

Nuclear Quantum Effects: Fast and Accurate

Présentée le 4 septembre 2020

à la Faculté des sciences et techniques de l'ingénieur
Laboratoire de science computationnelle et modélisation
Programme doctoral en science et génie des matériaux

pour l'obtention du grade de Docteur ès Sciences

par

Venkat KAPIL

Acceptée sur proposition du jury

Prof. D. Damjanovic, président du jury
Prof. M. Ceriotti, directeur de thèse
Prof. M. Tuckerman, rapporteur
Prof. S. Althorpe, rapporteur
Prof. S. Bonella, rapporteuse

Freedom from the desire for an answer is
essential to the understanding of a problem.

— Jiddu Krishnamurti

Acknowledgements

I would like to thank my supervisor Prof. Michele Ceriotti for giving me an opportunity of working in an exciting and challenging research area. His enthusiasm for research has been really inspiring, and has motivated me to give my hundred percent to my projects throughout the PhD. As a supervisor, he struck the balance by providing timely feedback and guidance, but also gave me space to approach problems and to develop my understanding of the research area in an independent manner. I also thank him for giving me ample opportunities to attend international schools and conferences, as well as for his invaluable suggestions for writing grant and fellowship applications.

I would like to thank my parents Mr. Shelleyandra Kapil, Dr. Mrs. Rupa Kapil and my sister Dr. Mrs. Vidushi Kapil, for their constant support. I would also like to thank all my relatives, especially my cousin Dr. Tejus Behl, for always making me feel like I was at home. I also thank my colleagues at the Laboratory of computational Science and Modelling (COSMO) for providing a fun and productive workplace, and my friends at Lausanne, and across the world, for all the fun times that we experienced together (both online and offline). A special mention to Andrea Anelli, Felix Musil, Dr. Max Veit, Dr. Soham Basu, Yash Singh and Tushar Thakur for the endless fun discussions over coffee and drinks.

I wish to also thank Dr. David Wilkins and Dr. Edgar Engel, for always being there for proof-reading my publications and documents related to fellowships. With Dr. Edgar Engel, I had the pleasant opportunity to work on a very difficult and long project, and I enjoyed learning from him, how to solve a big problem by breaking it into many small ones. I also enjoyed several very insightful discussion on properties and free energies of solids. I am really glad that I had the opportunity to learn from him.

Acknowledgements

In the summer of 2019, I went through some medical problems, and I would really like to thank Prof. Michele Ceriotti, Prof. Angelos Michaelides, and Dr. Mariana Rossi for supporting me during that time and for motivating me to not abandon writing my fellowship application. A big thanks to Jinggang Lan, who helped me during that time to write a grant for computational resources. I also thank my doctor Dr. Christian Winckler for speedy and effective treatment, and also for keeping me motivated. In this regard, I would also like to thank all my friends and colleagues – Nitica Sakharwade, Shreyan Chowdhury, Giulio Imbalzano, Andrea Anelli, Felix Musil, Dr. Max Veit, Natasha Lopanitsyna, Dr. Edgar Engel, Harshit Gupta, Dr. Soham Basu, and Tushar Thakur – for their help during that time.

And finally, I also thank Haosong Jiao for her constant care and support, especially for enduring me when I was at my worst during stressful times, and for making my time in Switzerland as memorable as it has been.

Abstract

Atomistic simulations are a bottom up approach that predict properties of materials by modelling the quantum mechanical behaviour of all electrons and nuclei present in a system. These simulations, however, routinely assume nuclei to be classical particles, which leads to incorrect predictions for systems that exhibit significant quantum delocalization and zero-point effects, such as those containing light nuclei. The path integral approach, the state of the art approach that models the exact quantum thermodynamics of this class of systems, is much more computationally expensive and harder to implement than the classical methods that evolve the system using classical statistical mechanics. This has prevented widespread modelling of the quantum mechanics of nuclei in atomistic simulations, especially in combination with computationally expensive interatomic potentials that model interparticle interactions at a high level of theory.

In this thesis, we present several new methods that dramatically reduce the computational cost of modelling the quantum nature of nuclei with respect to standard methods, and to existing cost reduction schemes. These methods are based on the realization that nuclear quantum effects can often be modelled using cheap short ranged interaction potential, or using high order splittings that decouple non-commuting potential and kinetic energy operators, or using generalized Langevin equations that can be used to mimic quantum fluctuations with correlated noise. We have also derived bespoke estimators of quantities such as the quantum heat capacity, the particle momentum distribution, and vibrational spectra that reduce the cost of calculating these properties, and allow direct comparisons with experiments. These methods have been implemented in the second release of an open source software `i-PI`, which allows them to be used in combination with widely used softwares that compute interatomic potentials. The availability of these methods has promoted routine incorporation

Abstract

of nuclear quantum effects in atomistic simulations. The relevance of these advances is underscored by the different properties and classes of materials to which we have applied these methods. For instance, we have computed PMD in different phases of water facilitating interpretation of Deep Inelastic Neutron Scattering experiments, and the understanding of the local environments of protons. Similarly, we have shown how the interplay of quantum effects and intermolecular interactions can be used to tune the heat capacity of methane loaded metal-organic frameworks, to increase, decrease or stay constant over a range of temperatures. We have also studied the impact of NQEs in affecting stabilities of pharmaceutically active molecular crystals using several computationally inexpensive methods that are routinely used to approximate quantum free energies. We have systematically studied their accuracy on a large set of solids, and concluded that free energy calculations that include the quantum nuclear motion exactly are the only reliable by way of predicting stabilities of molecular crystals.

List of publications

Publications directly related to this thesis

1. V. Kapil, J. VandeVondele, and M. Ceriotti, "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," *The Journal of Chemical Physics*, vol. 144, p. 054111, Feb. 2016
2. V. Kapil, J. Behler, and M. Ceriotti, "High order path integrals made easy," *The Journal of Chemical Physics*, vol. 145, p. 234103, Dec. 2016
3. M. Rossi, V. Kapil, and M. Ceriotti, "Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation," *The Journal of Chemical Physics*, vol. 148, p. 102301, July 2017
4. V. Kapil, A. Cuzzocrea, and M. Ceriotti, "Anisotropy of the Proton Momentum Distribution in Water," *The Journal of Physical Chemistry B*, vol. 122, pp. 6048–6054, June 2018

5. V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. V. Speybroeck, and M. Ceriotti, “i-PI 2.0: A universal force engine for advanced molecular simulations,” *Comput. Phys. Commun.*, vol. 236, pp. 214–223, oct 2019
6. V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. Van Speybroeck, and M. Ceriotti, “Modeling the Structural and Thermal Properties of Loaded Metal–Organic Frameworks. An Interplay of Quantum and Anharmonic Fluctuations,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 3237–3249, May 2019
7. V. Kapil, E. Engel, M. Rossi, and M. Ceriotti, “Assessment of Approximate Methods for Anharmonic Free Energies,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 5845–5857, Nov. 2019
8. V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, “Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats,” *The Journal of Chemical Physics*, vol. 152, p. 124104, Mar. 2020

Other publications submitted during the doctorate

1. K. Druzbicki, M. Krzystyniak, D. Hollas, V. Kapil, P. Slaviček, G. Romanelli, and F. Fernandez-Alonso, “Hydrogen dynamics in solid formic acid: Insights from simulations with quantum colored-noise thermostats,” *Journal of Physics: Conference Series*, vol. 1055, p. 012003, July 2018
2. J. Wieme, S. Vandenbrande, A. Lemaire, V. Kapil, L. Vanduyfhuys, and V. V. Speybroeck, “Thermal engineering of metal-organic frameworks for adsorption applications: A molecular simulations perspective,” *ACS Applied Materials & Interfaces*, 2019
3. M. Liu, L. Zhang, M. A. Little, V. Kapil[†], M. Ceriotti, S. Yang, L. Ding, D. L. Holden, R. Balderas-Xicohtencatl, D. He, R. Clowes, S. Y. Chong, G. Schütz, L. Chen, M. Hirscher, and A. I. Cooper, “Barely porous organic cages for hydrogen isotope separation,” *Science*, vol. 366, pp. 613–620, Oct. 2019

[†] first computational author

Abstract

Publications submitted before the doctorate

1. S. Awasthi, V. Kapil, and N. N. Nair, “Sampling free energy surfaces as slices by combining umbrella sampling and metadynamics,” *Journal of Computational Chemistry*, vol. 37, pp. 1413–1424, Apr. 2016

Contents

Acknowledgements	i
Abstract (English/Français/Deutsch)	iii
1 Introduction	1
1.1 Approximations in atomistic simulations	3
1.2 Classical thermodynamics	6
1.2.1 Molecular Dynamics for sampling classical thermodynamic ensembles .	7
1.2.2 Molecular dynamics for calculating classical dynamical properties . . .	11
1.2.3 Limitations of classical thermodynamics	13
1.3 Quantum thermodynamics	14
1.3.1 Path integral molecular dynamics for sampling quantum ensembles . .	17
1.3.2 Path integral methods for approximate quantum dynamics	22
1.3.3 Challenges associated with the path integral method	26
I Methodological Developments	29
2 Accelerated path integral methods	31
2.1 Exploiting multiple time / length scales	32
2.1.1 Ring polymer contraction (RPC)	32
2.1.2 Multiple timestepping (MTS)	33
2.1.3 Combining MTS and RPC	34
2.1.4 Benchmarks on q-TIP4P/f water	35
2.1.5 MTS / RPC combined with electronic structure methods	38
2.2 High-order path integrals	42
	vii

Contents

2.2.1	Challenges associated with high-order path integrals	43
2.2.2	Estimating high-order forces and virials	44
2.2.3	MTS integrators for high-order path integrals	45
2.2.4	A GLE thermostat for high-order path integrals	47
2.2.5	Benchmarks on water using a dissociable potential	49
3	Efficient calculation of complicated observables	59
3.1	Heat capacity	59
3.1.1	Estimators of heat capacity using standard PIMD	60
3.1.2	Estimators of heat capacity using high-order PIMD	62
3.1.3	Benchmarks on liquid water	64
3.2	Particle momentum distribution	65
3.2.1	Open paths for calculating the particle momentum distribution	65
3.2.2	A convenient algorithm for implementing open PIMD	67
3.2.3	An efficient estimator for the momentum distribution	68
3.2.4	Benchmarks on different phases of water	70
4	Accelerated methods for vibrational spectra	71
4.1	A Generalized Langevin Equation Thermostat	73
4.1.1	Controlling Classical Dynamics	74
4.1.2	Predicting and correcting the dynamical disturbance of a GLE	78
4.1.3	Dynamical properties from a quantum thermostat	82
4.1.4	Dynamical properties from a PIGLET thermostat	84
II	Applications	89
5	Anisotropy of the Proton Momentum Distribution in Water	91
5.1	Computational setup	93
5.2	Quantum kinetic energies	94
5.3	Anisotropic proton momentum distributions	95
5.4	Constant pressure ensemble	96
5.5	Discussion	97

6	Quantum spectroscopy of Water	99
6.1	Machine learning models	100
6.2	IR and Raman spectra	103
6.3	Discussion	106
7	Modeling structural and thermal properties of loaded metal-organic frameworks	107
7.1	Materials	110
7.2	First-principles-derived force fields	110
7.3	Calculation of thermodynamic observables	111
7.4	Computational Details	111
7.5	Structural properties	113
7.5.1	Heat Capacity	116
7.5.2	The interplay of gas loading, anharmonicities, and quantum effects . . .	121
7.6	Discussion	122
8	Assessment of approximate methods for anharmonic free energies	125
8.1	Theory	126
8.1.1	Harmonic Approximation	127
8.1.2	Independent Mode Framework	129
8.1.3	Vibrational Self-consistent Field	130
8.1.4	Self-consistent Phonons	131
8.1.5	Thermodynamic Integration	132
8.2	Implementation	134
8.2.1	Independent Mode Framework	135
8.2.2	Vibrational Self-Consistent Field	136
8.2.3	Self-Consistent Phonons	136
8.3	Results	140
8.3.1	Systems and Computational Details	141
8.3.2	Computational cost	142
8.3.3	Accuracy	144
8.4	Discussion	150

Contents

9 Conclusions	153
Bibliography	159
Curriculum Vitae	199

1 Introduction

The importance of materials is well understood and it is no surprise that several landmarks of our history – such as the stone, bronze and iron ages – are remembered by the materials used during those periods. The current age that we live in is often referred to as the Silicon age, due to the widespread use of Silicon in the electronics industry^[13]. Similarly, solutions of contemporary problems ranging from antibiotic resistance^[14], lack of sustainable and clean sources of energy^[15], to the saturation of Moore's law^[16] rely heavily on a concrete understanding of materials properties.

