on page 134
- [251] Stoyanov, E. S., Stoyanova, I. V., Tham, F. S. & Reed, C. A. H. *J. Am. Chem. Soc.*, 131:17540–17541, **2009**. Cited on page 134
- [252] Ciobanu, C. V., Ojamäe, L., Shavitt, I. & Singer, S. J. Structure and vibrational spectra of $\text{H}^+(\text{H}_2\text{O})_8$: Is the excess proton in a symmetrical hydrogen bond? *J. Chem. Phys.*, 113:5321, **2000**. Cited on page 141
- [253] Kornyshev, A. A., Kuznetsov, A. M., Spohr, E. & Ulstrup, J. Kinetics of proton transport in water. *J. Phys. Chem. B*, 107:3351–3366, **2003**. Cited on page 141
- [254] Tuckerman, M. E., Chandra, A. & Marx, D. A statistical mechanical theory of proton transport kinetics in hydrogen-bonded networks based on population correlation functions with applications to acids and bases. *The Journal of Chemical Physics*, 133:124108, **2010**. Cited on pages 150 & 159
- [255] Markovitch, O., Chen, H., Izvekov, S., Paesani, F., Voth, G. A. & Agmon, N. Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B*, 112(31):9456–9466, **2008**. URL <http://pubs.acs.org/doi/abs/10.1021/jp804018y>. Cited on page 154
- [256] Wilkins, D. M., Manolopoulos, D. E., Pipolo, S., Laage, D. & Hynes, J. T. Nuclear quantum effects in water reorientation and hydrogen-bond dynamics. *J. Phys. Chem. Lett.*, 8:2602–2607, **2017**. Cited on pages 154 & 156
- [257] Habershon, S., Markland, T. E. & Manolopoulos, D. E. Competing quantum effects in the dynamics of a flexible water model. *The Journal of Chemical Physics*, 131:024501, **2009**. Cited on page 154
- [258] Roberts, S. T., Petersen, P. B., Ramasesha, K., Tokmakoff, A., Ufimtsev, I. S. & Martinez, T. J. Observation of a zundel-like transition state during proton transfer in aqueous hydroxide solutions. *Proceedings of the National Academy of Sciences*, 106:15154–15159, **2009**. Cited on page 158
- [259] Wang, Y.-F. & Cheng, Y.-C. Molecular electrostatic potential on the proton-donating atom as a theoretical descriptor of excited state acidity. *Phys. Chem. Chem. Phys.*, 20:4351–4359, **2018**. Cited on page 158
- [260] Scharf, D., Klein, M. L. & Martyna, G. J. Pathâintegral monte carlo studies of. *The Journal of Chemical Physics*, 97:3590–3599, **1992**. Cited on page 160
- [261] Vañshteñ, A. I., Zakharov, V. I., Novikov, V. A. & Shifman, M. A. Abc of instantons. *Sov. Phys. Usp.*, 25:195–215, **1982**. Cited on page 160

- [262] Jiang, Z., Smelyanskiy, V. N., Isakov, S. V., Boixo, S., Mazzola, G., Troyer, M. & Neven, H. Scaling analysis and instantons for thermally assisted tunneling and quantum monte carlo simulations. *Phys. Rev. A*, 95:382, **2017**. *Cited on page 160*
- [263] Mazzola, G., Smelyanskiy, V. N. & Troyer, M. Quantum monte carlo tunneling from quantum chemistry to quantum annealing. *Phys. Rev. B*, 96:483, **2017**. *Cited on page 160*
- [264] Richardson, J. O. & Althorpe, S. C. Ring-polymer molecular dynamics rate-theory in the deep-tunneling regime: Connection with semiclassical instanton theory. *The Journal of Chemical Physics*, 131:214106, **2009**. *Cited on page 160*
- [265] Richardson, J. O. Ring-polymer instanton theory of electron transfer in the nonadiabatic limit. *J. Chem. Phys.*, 143:134116, **2015**. *Cited on page 160*
- [266] Fang, W., Chen, J., Rossi, M., Feng, Y., Li, X.-Z. & Michaelides, A. Inverse temperature dependence of nuclear quantum effects in dna base pairs. *J. Phys. Chem. Lett.*, 7:2125–2131, **2016**. *Cited on page 161*

Résumé

Il n'existe encore aucune théorie aujourd'hui capable de proposer une description précise et quantitative du transfert de proton en solution. En effet, la complexité de ce problème provient de la grande diversité des interactions existant dans l'eau liquide, à savoir : des interactions non liantes de type Van der Waals, des liaisons faiblement covalentes et des liaisons hydrogènes remarquablement fortes. Ces dernières sont à l'origine des nombreuses propriétés fascinantes de l'eau à l'échelle macroscopique. À cela s'ajoutent les effets quantiques nucléaires dus à la faible masse de l'hydrogène, qui modifient profondément la nature de la surface d'énergie potentielle et les effets thermiques. Dans cette thèse, nous proposons une approche tout quantique basée sur une description quasi exacte de la fonction d'onde électronique par l'utilisation de méthodes Monte Carlo Quantique (QMC). Notre nouvelle technique combine le QMC avec une dynamique moléculaire de type Langevin en utilisant le formalisme des intégrales de chemin de Feynman. Ceci permet de faire des simulations tout quantique avec une précision inédite de systèmes de taille relativement grande en phase gaz ou en solution, à température finie. Nous appliquons notre approche à des agrégats d'eau neutres ou protonés pour apporter de nouveaux éclaircissements sur les phénomènes microscopiques régissant la diffusion du proton dans de tels systèmes. En particulier, nous avons étudié l'hexamère d'eau protoné comme modèle décrivant un proton en excès en solution aqueuse. Nous avons découvert que le taux de transfert de proton est optimal pour des températures proches des conditions ambiantes, du fait de la compétition subtile entre les effets thermiques et quantiques nucléaires.

Abstract

There is no theory up to now able to provide an accurate and quantitative description of the proton transfer (PT) yet. Indeed, the complexity of this problem stems from the large diversity of the existing interactions in liquid water, namely: non bonding Van der Waals interactions, weakly covalent bonds and remarkably strong H-bonds. The latter ones are at the origin of the numerous fascinating properties of water at the macroscopic scale. In addition to such interactions, the nuclear quantum effects arising from the hydrogen light mass deeply modify the potential energy surface and thermal effects, and must be taken into account. In this thesis, we propose a fully quantum approach based on an almost exact description of the electronic wave function by means of Quantum Monte Carlo (QMC) methods. Our novel technique combines QMC with a Langevin-based Molecular Dynamics and the Feynman path integral formalism. This allows one to perform fully quantum simulations with an unprecedented level of accuracy of moderate size systems in gas or condensed phase, at finite temperature. We apply our approach to neutral or charged protonated water clusters to shed light on the microscopic phenomena driving the proton diffusion in such systems. In particular, we have studied the protonated water hexamer as a model describing an excess proton in aqueous solution. We discovered that the proton hopping is optimal for temperatures close to ambient conditions, due to the subtle competition between thermal and nuclear quantum effects.