For centuries, the design and discovery of materials has been largely driven by experimental studies, as the use of theory has been limited by our inability to solve the complex set of equations that govern the physics of materials^[17], and the processes they are involved in. However, since the last few decades, "Computation" is being viewed as the third pillar of science^[18] after theory and experiments, and due to the ever increasing availability of computational resources, it is becoming possible to simulate the physics of materials by numerically solving equations that were once considered intractable^[19,20]. Within this perspective, the use of atomistic simulations^[21,22] – a computational strategy which aims at simulating the properties of materials by modelling the quantum mechanics of all electrons and nuclei present in a system – shows great potential in guiding the design and discovery of material. These simulations are typically used to study systems at the length scale of nanometers, but provide access to most thermodynamic as well as transport properties of materials, in a rigorous manner^[21].

Chapter 1. Introduction

These simulations also provide the foundational step towards multi-scale modelling^[23] that use properties computed using atomistic simulations to simulate phenomena at larger length and timescales.

In this thesis, we have aimed to improve the accuracy of atomistic simulations by developing computational methods that go beyond standard Newtonian dynamics for evolving the positions of atoms, and rigorously account for quantum mechanical effects in their motion. We begin by introducing the fundamentals of atomistic simulations^[22,21] in Chapter 1, while focusing on the molecular dynamics approach for simulating classical thermodynamics of systems, and the errors it incurs while modeling systems that exhibit quantum behaviour. We then discuss Richard Feynman’s imaginary time path integral formulation of quantum statistical mechanics^[24], which allows one to model quantum nuclear motion in atomistic simulations, although at a high computational cost. We further examine the limitations of the path integral method, and the challenges that are needed to overcome to facilitate low cost modeling of a wide variety of physical observables.

In the next part of the thesis labeled “Methodological developments”, we present Chapters 2-4, that discuss an array of novel techniques^[1,2,6] that substantially reduce the computational cost of path integral methods, and novel estimators and methodologies for the estimation of complicated observables – quantum heat capacities^[6], particle momentum distributions^[4] (PMDs) and vibrational spectra^[8] – that have historically been considered challenging to compute. We also discuss the implementation of the methods in the second release of the open source software *i-PI*^[5].

We proceed to the part of the thesis labeled “Applications”, that includes Chapters 5-8, and present studies of a wide range of molecular materials ranging from fundamentally important phases of water to technologically relevant porous materials and pharmaceutically active molecular crystals. In Chapter 5, we computationally benchmark^[4] deep inelastic neutron scattering experiments^[25], that calculate PMDs of different phases of water and highlight cer-

tain anomalous properties^[26]. We show a good agreement between theory and experiments for certain phases of water by computing PMDs at unprecedented accuracy, and comment on the origin of remaining disagreements. In Chapter 6, we compute IR and Raman spectra of solid and liquid phases of water, using the methodology developed in Chapter 5 in combination with machine-learning predictions of dielectric responses. Our results show an overall good agreement with experimental results, and pave the way for modelling more advanced spectroscopic properties such as second-harmonic scattering, and sum-frequency generation. In Chapter 7, we study the quantum heat capacity of metal organic frameworks^[6], that show promise in the absorption of green house gases. We disentangle several components of the heat capacity as a function of gas loading into components arising from host and guest interactions, and present mechanism that can be utilized to tune the total heat capacity of the full system to increase, decrease or stay constant over a range of temperature. In Chapter 8, we study the role of the quantum nature of nuclei in stabilizing a wide range of solids^[7], including pharmaceutically relevant ones like paracetamol. Our results show that rigorous free energy calculations which include quantum motion of atoms are the *only* reliable and accurate way of studying free energy differences in crystal structure prediction applications of complex molecular solids. In Chapter 9, we summarize the developments and applications presented in the thesis, and discuss the possibility of applying the developed framework to the holy grails of computational chemistry and materials science.

Overall, our work has led to the development of a toolkit that allows efficient and accurate simulations at a low computational cost, which makes the study of quantum mechanical effects in a wide class of materials affordable, and systematically improves the accuracy and reliability of atomistic simulations in the prediction of materials properties.

1.1 Approximations in atomistic simulations

The theory of quantum mechanics which treats electrons and nuclei as fuzzy objects behaving like both waves and particles, instead of rigid billiard balls, is one of the most accurate theories of physics^[17]. In principle it allows most of the properties of materials to be estimated exactly

provided that its fundamental equation – the time dependent Schrödinger’s Equation^[27] (SE) – can be solved for all the electrons and nuclei present in a system. Unfortunately, the exact solution to the SE is possible only for the smallest and simplest of systems, due to which modern day atomistic simulations are based on approximate “quantum” descriptions of atoms^[21,22,28,29,30,31,32,33]. One of the biggest challenges in the field of atomistic simulations is to reduce the number of underlying approximations and to develop computationally efficient methods that enable a high quality description of the quantum mechanics of electrons and nuclei.

One of the foundational approximations behind modern day atomistic simulations is the Born-Oppenheimer (BO) approximation^[34], which relies on the large difference in the electronic and nuclear masses, and allows one to decouple the full SE into those for electrons and nuclei. Basically, the nuclear coordinates enter the electronic SE as constants given that the timescale of the nuclear motion is much larger than that of electrons. This allows one to define a BO potential energy landscape – the hypersurface of the lowest electronic energy eigenvalue at given nuclear coordinates – which governs nuclear motion. This approximation remains valid in most of the cases that are encountered in chemistry and materials science, except when small changes in the nuclear coordinates lead to large changes in the electronic wavefunctions^[35], during phenomena such as photo dissociation^[36] and electron transfer reactions^[37]. The remaining problem boils down to solving the SE for the nuclear degrees of freedom described by the Hamiltonian:

$$\hat{H} = \hat{T} + \hat{V} = \sum_{i=1}^N \frac{\hat{\mathbf{p}}_i^2}{2m_i} + V(\hat{\mathbf{q}}_1, \dots, \hat{\mathbf{q}}_N, \mathbf{h}) \quad (1.1)$$

where, \hat{T} and \hat{V} represent the kinetic and the BO potential energy operators, $\hat{\mathbf{p}}_i$, $\hat{\mathbf{q}}_i$ and m_i are respectively the momentum and position operators, and the mass of the i^{th} nucleus, $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ represent the momentum and the position of the full system, and \mathbf{h} is the cell tensor that limits the region occupied by the N nuclei. The eigenstates $|s; \mathbf{h}\rangle$ and the corresponding vibrational energy states $E_{(s)}(\mathbf{h})$ are estimated by solving the time independent SE

$$\hat{H}|s; \mathbf{h}\rangle = E_{(s)}(\mathbf{h})|s; \mathbf{h}\rangle, \quad (1.2)$$

and subsequently the theory of statistical mechanics can be used to estimate thermodynamic properties of the system by estimating the corresponding partition functions. For instance, the canonical partition function $Z(N, \mathcal{V}, \beta)$ at inverse temperature β and volume $\mathcal{V} = \text{Det}[\mathbf{h}]$, assuming that the nuclei are distinguishible, can be computed as

$$Z(N, \mathcal{V}, \beta) = \text{Tr}[\exp(-\beta \hat{H})] = \sum_s \exp(-\beta E_{(s)}(\mathbf{h})), \quad (1.3)$$

while the isothermal-isobaric partition function $\Delta(N, \mathcal{P}^{\text{ext}}, \beta)$ at inverse temperature β and external pressure \mathcal{P}^{ext} can be computed as:

$$\Delta(N, \mathcal{P}^{\text{ext}}, \beta) = \mathcal{V}_0^{-1} \int d\mathcal{V} \exp(-\beta \mathcal{P}^{\text{ext}} \mathcal{V}) Z(N, \mathcal{V}, \beta), \quad (1.4)$$

where \mathcal{V}_0 is a reference volume that ensures that the partition function is dimensionless. Every thermodynamic property can be derived from the corresponding partition functions by taking derivatives with respect to applied external fields in the limit of large N , also known as the thermodynamic limit. However, such calculations suffer from three notable limitations:

1. The typical length scales that are accessible in such calculations may not be enough to reach the thermodynamic limit. To eliminate this error it is essential to perform calculations with increasing N until convergence in the property of interest is achieved^[21].
2. The electronic SE is typically not solved in an exact manner as it is (so far) not possible to obtain exact results for an arbitrary many-body correlated system. Consequently, the errors while approximating the BO potential energy landscape^[28,29,30,31,32,33] can propagate into errors in thermodynamic properties. To reduce these errors it is essential to solve the electronic SE at a high level of theory.
3. It is also not possible to solve the nuclear SE and thus obtain the vibrational energy levels in an exact manner. Moreover, since the number of thermally accessible vibrational levels is much higher than the number of populated electronic levels, it is impractical to solve for all accessible excited states. Instead, a more efficient approach should be to sample the energy levels with a probability that is related to their population. However, the problem generally is typically approached from a completely different perspective.

1.2 Classical thermodynamics

Since the mass of the nuclei is at least three orders of magnitude higher than that of electrons, the former are routinely treated as classical particles. One usually proceeds by computing the trace in Eq. 1.3 in the position basis^[21] ¹

$$Z(N, \mathcal{V}, \beta) = \text{Tr} [\exp(-\beta \hat{H})] = \int d\mathbf{q} \langle \mathbf{q} | \exp(-\beta \hat{H}) | \mathbf{q} \rangle = \int d\mathbf{q} \langle \mathbf{q} | \exp(-\beta(\hat{T} + \hat{V})) | \mathbf{q} \rangle, \quad (1.5)$$

and assumes that the kinetic and potential energies commute, thus violating a fundamental principle of quantum mechanics

$$\begin{aligned} Z^{(\text{cl})}(N, \mathcal{V}, \beta) &= \int d\mathbf{q} \langle \mathbf{q} | \exp(-\beta \hat{T}) \exp(-\beta \hat{V}) | \mathbf{q} \rangle = \int d\mathbf{q} \langle \mathbf{q} | \exp(-\beta \hat{T}) | \mathbf{q} \rangle \exp(-\beta V(\mathbf{q}, \mathbf{h})) \\ &= \left(\frac{1}{2\pi\hbar} \right)^{3N} \prod_{i=1}^N (2\pi m_i \beta^{-1})^{\frac{3}{2}} \int d\mathbf{q} \exp(-\beta V(\mathbf{q}, \mathbf{h})) \\ &= \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\mathbf{p} d\mathbf{q} \exp \left(-\beta \left(\sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + V(\mathbf{q}, \mathbf{h}) \right) \right) \\ &= \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\mathbf{p} d\mathbf{q} \exp(-\beta H(\mathbf{p}, \mathbf{q})). \end{aligned} \quad (1.6)$$

Here, \mathbf{p} are a set of auxiliary momenta and $H(\mathbf{p}, \mathbf{q})$ is the classical Hamiltonian of the physical system. Note that the momentum part of the Hamiltonian, which can be factored out by computing Gaussian integrals, only contributes trivially to the partition function and in general does not affect thermodynamic averages of position dependent observables

$$\begin{aligned} \langle A(\mathbf{q}) \rangle^{(\text{cl})} &= -(\beta Z^{(\text{cl})}(N, \mathcal{V}, \beta))^{-1} \frac{\partial}{\partial \lambda} Z^{(\text{cl})}(N, \mathcal{V}, \beta)_{V(\mathbf{q}) \rightarrow V(\mathbf{q}) + \lambda A(\mathbf{q})} \Big|_{\lambda=0} \\ &= \frac{\int d\mathbf{p} d\mathbf{q} A(\mathbf{q}) \exp(-\beta H(\mathbf{p}, \mathbf{q}))}{\int d\mathbf{p} d\mathbf{q} \exp(-\beta H(\mathbf{p}, \mathbf{q}))} = \frac{\int d\mathbf{q} A(\mathbf{q}) \exp(-\beta V(\mathbf{q}, \mathbf{h}))}{\int d\mathbf{q} \exp(-\beta V(\mathbf{q}, \mathbf{h}))}. \end{aligned} \quad (1.7)$$

A brute force numerical evaluation of such integrals has a cost that scales exponentially with the number of nuclei N , rendering such calculations prohibitively expensive for most systems

¹The trace of the Boltzmann operator is invariant w.r.t the choice of the basis set provided that it is complete

of interest. A practical way of estimating such integral is using the “importance sampling” Monte Carlo method^[38] that draws samples $\mathbf{q}(t')$ from the normalized Boltzmann probability $\propto \exp(-\beta V(\mathbf{q}, \mathbf{h}))$, and estimates thermodynamic averages as time averages:

$$\langle A(\mathbf{q}) \rangle^{(\text{cl})} = \frac{\int d\mathbf{q} A(\mathbf{q}) \exp(-\beta V(\mathbf{q}, \mathbf{h}))}{\int d\mathbf{q} \exp(-\beta V(\mathbf{q}, \mathbf{h}))} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' A(\mathbf{q}(t')), \quad (1.8)$$

under the assumption known as the ergodic hypothesis that assumes that the random process used to sample $\mathbf{q}(t')$ covers all the regions with a high Boltzmann probability within the domain of the positions constrained by the volume \mathcal{V} .

1.2.1 Molecular Dynamics for sampling classical thermodynamic ensembles

While the key algorithm to perform “importance sampling” is the Metropolis Markov Chain Monte Carlo method^[39], we will focus mainly on the molecular dynamics (MD) method^[19]. This is because the former either requires system dependent heuristics or has an efficiency that either scales poorly with system size; MD on the other hand can be applied to any arbitrary potential energy landscape and doesn’t have a scaling dependence on the system size aside from the cost of computing the potential, forces and the virial tensor.

The basic idea behind MD is to define “equations of motion” (EOMs) of the system (and of a bath) that conserve the thermodynamic ensemble of interest. The EOMs are then numerically integrated using a fixed timestep $\Delta t'$ to yield a discretized time series of the momentum and position $(\mathbf{p}(t'), \mathbf{q}(t'))$ over which any observable can be averaged. Within the ergodic hypothesis and in the limit of $\Delta t' \rightarrow 0$, the MD method is exact.

Sampling the classical canonical distribution

Let us focus on the canonical ensemble; the simplest set of equations of motion that conserve the canonical ensemble are the Hamilton’s EOMs:

$$\dot{\mathbf{p}}_i = \mathbf{f}_i, \quad \dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m_i}, \quad (1.9)$$

Chapter 1. Introduction

where, $\mathbf{f} = -\nabla_{\mathbf{q}} V(\mathbf{q}, \mathbf{h})$ is the total force acting on the system. To study the conservation of the ensemble, it is useful to define the Liouvillian

$$iL^{(\text{cl})} = \sum_{i=1}^N \left[\dot{\mathbf{p}}_i \frac{\partial}{\partial \mathbf{p}_i} + \dot{\mathbf{q}}_i \frac{\partial}{\partial \mathbf{q}_i} \right] = \sum_{i=1}^N \left[\mathbf{f}_i \frac{\partial}{\partial \mathbf{p}_i} + m_i^{-1} \mathbf{p}_i \frac{\partial}{\partial \mathbf{q}_i} \right] \quad (1.10)$$

which is essentially the total derivative with respect to time^[22]. It is easy to show that the canonical Boltzmann distribution is stationary with respect to the Liouvillian of Eq. 1.9 i.e. $\frac{\partial}{\partial t} \exp(-\beta H(\mathbf{p}, \mathbf{q})) = iL^{(\text{cl})} \exp(-\beta H(\mathbf{p}, \mathbf{q})) = 0$. Another use of the Liouvillian is to define the classical time evolution operator $(\mathbf{p}(t' + \Delta t'), \mathbf{q}(t' + \Delta t')) = \exp(iL^{(\text{cl})} \Delta t') (\mathbf{p}(t'), \mathbf{q}(t'))$ which can be approximated to propagate the positions and velocities over one step of the simulation^[40]. It remains useful to split Eq. 1.10 into the Liouvillian's associated with the momentum $iL_{\mathbf{p}}^{(\text{cl})} = \sum_{i=1}^N \mathbf{f}_i \frac{\partial}{\partial \mathbf{p}_i}$, and with the position $iL_{\mathbf{q}}^{(\text{cl})} = \sum_{i=1}^N m_i^{-1} \mathbf{p}_i \frac{\partial}{\partial \mathbf{q}_i}$, and decouple the propagators of the position and momentum using the second-order Trotter scheme^[40]

$$\begin{aligned} (\mathbf{p}(t' + \Delta t'), \mathbf{q}(t' + \Delta t')) &= \exp(iL_{\mathbf{p}}^{(\text{cl})} \Delta t' + iL_{\mathbf{q}}^{(\text{cl})} \Delta t') (\mathbf{p}(t'), \mathbf{q}(t')) \\ &\approx \exp(iL_{\mathbf{p}}^{(\text{cl})} \Delta t' / 2) \exp(iL_{\mathbf{q}}^{(\text{cl})} \Delta t') \exp(iL_{\mathbf{p}}^{(\text{cl})} \Delta t' / 2) (\mathbf{p}(t'), \mathbf{q}(t')). \end{aligned} \quad (1.11)$$

This amounts to the update of the momentum for half a timestep, followed by the update of the position for a full timestep, concluded by the update of momentum for another half timestep

$$\begin{aligned} \mathbf{p}_i &\rightarrow \mathbf{p}_i + \mathbf{f}_i \Delta t' / 2 \\ \mathbf{q}_i &\rightarrow \mathbf{q}_i + \mathbf{p}_i \Delta t' / m_i \\ \mathbf{p}_i &\rightarrow \mathbf{p}_i + \mathbf{f}_i \Delta t' / 2. \end{aligned} \quad (1.12)$$

The integration scheme thus obtained, popularly known as velocity-Verlet scheme^[41], conserves properties of Hamiltonian dynamics such as time reversibility, and symplecticity.

A major limitation of Hamiltonian dynamics is that the equations of motion are not necessarily ergodic. For instance, for a harmonic potential a system subject to Hamiltonian dynamics will keep moving periodically on the isosurface of the energy, and not sample all the relevant regions of position's domain. To avoid such problems, one typically applies a "thermostat" which means either coupling the EOMs of the system to those of some "bath" degrees of

freedom, or modifying the EOM of momentum to make the dynamics more chaotic. An easy way to ensure ergodicity is to apply a Langevin thermostat^[42] which yields the following EOMs

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \gamma \mathbf{p}_i + \sqrt{2\gamma m_i \beta} \mathbf{R}_i, \quad \dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m_i}, \quad (1.13)$$

where, γ is a friction parameter, and \mathbf{R}_i is a random force drawn from a three dimensional normal distribution. The corresponding Liouvillian is^[42]

$$iL^{(\text{cl-LE})} = iL^{(\text{cl})} + \sum_{i=1}^N \left[-\gamma \left(\frac{\partial}{\partial \mathbf{p}_i} \mathbf{p}_i + m_i \beta^{-1} \frac{\partial^2}{\partial \mathbf{p}_i^2} \right) \right] = iL_{\mathbf{p}}^{(\text{cl})} + iL_{\mathbf{q}}^{(\text{cl})} + iL_{\gamma}^{(\text{cl-LE})}, \quad (1.14)$$

where $iL_{\gamma}^{(\text{cl-LE})}$ is the propagator of the momentum due to the modified Langevin dynamics. It is easy to verify that Langevin dynamics conserves the canonical ensemble as both $iL^{(\text{cl})}$ and $iL_{\gamma}^{(\text{cl-LE})}$ individually conserve $\exp(-\beta H(\mathbf{p}, \mathbf{q}))$. The time propagation over one simulation step is defined as

$$\begin{aligned} (\mathbf{p}(t' + \Delta t'), \mathbf{q}(t' + \Delta t')) &= \exp \left(iL_{\gamma}^{(\text{cl-LE})} \Delta t' + iL_{\mathbf{p}}^{(\text{cl})} \Delta t' + iL_{\mathbf{q}}^{(\text{cl})} \Delta t' \right) (\mathbf{p}(t'), \mathbf{q}(t')) \\ &\approx \exp \left(iL_{\gamma}^{(\text{cl-LE})} \Delta t' / 2 \right) \exp \left(iL_{\mathbf{p}}^{(\text{cl})} \Delta t' / 2 \right) \exp \left(iL_{\mathbf{q}}^{(\text{cl})} \Delta t' \right) \\ &\quad \exp \left(iL_{\mathbf{p}}^{(\text{cl})} \Delta t' / 2 \right) \exp \left(iL_{\gamma}^{(\text{cl-LE})} \Delta t' / 2 \right) (\mathbf{p}(t'), \mathbf{q}(t')). \end{aligned} \quad (1.15)$$

implying time evolution of the momenta using the thermostat for half the timestep, followed by the velocity-Verlet scheme, finally concluded by time evolution using the thermostat for another half a timestep. We have deliberately not mathematically expressed the thermostat step, as there exist numerous thermostats^[43,44,45,46,47] which could be used in place of the Langevin thermostat in an actual simulation.

Sampling the classical isothermal-isobaric distribution

In order to sample the isothermal isobaric ensemble, it is important to introduce the volume \mathcal{V} and – in the spirit of molecular dynamics – its conjugate momentum α as a dynamical variables, and to modify Eq. 1.13 with terms that couple the system's position and momentum with volume fluctuations. Several attempts were made to define EOMs^[48,49,43,50,51,52] that

rigorously conserve the isothermal-isobaric ensemble. Here, we list one such set of EOMs proposed by Bussi Zykova and Parinello^[52] (BZP)

$$\begin{aligned}\dot{\mathbf{p}}_i &= \mathbf{f}_i - \gamma \mathbf{p}_i + \sqrt{2\gamma m_i \beta} \mathbf{R}_i - \alpha \mu^{-1} \mathbf{p}_i, \quad \dot{\mathbf{q}}_i = \frac{\mathbf{p}_i}{m_i} + \alpha \mu^{-1} \mathbf{q}_i, \\ \dot{\mathcal{V}} &= 3\alpha \mu^{-1} \mathcal{V}, \quad \dot{\alpha} = [3\mathcal{V}(\mathcal{P}^{\text{int}} - \mathcal{P}^{\text{ext}}) + \beta^{-1}] - \gamma \alpha + \sqrt{2\gamma \alpha \mu \beta} \mathbf{R}'_i\end{aligned}\tag{1.16}$$

where,

$$\mathcal{P}^{\text{int}} = \frac{1}{3\mathcal{V}} \text{Tr} \left[\sum_{i=1}^N m_i^{-1} \mathbf{p}_i \otimes \mathbf{p}_i + \sum_{i=1}^N \mathbf{q}_i \otimes \mathbf{f}_i - \frac{\partial V(\mathbf{q}, \mathbf{h})}{\partial \mathbf{h}} \mathbf{h}^T \right] = \frac{1}{3\mathcal{V}} \text{Tr} \left[\sum_{i=1}^N m_i^{-1} \mathbf{p}_i \otimes \mathbf{p}_i - \Xi \right], \tag{1.17}$$

is the instantaneous internal pressure of the system, Ξ is the virial tensor, and \otimes defines an outer product. As in the case of the canonical ensemble, it is straightforward (but tedious) to show that the Liouvillian, $iL^{(\text{cl-BZP})} = \sum_{i=1}^N \left[\dot{\mathbf{p}}_i \frac{\partial}{\partial \mathbf{p}_i} + \dot{\mathbf{q}}_i \frac{\partial}{\partial \mathbf{q}_i} \right] + \dot{\mathcal{V}} \frac{\partial}{\partial \mathcal{V}} + \dot{\alpha} \frac{\partial}{\partial \alpha}$, conserves the isothermal-isobaric distribution^[52] $\exp \left(-\beta \left(H(\mathbf{p}, \mathbf{q}) + \frac{\alpha^2}{2\mu} + \mathcal{P}^{\text{ext}} \mathcal{V} \right) \right)$.

To determine the integration scheme, it is useful to split the Liouvillian of Eq. 1.16 into $iL^{(\text{cl-BZP})} = iL_{\gamma}^{(\text{cl-BZP})} + iL_{\mathbf{p}, \alpha}^{(\text{cl-BZP})} + iL_{\mathbf{q}, \mathbf{p}, \mathcal{V}}^{(\text{cl-BZP})}$, where $iL_{\gamma}^{(\text{cl-BZP})}$ is associated with the thermostat of the system and the volume, $iL_{\mathbf{p}, \alpha}^{(\text{cl-BZP})} = \sum_{i=1}^N \mathbf{f}_i \frac{\partial}{\partial \mathbf{p}_i} + [3\mathcal{V}(\mathcal{P}^{\text{int}} - \mathcal{P}^{\text{ext}}) + \beta^{-1}] \frac{\partial}{\partial \alpha}$ is the Liouvillian of the system's and volume's momentum associated with the pressure and the forces, and $iL_{\mathbf{q}, \mathbf{p}, \mathcal{V}}^{(\text{cl-BZP})} = \sum_{i=1}^N \left[m_i^{-1} \mathbf{p}_i \frac{\partial}{\partial \mathbf{q}_i} + \alpha \mu^{-1} \mathbf{q}_i \frac{\partial}{\partial \mathbf{q}_i} - \alpha \mu^{-1} \mathbf{p}_i \frac{\partial}{\partial \mathbf{p}_i} \right] + 3\alpha \mu^{-1} \mathcal{V} \frac{\partial}{\partial \mathcal{V}}$ is the position's and volume's Liouvillian. The time propagation over one simulation step is defined as^[52]

$$\begin{aligned}& (\mathbf{p}(t' + \Delta t'), \mathbf{q}(t' + \Delta t'), \alpha(t' + \Delta t'), \mathcal{V}(t' + \Delta t')) \\ &= \exp \left(iL_{\gamma}^{(\text{cl-BZP})} \Delta t' + iL_{\mathbf{p}, \alpha}^{(\text{cl-BZP})} \Delta t' + iL_{\mathbf{q}, \mathbf{p}, \mathcal{V}}^{(\text{cl-BZP})} \Delta t' \right) (\mathbf{p}(t'), \mathbf{q}(t')) \\ &\approx \exp \left(iL_{\gamma}^{(\text{cl-BZP})} \Delta t' / 2 \right) \exp \left(iL_{\mathbf{p}, \alpha}^{(\text{cl-BZP})} \Delta t' / 2 \right) \exp \left(iL_{\mathbf{q}, \mathbf{p}, \mathcal{V}}^{(\text{cl-BZP})} \Delta t' \right) \\ &\quad \exp \left(iL_{\mathbf{p}, \alpha}^{(\text{cl-BZP})} \Delta t' / 2 \right) \exp \left(iL_{\gamma}^{(\text{cl-BZP})} \Delta t' / 2 \right) (\mathbf{p}(t'), \mathbf{q}(t'), \alpha(t'), \mathcal{V}(t')), \end{aligned}\tag{1.18}$$

implying the time evolution of the system's and volume's momentum using the thermostat for half the timestep, followed by the update of the system's and volume's momenta using the

forces and the pressure, respectively, for half a timestep^[52]

$$\begin{aligned}\alpha &\rightarrow \alpha + (\delta t'/2) [3\mathcal{V}(\mathcal{P}^{\text{int}} - \mathcal{P}^{\text{ext}}) + \beta^{-1}] \\ &+ (\delta t'/2)^2 \sum_{i=1}^N m_i^{-1} \mathbf{p}_i \cdot \mathbf{f}_i + \frac{1}{3} (\delta t'/2)^3 \sum_{i=1}^N m_i^{-1} \mathbf{f}_i \cdot \mathbf{f}_i, \\ \mathbf{p}_i(t' + \Delta t'/2) &\rightarrow \mathbf{p}_i(t') + \mathbf{f}_i \Delta t'/2,\end{aligned}\tag{1.19}$$

followed by the update of the volume, and the system's position and momentum for a full timestep^[52]

$$\begin{aligned}\mathbf{q}_i &\rightarrow \exp(\mu^{-1} \alpha \Delta t') \mathbf{q}_i + \frac{\sinh(\mu^{-1} \alpha \Delta t')}{\mu^{-1} \alpha} m_i^{-1} \mathbf{p}_i(t') \\ \mathbf{p}_i &\rightarrow \exp(-\mu^{-1} \alpha \Delta t') \mathbf{p}_i, \\ \mathcal{V} &\rightarrow \exp(3\mu^{-1} \alpha \Delta t') \mathcal{V}\end{aligned}\tag{1.20}$$

continuing with the second part of the integrator applying Eq. 1.19 and the thermostat steps for half timesteps.

1.2.2 Molecular dynamics for calculating classical dynamical properties

So far we have tried to estimate thermodynamic properties of systems, by modelling their equilibrium behaviour in the absence of external perturbations. These properties are calculated by sampling from relevant thermodynamic ensembles by simulating Hamiltonian dynamics coupled to an external bath. While in equilibrium and within the ergodic hypothesis, it is assumed that a system loses memory of its initial microstate, making it possible to estimate its thermodynamic properties as averages over phase space, without necessarily reproducing the real time dynamics of the modelled system. Consequently, thermodynamic properties are also known as static properties.

However, there exist many "dynamical" properties which are probed by perturbing the system from its equilibrium state^[22]. For instance, the diffusion coefficient can be estimated by applying a concentration gradient and tracking the motion of atoms. Similarly, vibrational spectra can be estimated by subjecting the system to external electromagnetic fields and studying the response of the system. Since some of these perturbations can be time dependent,

Chapter 1. Introduction

the resulting response can depend on the frequency of the perturbations and therefore be time dependent as well. One would think that it would be essential to simulate the steady state behaviour of the system within non-equilibrium ensembles in the presence of external fields. However, under the assumption of weak external fields, linear response theory^[53] allows one to estimate the expectation values over these non-equilibrium ensembles in terms of time correlation functions (TCF) estimated from equilibrium simulations. TCFs therefore provide the link between the dynamics of the system in equilibrium and experimental “dynamical” observables.

Within the classical limit, the equilibrium canonical TCF between the observables $A(\mathbf{p}, \mathbf{q})$ and $B(\mathbf{p}, \mathbf{q})$ for the system described by Eq. 1.1 is

$$\begin{aligned} C_{AB}^{\text{cl}}(t) &= Z^{\text{cl}}(N, \mathcal{V}, \beta)^{-1} \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\mathbf{p} d\mathbf{q} \exp(-\beta H(\mathbf{p}, \mathbf{q})) A(\mathbf{p}(0), \mathbf{q}(0)) \exp\left(iL^{(\text{cl})}t\right) B(\mathbf{p}(0), \mathbf{q}(0)) \\ &= Z^{\text{cl}}(N, \mathcal{V}, \beta)^{-1} \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\mathbf{p} d\mathbf{q} \exp(-\beta H(\mathbf{p}, \mathbf{q})) A(\mathbf{p}(0), \mathbf{q}(0)) B(\mathbf{p}(t), \mathbf{q}(t)) \\ &= \langle A(\mathbf{p}(0), \mathbf{q}(0)) B(\mathbf{p}(t), \mathbf{q}(t)) \rangle^{(\text{cl})} \end{aligned} \quad (1.21)$$

By expanding the classical propagator $\exp(iL^{(\text{cl})}t)$ as a Taylor series, applying integration by parts repeatedly in position and momentum coordinates, and exploiting the fact that the classical Liouvillian conserves the canonical Boltzmann distribution, one can prove that the detailed balance condition

$$C_{AB}^{\text{cl}}(t) = \langle A(\mathbf{q}(0)) B(\mathbf{q}(t)) \rangle^{(\text{cl})} = \langle B(\mathbf{q}(0)) A(\mathbf{q}(-t)) \rangle^{(\text{cl})} = C_{BA}^{\text{cl}}(-t). \quad (1.22)$$

if fulfilled. Setting $A = 1$ yields another detailed balance condition

$$C_B^{\text{cl}}(t) = C_B^{\text{cl}}(-t) = \langle B(\mathbf{q}) \rangle^{(\text{cl})}, \quad (1.23)$$

that links thermodynamic expectation values and TCFs. Eq. 1.23 is a direct consequence of the fact that classical Liouvillian of the Hamiltonian conserves the canonical Boltzmann distribution.

In principle, for the computation of TCFs given by Eq. 1.21, a single Hamiltonian dynamics trajectory should suffice, as it simultaneously conserves the canonical Boltzmann distribution, and also classical evolves the system in real time. However, for many systems (such as harmonic oscillators) Hamiltonian dynamics is not ergodic and can lead to insufficient sampling of the configuration space. Therefore, in practice, TCFs are estimated by sampling initial values ($\mathbf{p}(0), \mathbf{q}(0)$) from the canonical Boltzmann distribution using molecular dynamics in combination with thermostats, performing Hamiltonian dynamics using Eq. 1.11 from each of the starting points, computing TCFs for all trajectories, and subsequently averaging them^[54]. Similarly, time correlation functions can be estimated in the isothermal-isobaric ensemble by evaluating the thermal average in Eq. 1.21 in the isothermal-isobaric ensemble, which in practice leads to sampling of initial values of ($\mathbf{p}(0), \mathbf{q}(0)$) using constant pressure molecular dynamics simulations. This approach, however, can be computationally expensive due to the need of performing multiple simulations, and tedious as one must check for errors arising from lack of ergodicity.

A more convenient but less rigorous way of estimating time correlation functions is by using a single thermostatted trajectory to compute Eq. 1.21. However, care must be taken while selecting the thermostat so that the perturbation caused upon the real time dynamics of the system is minimized. To this extent, global thermostats such as “stochastic velocity rescaling”^[46] and “Nose-Hoover” thermostat^[48], which do not act on individual atoms but only rescale the total momentum of the system, do not lead to a significant disruption of the real time dynamical properties of the system, and therefore can usually be readily applied to the computation of TCFs.

1.2.3 Limitations of classical thermodynamics

Thermodynamic properties computed with classical nuclei simulations are valid in the high temperature limit or for system containing heavy elements. However, there exist many systems containing light nuclei such as hydrogen for which the classical approximation is no longer valid even at ambient temperature^[55,56,57]. For instance, the pH of water is 7 because of the quantum nature of protons. The protons experience large fluctuations along the OH bond

and trigger auto-protolysis events^[58], due to the high zero point energy of the O–H bond. If the protons are simulated as classical particles the fluctuations along OH are relatively small, the probability of auto-protolysis is reduced by many orders of magnitude^[58], and therefore the pH increases to about 9. Similarly, the acidity of active sites of enzymes in our body can be under-estimated by many orders of magnitude if the nuclei are assumed to be classical particles^[59]. These examples underscore the importance of the quantum nature of nuclei in the maintenance of aqueous biological reactions within our bodies that sustain our lives. Other systems that exhibit NQEs are helium^[60], the record high- T_c conventional superconductor SH_3 ^[61], metal-organic frameworks^[62,63], as well as most molecular materials^[64,65].

There are also many properties which can not be modelled using classical statistical mechanics. Recall that position dependent properties observables calculated using Eq. 1.8 do not depend on the mass of the particles, but only on the BO potential energy landscape. Classical nuclei simulations, therefore, will give the same results when applied to study systems containing different isotopes of an element. We shall later see that these simulations actually correspond to quantum nuclei simulations with hypothetical isotopes of elements that have infinite masses. We can therefore say that difference in thermodynamical properties of isotopomers is a "pure" quantum effect that can not be described classically. Similarly, since the momenta used in classical molecular dynamics are just auxiliary variables used to sample the Boltzmann distribution of the positions, momentum dependent properties such as particle momentum distributions, kinetic energies, heat capacities etc. are not targetted by these simulations.

Therefore, it is important to turn to quantum statistical mechanics which allows one to study all systems at low temperature, and systems containing light, as well as calculate isotope dependent observables, and those that depend explicitly on momenta.

1.3 Quantum thermodynamics

Fortunately, it is possible to treat the quantum statistical mechanics of distinguishable nuclei in a numerically exact manner, without painstakingly computing all the vibrational energy

levels. The imaginary time path integral method allows one to sample configurations from the quantum (Boltzmann) distributions of thermodynamics ensembles, just as in the case of classical statistical mechanics, although at a much higher computational cost. Nevertheless, it is the most effective framework for treating the quantum mechanical behavior of distinguishable particles. Within the path integral formalism, the quantum partition function is shown to be equivalent to the classical partition function of an extended classical system made of several replicas of the physical systems at an elevated temperature^[66]. This is shown by first writing the canonical Boltzmann distribution at β as the product of P high temperature distributions at $\beta_P = \frac{\beta}{P}$

$$Z(N, \mathcal{V}, \beta) = \text{Tr} [\exp(-\beta \hat{H})] = \text{Tr} [(\exp(-\beta_P \hat{H}))^P]$$

and subsequently performing a Trotter splitting of the high temperature Boltzmann distributions

$$\begin{aligned} Z(N, \mathcal{V}, \beta) &= \text{Tr} [(\exp(-\beta_P \hat{H}))^P] \\ &= \text{Tr} [(\exp(-\beta_P \hat{V}/2) \exp(-\beta_P \hat{T}) \exp(-\beta_P \hat{V}/2))^P] + \mathcal{O}(P^{-2}). \end{aligned}$$

As in the case of classical statistical mechanics, the trace is estimated in the position basis $|\mathbf{q}^{(1)}\rangle$ and $P-1$ additional completeness relations $\mathbf{1} = \int d\mathbf{q}^{(j)} |\mathbf{q}^{(j)}\rangle \langle \mathbf{q}^{(j)}|$ for $j \in \{2, \dots, P\}$ are inserted between the high temperature Boltzmann operators to estimate the second-order canonical partition function

$$Z_P^{(2)}(N, \mathcal{V}, \beta) = \int d\{\mathbf{q}^{(j)}\} \prod_{j=1}^P \langle \mathbf{q}^{(j)} | \exp(-\beta_P \hat{V}/2) \exp(-\beta_P \hat{T}) \exp(-\beta_P \hat{V}/2) | \mathbf{q}^{(j)} \rangle$$

where, $\{\mathbf{q}^{(j)}\}$ is short hand for extended vector $\{\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(P)}\}$, and $\mathbf{q}^{(j+P)} \equiv \mathbf{q}^{(j)}$ is implied. It is easy to show that the second-order canonical partition function is

$$\begin{aligned} Z_P^{(2)}(N, \mathcal{V}, \beta) &= \left(\frac{1}{2\pi\hbar}\right)^{3NP} \left(\frac{1}{2\pi\hbar}\right)^{3NP} \prod_{i=1}^N (2\pi m_i \beta_P^{-1})^{\frac{3P}{2}} \\ &\int d\{\mathbf{q}^{(j)}\} \exp\left(-\beta_P \sum_{j=1}^P \left[\sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 (\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)})^2 + V(\mathbf{q}^{(j)}) \right]\right). \end{aligned} \quad (1.24)$$

under the assumption that the dimensions of the cell are much larger than the delocalization of the particles. ^[67] As in the case of classical statistical mechanics, it is possible to convert the prefactor into Gaussian integrals of fictitious momenta of the atoms in each replica of the system, leading to

$$Z_P^{(2)}(N, \mathcal{V}, \beta) = \left(\frac{1}{2\pi\hbar} \right)^{3NP} \int d\{\mathbf{p}^{(j)}\} d\{\mathbf{q}^{(j)}\} \exp\left(-\beta_P H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\})\right) \quad (1.25)$$

which equals the classical canonical partition function of an extended classical system – referred to as the ring polymer – at inverse temperature β_P . The ring polymer Hamiltonian is made of P replicas of the system such that the atoms in consecutive replicas are connected with springs of frequency $\omega_P = \hbar^{-1} \beta_P^{-1}$

$$H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\}) = \sum_{j=1}^P \left[\sum_{i=1}^N \frac{1}{2} m_i^{-1} |\mathbf{p}_i^{(j)}|^2 + \sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 (\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)})^2 + V(\mathbf{q}^{(j)}) \right]. \quad (1.26)$$

Note that the dynamical masses in Eq. 1.26 could in principle be a generalized mass matrix ^[68], however, for simplicity we choose a diagonal matrix of physical masses. This Hamiltonian is also known as the imaginary time path integral Hamiltonian as it can also be derived by writing the canonical partition function as an imaginary time propagator in the path integral representation, and discretising the action integral ^[24,22]. The spring term allows the particles to be delocalized in space, introducing a net uncertainty in their positions. Since the spring forces felt by the atoms depends explicitly on their masses, the heavy isotope of an atom would experience stiffer springs and therefore smaller quantum delocalization. Note that in the limits of infinite masses and temperature, the stiffness of the spring shoots up to infinity and all the replicas collapse to the centroid leading to a fully classical description.

The thermodynamic average of any position dependent observable \hat{A} can either be estimated as ^[66] $\langle A(\mathbf{q}) \rangle_P^{(2)} = Z(N, \mathcal{V}, \beta)^{-1} \text{Tr}[\exp(-\beta \hat{H}) \hat{A}]$, or by deriving the partition function as in Eq.

1.7. Both methods lead to the estimator

$$\begin{aligned} \langle A(\mathbf{q}) \rangle_P^{(2)} &= Z_P^{(2)}(N, \mathcal{V}, \beta)^{-1} \int d\{\mathbf{p}^{(j)}\} d\{\mathbf{q}^{(j)}\} \left[\frac{1}{P} \sum_{j=1}^P A(\mathbf{q}^{(j)}) \right] \exp\left(-\beta_P H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\})\right) \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt' \frac{1}{P} \sum_{j=1}^P A(\mathbf{q}^{(j)}(t')), \end{aligned}$$

(1.27)

a time average of the average over beads of the observable, sampled using the classical Boltzmann distribution of the second-order ring polymer Hamiltonian. Note that in the limit of $P \rightarrow \infty$, the estimator converges to the exact quantum expectation value with an error that scales as $\mathcal{O}(P^{-2})$. Momentum dependent observables can not be calculated by merely averaging over the fictitious momenta of the ring polymer, but require sampling of open chain polymers^[69] instead of closed ring polymers. As we shall see later there are several challenges associated with calculating momentum dependent observables that need to be addressed.

1.3.1 Path integral molecular dynamics for sampling quantum ensembles

Expectation values of observables in Eq. 1.27 can be estimated, as in the case of classical statistical mechanics, using the molecular dynamics technique by simulating the $3N \times P$ dimensional ring polymer Hamiltonian^[70], and by coupling its equations of motion with those of thermostats^[67,71] and barostats^[72,73]. The generic term used to describe molecular dynamics for sampling thermodynamic ensembles, obtained within the imaginary time path integral method, is known as path integral molecular dynamics (PIMD). Let us briefly go through the algorithms that enable PIMD for sampling the canonical, and the isothermal-isobaric ensembles.

Sampling the quantum canonical distribution

The ring polymer Hamiltonian has a wide range of normal mode frequencies. For instance the characteristic frequencies of a free ring polymer i.e. one without any potential, are $\omega_P^{(j)} = 2\omega_P \sin \frac{(j-1)\pi}{P}$ for $j \in \{1, \dots, P\}$, which means the friction parameter of a single Langevin thermostat applied to the full ring polymer, may be too small or too large for some of the normal modes^[74]. Both the regimes lead to non-ergodic sampling as a small friction parameter implies less chaotic dynamics, while a large parameter implies overdamped motion which prevents sampling new regions of phase space^[74]. One of the ways to ensure ergodic sampling, as proposed by Ceriotti and co-workers, is to apply P thermostats to the different normal

modes of the ring polymer Hamiltonian leading to the EOMs^[71]

$$\begin{aligned}\dot{\tilde{\mathbf{p}}}_i^{(j)} &= \tilde{\mathbf{f}}_i^{(j)} - \gamma^{(j)} \tilde{\mathbf{p}}_i^{(j)} - m_i \omega_P^2 \tilde{\mathbf{q}}_i^{(j)} + \sqrt{2m_i \gamma^{(j)}} \beta_P^{-1} \mathbf{R}_i^{(j)}, \\ \dot{\tilde{\mathbf{q}}}_i^{(j)} &= m_i^{-1} \tilde{\mathbf{p}}_i^{(j)},\end{aligned}\tag{1.28}$$

where $\{\tilde{\mathbf{p}}^{(j)}, \tilde{\mathbf{q}}^{(j)}\}$ are the momentum and position written in the normal mode representation of the free ring polymer. The choice $\gamma^{(j)}|_{j \neq 0} = 2\tilde{\omega}^{(j)}$ leads to optimal sampling^[74] i.e. the middle ground between near-Hamiltonian dynamics and over damped motion, for the Hamiltonian of free particles. In addition a ‘‘centroid’’ thermostat is attached to the first normal mode with zero frequency also known as the centroid mode. The thermostating scheme is known as the path integral Langevin equation (PILE- \square)^[71], where \square is G or L depending on whether the centroid thermostat is global (G) or local (L). Note that the use of multiple thermostats within PILE- \square conserves the canonical distribution as the kinetic energy is trivially decoupled into independent normal mode components. Although derived in the free particle limit, the PILE thermostat remains useful for most realistic systems.

A straightforward integration of Eq. 1.28 using the velocity Verlet scheme, would require a very small timestep as the springs frequencies are usually higher than the physical characteristic frequencies of most systems. One way to avoid this problem is by using a mass matrix that reduces the frequencies of the free ring polymer. The other way, which allows physical masses to be used for the ring polymer, requires the total Liouvillian of Eq. 1.28 to be written as:

$$iL^{(\text{PI-LE})} = \sum_{j=1}^P iL_{\mathbf{pq}^{(j)}}^{(\text{PI})} + iL_{\mathbf{p}}^{(\text{PI})} + iL_{\gamma}^{(\text{PI-LE})},\tag{1.29}$$

where $iL_{\gamma}^{(\text{PI-LE})}$ is associated with the thermostat, $iL_{\mathbf{p}}^{(\text{PI})} = \sum_{j=1}^P \sum_{i=1}^N \tilde{\mathbf{f}}_i^{(j)} \frac{\partial}{\partial \tilde{\mathbf{p}}_i^{(j)}}$ is the Liouvillian of the momenta due to the physical forces, and $\sum_{j=1}^P iL_{\mathbf{pq}^{(j)}}^{(\text{PI})}$ is the Liouvillian of the free ring polymer that only experiences spring forces. The time propagation for a timestep $\Delta t'$ can then

be performed as^[71]

$$\begin{aligned} \exp(iL^{(\text{PI-LE})}\Delta t') \left(\{\mathbf{p}^{(j)}\}(t' + \Delta t'), \{\mathbf{q}^{(j)}\}(t' + \Delta t') \right) &= \exp(iL_{\gamma}^{(\text{PI-LE})}\Delta t'/2) \exp(iL_{\mathbf{p}}^{(\text{PI})}\Delta t'/2) \\ &\exp\left(\sum_{j=1}^P iL_{\mathbf{p}\mathbf{q}^{(j)}}^{(\text{PI})}\Delta t'\right) \exp(iL_{\mathbf{p}}^{(\text{PI})}\Delta t'/2) \exp(iL_{\gamma}^{(\text{PI-LE})}\Delta t'/2) \left(\{\mathbf{p}^{(j)}\}(t'), \{\mathbf{q}^{(j)}\}(t') \right). \end{aligned} \quad (1.30)$$

which implies update of momenta due to the thermostat for half a timestep, followed by the update of momenta due to the physical forces for another half a timestep, followed by the exact propagation (that is analytically known) using the free particle Liouvillian for a full timestep, followed by the other symmetric part of the integration scheme. The exact propagation typically enables the use of a time step that is same as or half of the one used in classical nuclei simulations. Recent developments, such as the *BAOAB* integration scheme^[75] which basically applies the thermostat step in the middle, instead of two thermostat steps at the ends, and numerically integrates the spring forces allows one to use – in most circumstances – the same timesteps that one would in classical nuclei simulations^[76]. And finally, if one runs into problems such as numerical instability or poor ergodicity due to the numerical integration scheme, especially in the limit of large number of replicas, the recently developed Cayley modified integration schemes^[77,78] can be used.

Sampling the quantum isothermal-isobaric distribution

The isothermal-isobaric ensemble can be sampled by simulating classical constant pressure dynamics for the ring polymer Hamiltonian at inverse temperature β_P and P times the external pressure $\mathcal{P}_P^{\text{ext}} = \mathcal{P}^{\text{ext}} P$ as

$$\begin{aligned} \Delta_P^{(2)}(N, \mathcal{P}^{\text{ext}}, \beta) &= \mathcal{V}_0^{-1} \int d\mathcal{V} \exp(-\beta \mathcal{P}^{\text{ext}} \mathcal{V}) Z_P^{(2)}(N, \mathcal{V}, \beta) \\ &= \mathcal{V}_0^{-1} \int d\mathcal{V} \exp(-\beta_P \mathcal{P}_P^{\text{ext}} \mathcal{V}) Z_P^{(2)}(N, \mathcal{V}, \beta). \end{aligned} \quad (1.31)$$

To sample the isothermal-isobaric ensemble in an efficient manner, it is useful to assume that the fluctuations of the centroid mode are much larger than those of the normal modes, and only the former are scaled during volume fluctuations^[72]. This results in the following

reduced ensemble^[72]

$$\Delta_P^{(2)}(N, \mathcal{P}^{\text{ext}}, \beta) \propto \int d\mathcal{V} \int d\mathbf{p}^{(1)} d\mathbf{q}^{(1)} \int d\{\mathbf{p}^{(j)}\}_{j \neq 1} d\{\mathbf{q}^{(j)}\}_{j \neq 1} \exp\left(-\beta_P \left(H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\}) + \mathcal{P}_P^{\text{ext}} \mathcal{V}\right)\right) \quad (1.32)$$

where only the integral over the centroid position is performed over the domain of \mathcal{V} . We present the EOMs proposed by Ceriotti and co-workers^[73], which are modifications of the classical isothermal-isobaric EOMs described by Eq. 1.16 that sample the reduced ensemble using the PILE thermostat

$$\begin{aligned} \dot{\tilde{\mathbf{p}}}_i^{(j)} &= \tilde{\mathbf{f}}_i^{(j)} - \gamma \tilde{\mathbf{p}}_i^{(j)} - m_i \omega_P^2 \tilde{\mathbf{q}}_i^{(j)} + \sqrt{2\gamma m_i \beta_P} \mathbf{R}_i^{(j)}, \quad \text{for } j = \{2, \dots, P\} \\ \tilde{\mathbf{q}}_i^{(j)} &= m_i^{-1} \tilde{\mathbf{p}}_i^{(j)}, \quad \text{for } j = \{2, \dots, P\} \\ \dot{\tilde{\mathbf{p}}}_i^{(1)} &= \tilde{\mathbf{f}}_i^{(1)} - \gamma \tilde{\mathbf{p}}_i^{(1)} + \sqrt{2\gamma m_i \beta_P} \mathbf{R}_i^{(1)} - \alpha \mu^{-1} \tilde{\mathbf{p}}_i^{(1)}, \\ \dot{\tilde{\mathbf{q}}}_i^{(1)} &= m_i^{-1} \tilde{\mathbf{p}}_i^{(1)} + \alpha \mu^{-1} \tilde{\mathbf{q}}_i^{(1)}, \\ \dot{\mathcal{V}} &= 3\alpha \mu^{-1} \mathcal{V}, \\ \dot{\alpha} &= [3\mathcal{V}(\mathcal{P}_P^{\text{int}} - \mathcal{P}_P^{\text{ext}}) + \beta_P^{-1}] - \gamma_\alpha \alpha + \sqrt{2\gamma_\alpha \mu \beta_P} \mathbf{R}'_i \end{aligned} \quad (1.33)$$

where,

$$\mathcal{P}_P^{\text{int}} = \frac{1}{3\mathcal{V}} \text{Tr} \left[\sum_{i=1}^N m_i^{-1} \tilde{\mathbf{p}}_i^{(1)} \otimes \tilde{\mathbf{p}}_i^{(1)} - \frac{1}{P} \sum_{j=1}^P \sum_{i=1}^N (\tilde{\mathbf{q}}_i^{(j)} - \bar{\mathbf{q}}_i) \otimes \tilde{\mathbf{f}}_i^{(j)} - \frac{1}{P} \sum_{j=1}^P \Xi^{(j)} \right], \quad (1.34)$$

with $\mathcal{P}_P^{\text{int}}$ the total internal pressure of the Hamiltonian as estimated by the centroid-virial estimator^[72], and $\bar{\mathbf{q}}_i$ the centroid of the i^{th} atom. The total Liouvillian of Eq. 1.33

$$iL^{(\text{PI-BZP})} = \sum_{j=2}^P iL_{\mathbf{p}\mathbf{q}^{(j)}}^{(\text{PI})} + iL_{\mathbf{p},\alpha}^{(\text{PI-BZP})} + iL_{\mathbf{p},\mathbf{q},\mathcal{V}}^{(\text{PI-BZP})} + iL_{\gamma}^{(\text{PI-BZP})}, \quad (1.35)$$

with $iL_{\gamma}^{(\text{PI-BZP})}$ the Liouvillian associated with the thermostat of the system and the volume, $\sum_{j=2}^P iL_{\mathbf{p}\mathbf{q}^{(j)}}^{(\text{PI})}$ the Liouvillian of the free ring polymer Hamiltonian with all modes but the centroid, $iL_{\mathbf{p},\alpha}^{(\text{PI-BZP})} = \sum_{j=1}^P \sum_{i=1}^N \tilde{\mathbf{f}}_i^{(j)} \frac{\partial}{\partial \tilde{\mathbf{p}}_i^{(j)}} + \dot{\alpha} \frac{\partial}{\partial \alpha}$, the Liouvillian associated with the system's and

volume's momenta due to the force and the internal pressure, respectively, and

$$iL_{\mathbf{p},\mathbf{q},\mathcal{V}}^{(\text{PI-BZP})} = \sum_{i=1}^N \left[m_i^{-1} \mathbf{p}_{\tilde{i}}^{(1)} \frac{\partial}{\partial \mathbf{q}_{\tilde{i}}^{(1)}} + \alpha \mu^{-1} \mathbf{q}_{\tilde{i}}^{(1)} \frac{\partial}{\partial \mathbf{q}_{\tilde{i}}^{(1)}} - \alpha \mu^{-1} \mathbf{p}_{\tilde{i}}^{(1)} \frac{\partial}{\partial \mathbf{p}_{\tilde{i}}^{(1)}} \right] + 3\alpha \mu^{-1} \mathcal{V} \frac{\partial}{\partial \mathcal{V}}$$

the Liouvillian associated with the centroid mode of the ring polymer, the volume, and the rest of the terms associated with the centroid's momentum and position. It is tedious but straightforward to verify that Eq. 1.33 conserve the reduced isothermal-isobaric distribution i.e. $iL^{(\text{PI-BZP})} \exp\left(-\beta_P \left(H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\}) + \frac{\alpha^2}{2\mu} + \mathcal{P}_P^{\text{ext}} \mathcal{V}\right)\right) = 0$. The time propagation for the EOMs is performed as

$$\begin{aligned} & \exp\left(iL^{(\text{PI-BZP})} \Delta t'\right) \left(\{\mathbf{p}^{(j)}\}(t' + \Delta t'), \{\mathbf{q}^{(j)}\}(t' + \Delta t'), \alpha(t' + \Delta t'), \mathcal{V}(t' + \Delta t')\right) \\ &= \exp\left(iL_{\gamma}^{(\text{PI-BZP})} \Delta t'/2\right) \exp\left(iL_{\mathbf{p},\alpha}^{(\text{PI-BZP})} \Delta t'/2\right) \exp\left(\left(\sum_{j=2}^P iL_{\mathbf{p}\mathbf{q}^{(j)}}^{(\text{PI})} + iL_{\mathbf{p},\mathbf{q},\mathcal{V}}^{(\text{PI-BZP})}\right) \Delta t'\right) \\ & \exp\left(iL_{\mathbf{p},\alpha}^{(\text{PI-BZP})} \Delta t'/2\right) \exp\left(iL_{\gamma}^{(\text{PI-BZP})} \Delta t'/2\right) \left(\{\mathbf{p}^{(j)}\}(t'), \{\mathbf{q}^{(j)}\}(t'), \alpha(t'), \mathcal{V}(t')\right). \end{aligned} \quad (1.36)$$

which implies, the thermostat step for both the system and the volume, the update of the momenta of the system and the volume due to forces and internal pressure, respectively, for half a timestep^[73]

$$\begin{aligned} & \alpha \rightarrow \alpha + (\delta t'/2) [3\mathcal{V} (\mathcal{P}_P^{\text{int}} - \mathcal{P}_P^{\text{ext}}) + \beta_P^{-1}] \\ & + (\delta t'/2)^2 \sum_{i=1}^N m_i^{-1} \mathbf{p}_{\tilde{i}}^{(1)} \cdot \mathbf{f}_{\tilde{i}}^{(1)} + \frac{1}{3} (\delta t'/2)^3 \sum_{i=1}^N m_i^{-1} \mathbf{f}_{\tilde{i}}^{(1)} \cdot \mathbf{f}_{\tilde{i}}^{(1)}, \\ & \mathbf{p}_{\tilde{i}}^{(j)} \rightarrow \mathbf{p}_{\tilde{i}}^{(j)} + \mathbf{f}_{\tilde{i}}^{(j)} \Delta t'/2, \end{aligned} \quad (1.37)$$

followed by the update of the volume, and the position and momentum of the system for a full timestep^[73]

$$\begin{aligned} & \mathbf{q}_{\tilde{i}}^{(1)} \rightarrow \exp(\mu^{-1} \alpha \Delta t') \mathbf{q}_{\tilde{i}}^{(1)} + \frac{\sinh(\mu^{-1} \alpha \Delta t')}{\mu^{-1} \alpha} m_i^{-1} \mathbf{p}_{\tilde{i}}^{(1)} \\ & \mathbf{p}_{\tilde{i}}^{(1)} \rightarrow \exp(-\mu^{-1} \alpha \Delta t') \mathbf{p}_{\tilde{i}}^{(1)}, \\ & \mathcal{V} \rightarrow \exp(3\mu^{-1} \alpha \Delta t') \mathcal{V} \end{aligned} \quad (1.38)$$

and the exact evolution of all the modes but the centroid of the free ring polymer for a full timestep, continuing with the second part of the integrator applying Eq. 1.37 and the thermostat steps for half timesteps.

1.3.2 Path integral methods for approximate quantum dynamics

As in the classical case, quantum dynamical properties can be estimated in terms of quantum TCFs, by invoking linear response theory^[22]. The standard TCF in the canonical ensemble between position dependent operators \hat{A} and \hat{B} for the system described by Eq. 1.1 is

$$\begin{aligned} C_{AB}(t) &= Z(N, \mathcal{V}, \beta)^{-1} \text{Tr} [\exp(-\beta \hat{H}) \hat{A} \exp(i\hbar^{-1} \hat{H} t) \hat{B} \exp(-i\hbar^{-1} \hat{H} t)] \\ &= Z(N, \mathcal{V}, \beta)^{-1} \text{Tr} [\exp(-\beta \hat{H}) \hat{A}(0) \hat{B}(t)], \end{aligned} \quad (1.39)$$

which can be estimated if the eigenstates of the Hamiltonian are known. Let us try to apply the path integral trick to avoid the eigenvalue problem. This is done by estimating the trace in Eq. 1.39, in the position basis, leading to^[22]

$$\begin{aligned} C_{AB}(t) &= Z(N, \mathcal{V}, \beta)^{-1} \int d\mathbf{q}^{(j')} \langle \mathbf{q}^{(j')} | \exp(-\beta \hat{H}) \hat{A} \exp(i\hbar^{-1} \hat{H} t) \hat{B} \exp(-i\hbar^{-1} \hat{H} t) | \mathbf{q}^{(j')} \rangle \\ &= Z(N, \mathcal{V}, \beta)^{-1} \int d\mathbf{q}^{(j')} d\mathbf{q}^{(j'')} d\mathbf{q}^{(j''')} \langle \mathbf{q}^{(j')} | \exp(-\beta \hat{H}) | \mathbf{q}^{(j'')} \rangle A(\mathbf{q}^{(j'')}) \\ &\quad \langle \mathbf{q}^{(j'')} | \exp(i\hbar^{-1} \hat{H} t) | \mathbf{q}^{(j''')} \rangle B(\mathbf{q}^{(j''')}) \langle \mathbf{q}^{(j''')} | \exp(-i\hbar^{-1} \hat{H} t) | \mathbf{q}^{(j')} \rangle. \end{aligned} \quad (1.40)$$

Performing a Trotter discretization of the matrix elements and inserting completeness identities in $\langle \mathbf{q}^{(j')} | \exp(-\beta \hat{H}) | \mathbf{q}^{(j'')} \rangle$, $\langle \mathbf{q}^{(j'')} | \exp(i\hbar^{-1} \hat{H} t) | \mathbf{q}^{(j''')} \rangle$ and $\langle \mathbf{q}^{(j''')} | \exp(-i\hbar^{-1} \hat{H} t) | \mathbf{q}^{(j')} \rangle$ leads to the path integral expression of the standard time correlation function. However, since the matrix elements of the forward and backward time propagators $\exp(\pm i\hbar^{-1} \hat{H} t)$ are not positive definite^[22], they can not be sampled using path integral molecular dynamics methods.

An alternative to the standard TCF is the Kubo transformed TCF, is the Kubo transformed

correlation function^[53]

$$\begin{aligned} K_{AB}(t) &= Z(N, \mathcal{V}, \beta)^{-1} \int d\lambda \operatorname{Tr} [\exp(-(\beta - \lambda)\hat{H}) \hat{A} \exp(-\lambda\hat{H}) \exp(i\hbar^{-1}\hat{H}t) \hat{B} \exp(-i\hbar^{-1}\hat{H}t)] \\ &= Z(N, \mathcal{V}, \beta)^{-1} \int d\lambda \operatorname{Tr} [\exp(-(\beta - \lambda)\hat{H}) \hat{A}(0) \exp(-\lambda\hat{H}) \hat{B}(t)], \end{aligned} \quad (1.41)$$

which has the advantage that it is real, and that it reduces to the classical time correlation function in the classical (high temperature) and in harmonic limits, and that it satisfies the detailed balance condition $K_{AB}(t) = K_{BA}(-t)$ ^[53], unlike the standard quantum TCF. These properties open the possibility of approximating the Kubo transformed TCF using the path integral method, which relies on classical dynamics. Moreover, the Fourier transform (FT) of the Kubo transformed TCF is related to that of the standard time correlation function,

$$\operatorname{FT}[C_{AB}(t)](\omega) = \left[\frac{\beta\hbar\omega}{1 - \exp(-\beta\hbar\omega)} \right] \operatorname{FT}[K_{AB}(t)](\omega) \quad (1.42)$$

meaning that the Kubo transformed TCF can be just as easily related to experimental dynamical observables, as the standard TCF.

Formal derivation of path integral methods that deliver approximate quantum dynamics, requires the theory of Matsubara dynamics^[79] which approximates the Kubo transformed TCF by a classical TCF of an extended system (the Matsubara Hamiltonian), whose dynamics conserves the quantum Boltzmann distribution. Unfortunately, the Matsubara TCF contains a rapidly oscillating phase term, which makes its calculation very hard even for small systems. However, Matsubara dynamics provides the theoretical framework for the derivation of approximate methods, such as centroid molecular dynamics (CMD)^[80] and ring polymer molecular dynamics (RPMD)^[81] that do not suffer from the aforementioned phase problem, and can be used to compute TCFs of systems of practical interest^[82]. While the full derivation of Matsubara dynamics and its simplification to other approximate methods, is beyond the scope of this thesis, we refer the interested reader to Ref.^[79,83,84], for derivations and proceed to the description of CMD and RPMD methods.

Centroid molecular dynamics

Centroid molecular dynamics^[80] is an approximate method for computing quantum TCFs in terms of the dynamics of the centroid of the imaginary time path integral Hamiltonian, generated classically by the mean thermodynamic force acting on the centroid. Within CMD, the Kubo transformed TCF of linear position dependent operators \hat{A} and \hat{B} is approximated as

$$C_{AB}^{\text{CMD}}(t) = Z_P^2(N, \mathcal{V}, \beta)^{-1} \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\tilde{\mathbf{p}}_1' d\tilde{\mathbf{q}}_1' \exp \left(-\beta H^{\text{CMD}}(\tilde{\mathbf{p}}_1', \tilde{\mathbf{q}}_1') \right) A(\tilde{\mathbf{q}}_1'(0)) \exp(iL^{(\text{CMD})}t) B(\tilde{\mathbf{q}}_1'(0)), \quad (1.43)$$

where the CMD Hamiltonian is defined as $H^{\text{CMD}}(\tilde{\mathbf{p}}_1', \tilde{\mathbf{q}}_1') = \sum_{i=1}^N \frac{\tilde{\mathbf{p}}_i'^2}{2m_i} + A(\tilde{\mathbf{q}}_1')$, with $A(\tilde{\mathbf{q}}_1') = -\beta^{-1} \log \int d\{\mathbf{p}^{(j)}\} d\{\mathbf{q}^{(j)}\} \delta \left(\frac{1}{P} \sum_{j=1}^P \mathbf{q}^{(j)} - \tilde{\mathbf{q}}_1' \right) \exp \left(-\beta_P H_P^{(2)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\}) \right)$ the free energy landscape (or the potential of mean thermodynamic force) of the centroid of the ring polymer, and $iL^{(\text{CMD})}$ is the Liouvillian of the CMD Hamiltonian. It is straightforward to show that^[80] $\lim_{t \rightarrow 0} C_{AB}^{\text{CMD}}(t) = K_{AB}(0)$, and that in the classical and harmonic limits $C_{AB}^{\text{CMD}}(t) = K_{AB}(t)$ for sufficient number of replicas P . Moreover, since the CMD Liouvillian conserves the quantum canonical Boltzmann distribution^[83], $C_{AB}^{\text{CMD}}(t)$, like the Kubo transformed TCF, also satisfies the detailed balance condition. These properties are used to approximate $K_{AB}(t)$ using CMD away from the limits where it is exact. For general anharmonic potentials, $C_{AB}^{\text{CMD}}(t)$ is accurate up to $\mathcal{O}(t^6)$ ^[85], while for non-linear operators, CMD does not even recover the correct $t \rightarrow 0$ limit^[86].

Performing CMD in its native form is computationally expensive as each force evaluation requires a path integral simulation. Therefore, one normally performs an adiabatic approximation to separate the time scales of the motion of the centroid and the rest of the modes, by specifying lighter dynamical masses to the non centroid modes, allowing an on the fly estimation^[22] of the CMD forces acting on the centroid. This variant of CMD, also known as partially adiabatic (PA-) CMD^[87], is generally used to study systems of practical interest.

Ring polymer molecular dynamics

Within ring polymer molecular dynamics^[88], approximate quantum dynamics is obtained by performing microcanonical dynamics of the standard ring polymer Hamiltonian, and computing TCFs of bead averages of operators. The RPMD approximation to the Kubo transformed TCF of general non-linear position dependent operators \hat{A} and \hat{B} is

$$C_{AB}^{\text{RPMD}}(t) = Z_P^2(N, \mathcal{V}, \beta)^{-1} \left(\frac{1}{2\pi\hbar} \right)^{3N} \int d\mathbf{p}^{(j)} d\mathbf{q}^{(j)} \exp\left(-\beta H_P^{(2)}(\mathbf{p}^{(j)}, \mathbf{q}^{(j)})\right) A^{\text{RPMD}}(\mathbf{q}(0)) \exp(iL^{(\text{RPMD})}t) B^{\text{RPMD}}(\mathbf{q}(0)), \quad (1.44)$$

where $iL^{(\text{RPMD})} = \sum_{j=1}^P iL_{\mathbf{p}\mathbf{q}^{(j)}}^{(\text{PI})} + iL_{\mathbf{p}}^{(\text{PI})}$ is the Liouvillian of the standard path integral Hamiltonian, and $\square^{\text{RPMD}}(\mathbf{q}) = \frac{1}{P} \sum_{j=1}^P \square(\mathbf{q}^{(j)})$ is the bead average of the operator \square . Just like its counterpart CMD, the RPMD TCF satisfies detailed balance^[81], is exact in the classical and the harmonic limits, but for general position dependent operators displays $\lim_{t \rightarrow 0} C_{AB}^{\text{RPMD}}(t) = K_{AB}(0)$, allowing one to approximately compute $K_{AB}(t)$ using RPMD, for general anharmonic potentials and non-linear operators. The position autocorrelation function within RPMD remains accurate up to $\mathcal{O}(t^8)$ ^[85] while, TCFs of non-linear position dependent operators are accurate up to $\mathcal{O}(t^2)$ ^[85].

Practical evaluation of RPMD TCFs requires sampling of initial structures using thermostatted path integral molecular dynamics, and an averaging of TCFs computed over the initial configurations. Since the standard imaginary time Hamiltonian contains a harmonic spring term, RPMD dynamics can be poorly ergodic. A more efficient way of running RPMD is by applying an optimally tuned PILE thermostat^[71] on the non-centroid modes, together with a weak global thermostat on the centroid^[89]. This variant of RPMD, also known as thermostatted (T) RPMD, is much more ergodic, and also recovers the exact Kubo transformed TCF in the classical and the harmonic limits. Furthermore, as we shall see later, it also removes some of the artefacts of RPMD on vibrational spectra^[90].

1.3.3 Challenges associated with the path integral method

The quantum nature of distinguishable nuclei can be simulated in a classical fashion by running path integral molecular dynamics (PIMD), i.e. molecular dynamics for the ring polymer Hamiltonian made of P replicas of the system. This also means that the computational cost of a PIMD simulation is roughly P times that of a standard classical nuclei simulation, due to the cost of computing the energetics of P replicas of the system. The number of replicas required to converge a property depends on the inverse temperature β , the highest characteristic frequency of the system ω^{\max} , and on the property of interest. As a rule of thumb^[91], to converge the energy to less than 10% of its quantum limit, one should choose $P > \beta \hbar \omega^{\max}$. This amounts to an overhead that is between 10 to 50 at room temperature for molecular systems containing hydrogen atoms. More sophisticated properties such as heat capacities – which depend on second derivatives of the partition function – require $P > 2\beta \hbar \omega^{\max}$ to get to the same level of accuracy, leading to an overhead that is typically over a factor of 100^[92]. These estimates illustrate the large cost of modelling the quantum nature of nuclei over classical simulations, and the need for the development of cost reduction schemes, especially when PIMD is to be performed with *ab initio* methods of electronic structure theory.

Over the last two decades, several methods such as ring polymer contraction (RPC)^[91] (see Chapter 2), high order path integrals^[93,94,95] (see Chapter 2), and generalized Langevin equation (GLE) thermostats^[96,71] (see Chapter 4), have been developed which reduce the computational expense of performing PIMD simulations, and have facilitated several important studies^[59,58,97,64,98] that highlight the role of quantum fluctuations across different systems. These methods, however, suffer from several notable limitations. For instance, RPC method which relies on range separation of the potential, into slow and fast term is so far only limited to analytic force-fields. While GLE thermostats, such as PIGLET, can be used in combination with a first principles interatomic potentials, they are applicable only to the study of simple structural properties, and the potential and kinetic energies. More complex observables such as isotope-fractionation ratios^[99] and heat capacities^[92], which require a large number of replicas, are not targetted by GLE thermostats. The high order path integral method can accelerate the convergence of all thermodynamic observables, but requires sec-

ond derivatives of the potential for performing PIMD.

There exist many other properties, such as those that depend explicitly on the momentum operator (with an exception of the kinetic energy) that can not be directly obtained from standard path integral simulations. The particle momentum distribution is one such property which remains fundamentally relevant as it sheds light on the chemical environment of particles^[100], and can be measured quite accurately for protons in different phases of water^[25]. Calculation of this property within the path integral method, requires opening of the otherwise closed path of the target atoms, and the calculation of their end-to-end distribution^[69]. Unfortunately, converging this property requires very long trajectories^[69,95] within the standard open path integral method, and existing cost reduction schemes either compute it in an approximate manner^[71], or require multiple simulations that constrain/restrain path opening in a controlled manner^[101].

Other properties which are directly not related to the observables that we have considered within equilibrium path integral simulations are dynamical properties such as time correlation functions that explicitly depend on physical time. For instance, vibrational spectra^[54,87,82], diffusion coefficients, reaction rates^[102,103], thermal conductivity etc. are a few dynamical properties which are influenced by the quantum nature of nuclei. These properties can be estimated approximately using many schemes based on the path integral method^[81,89,80,79,104,105,106,86,107], all of which have the same limitation (amongst many others) that their cost scales with the number of replicas required to converge the dynamical properties of interest.

Another problem which becomes prominent at ultra low temperatures, is the indistinguishability of particles. So far we have assumed that our particles are "Boltzmannons" meaning that they are distinguishable and follow Boltzmann statistics. However, real particles (at low temperatures) behave either as "Bosons" or "Fermions" meaning that their wavefunction remains symmetric or anti-symmetric, respectively, to the exchange of identical particles.

Including the statistics of exchange within the path integral framework, requires one to make the partition function of the system, symmetric or antisymmetric with respect to exchange of identical particles, by including larger ring polymers that combine combinations of identical particles, contributing to the total partition function either with positive or negative weights (depending on whether the system is Bosonic or Fermionic)^[66]. These larger ring polymers can be sampled in a Monte Carlo fashion, by performing a random walk in permutation space, however, the methodology remains problematic for Fermions for which the sampled probability distribution is not positive definite^[108]. Aside from the challenges associated with sampling Bosonic^[109] and Fermionic^[110] statistical distributions, these methods also suffer from the large computational cost arising from the use of a very large number of replicas, as exchange effects are prominent only at ultra-low temperatures.

It would therefore be beneficial to develop methods that cater to the deficiencies of existing cost reduction schemes, so that path integral methods can be applied to a wide range of estimators, as well as for the description of dynamical properties in a computationally efficient manner. The following three chapters of the thesis, describe novel methodological developments that act on the limitations of RPC, high-order path integral and GLE thermostats, and efficient estimators for the heat capacity, particle momentum distribution and vibrational spectra.

Methodological Developments **Part I**

2 Accelerated path integral methods

The computational cost of a path integral simulation is approximately P times the cost of a classical nuclei simulation, as the former require energetics of P replicas of the system. For most systems of interest, path integral simulations can require up to 10 – 100 times more computational resources at room temperature, which increases linearly as the temperature is reduced. When combined with first principles methods of electronic structure theory that evaluate of energies, forces and virials, these simulations can get prohibitively expensive. Consequently, a number of methods have been developed that reduce the computational cost of path integral simulations. In this chapter, we shall examine the strengths and weaknesses of current cost reduction schemes, and discuss the new developments made by us that solve the limitations of these methods.

The following chapter is partially based on the journal articles:

1. V. Kapil, J. VandeVondele, and M. Ceriotti, "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," *The Journal of Chemical Physics*, vol. 144, p. 054111, Feb. 2016. VK was involved in the development and implementation of the method, in running calculations, in performing analysis and in the writing of the manuscript.
2. V. Kapil, J. Behler, and M. Ceriotti, "High order path integrals made easy," *The Journal of Chemical Physics*, vol. 145, p. 234103, Dec. 2016 VK was involved in the design of the research, in the development and implementation of the method, in running calculations, in performing analysis and in the writing of the manuscript.
3. V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. Van Speybroeck, and M. Ceriotti, "Modeling the Structural and Thermal Properties of Loaded Metal–Organic Frameworks. An Interplay of Quantum and Anharmonic Fluctuations," *Journal of Chemical Theory and Computation*, vol. 15, pp. 3237–3249, May 2019 VK was involved in the design of the research, in the development and implementation of the method, in running some of the calculations, in performing analysis and in the writing of the manuscript.

2.1 Exploiting multiple time / length scales

We assume that the BO potential energy can be written as $V(\mathbf{q}, \mathbf{h}) = V^{\text{short}}(\mathbf{q}, \mathbf{h}) + V^{\text{long}}(\mathbf{q}, \mathbf{h})$, a sum of a short-ranged but cheap to evaluate component $V^{\text{short}}(\mathbf{q}, \mathbf{h})$, and a long-ranged but expensive component $V^{\text{long}}(\mathbf{q}, \mathbf{h})$. Such splittings are usually possible when systems are studied using force-fields that model the short and long ranged interactions using separate analytic functions. Long ranged interactions that arise from dispersion and electrostatics interactions are typically computationally more expensive than short-ranged interactions between neighbouring atoms.

2.1.1 Ring polymer contraction (RPC)

Let's say that the number of beads required to separately converge the short and long ranged components are P and P' , respectively. Since $P' < P$, a simulation that uses the full potential would wastefully compute the expensive $V^{\text{long}}(\mathbf{q}, \mathbf{h})$ component on P replicas. The ideal solution would be to estimate $V^{\text{long}}(\mathbf{q}, \mathbf{h})$ on a smaller ring polymer with P' replicas. The method proposed by Markland and Manolopolous called Ring Polymer Contraction (RPC) describes an efficient way to compute different components of the potential on ring polymers of different sizes. Within this scheme, the coordinates of the ring polymer are transformed into the free ring polymer representation $\{\mathbf{q}^{(j)}\} \rightarrow \{\tilde{\mathbf{q}}^{(j)}\}$, the highest $P-P'$ components are discarded to yield a smaller ring polymer $\{\tilde{\mathbf{q}}^{(j)}\} \rightarrow \{\tilde{\mathbf{q}}^{(j')}\}$, and the potential is computed after transforming back to Cartesian coordinates $\{\tilde{\mathbf{q}}^{(j')}\} \rightarrow \{\mathbf{q}^{(j')}\}$. The total physical potential of the ring polymer is approximated as

$$\begin{aligned} \sum_{j=1}^P V(\mathbf{q}^{(j)}, \mathbf{h}) &= \sum_{j=1}^P V^{\text{short}}(\mathbf{q}^{(j)}, \mathbf{h}) + \sum_{j=1}^P V^{\text{long}}(\mathbf{q}^{(j)}, \mathbf{h}) \\ &\approx \sum_{j=1}^P V^{\text{short}}(\mathbf{q}^{(j)}, \mathbf{h}) + \frac{P}{P'} \sum_{j=1}^{P'} V^{\text{long}}(\mathbf{q}^{(j')}, \mathbf{h}) \end{aligned} \quad (2.1)$$

where, the transformation $\{\mathbf{q}^{(j')}\} = \sum_{j=1}^N T_{j'j} \{\mathbf{q}^{(j)}\}$ can be easily estimated in terms of the eigenvectors of the free ring polymer as discussed in Ref. 91. Similarly the total physical force acting on the j^{th} replica is $\mathbf{f}^{(j)} = \frac{P}{P'} \sum_{j'=1}^{P'} T_{j'j} \mathbf{f}^{(j')}$. This scheme leads to a speedup of factor of $\frac{P}{P'}$, assuming that the computational cost of the short-ranged interactions is negligible in

comparison to that of the long-ranged interactions.

2.1.2 Multiple timestepping (MTS)

Yet another way of exploiting the splitting of the potential is by integrating the fast and slow forces using different timesteps $\delta t'$ and $\Delta t'$, respectively with $\Delta t' > \delta t'$, so that the expensive component that varies slowly in time can be computed less frequently than the cheap one. In the context of classical molecular dynamics, this is achieved by the methodology of Tuckerman *et. al* who suggested writing the total Liouvillian associated with Hamiltonian dynamics as

$$iL^{(\text{cl})} = \sum_{i=1}^N \left[\mathbf{f}_i^{\text{short}} \frac{\partial}{\partial \mathbf{p}_i} + \mathbf{f}_i^{\text{long}} \frac{\partial}{\partial \mathbf{p}_i} + m_i^{-1} \mathbf{p}_i \frac{\partial}{\partial \mathbf{q}_i} \right] = iL_{\mathbf{p}}^{(\text{cl-short})} + iL_{\mathbf{p}}^{(\text{cl-long})} + iL_{\mathbf{q}}^{(\text{cl})} \quad (2.2)$$

and integrating the equations of motion by approximating the propagator as

$$\begin{aligned} \exp(iL^{(\text{cl})} \Delta t') &= \exp(iL_{\mathbf{p}}^{(\text{cl-short})} \Delta t' + iL_{\mathbf{p}}^{(\text{cl-long})} \Delta t' + iL_{\mathbf{q}}^{(\text{cl})} \Delta t') \\ &\approx \exp(iL_{\mathbf{p}}^{(\text{cl-long})} \Delta t') \left[\exp(iL_{\mathbf{p}}^{(\text{cl-short})} \Delta t' / M + iL_{\mathbf{q}}^{(\text{cl})} \Delta t' / M) \right]^M \exp(iL_{\mathbf{p}}^{(\text{cl-long})} \Delta t') \\ &= \exp(iL_{\mathbf{p}}^{(\text{cl-long})} \Delta t') \left[\exp(iL_{\mathbf{p}}^{(\text{cl-short})} \delta t' / 2) \exp(iL_{\mathbf{q}}^{(\text{cl})} \delta t') \exp(iL_{\mathbf{p}}^{(\text{cl-short})} \delta t' / 2) \right]^M \\ &\quad \exp(iL_{\mathbf{p}}^{(\text{cl-long})} \Delta t') \end{aligned} \quad (2.3)$$

where, $\Delta t' = M \delta t'$. This amounts to integration of momenta using the long-ranged force with a large timestep, followed by M steps of velocity-Verlet using the short-ranged force and a short timestep, concluded by integration of the momenta using the long-ranged force with a large timestep. The same methodology can be extended to constant temperature path integral molecular dynamics by replacing the Liouvillian associated with the position in Eq. 2.3 with that of the free ring polymer, and by applying thermostat steps as described in Eq. 1.30. Assuming that the computational cost of the short-ranged interactions is negligible in comparison to that of the long-ranged interactions, this methodology leads to a speedup of a factor of M .

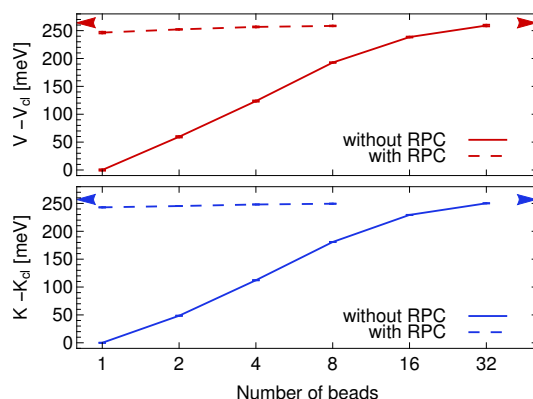


Figure 2.1 – Difference between quantum and classical expectation values of potential (top) and kinetic energy (bottom) per molecule as a function of the number of path integral replicas for a simulation of q-TIP4P/f liquid water at room temperature. Solid curves correspond to full PIMD, while dashed curves correspond to a RPC setup in which the bonded (stretch and bend) terms were computed on 32 replicas, and the non-bonded (electrostatic and dispersion) terms were computed on a reduced number of replicas, that corresponds to the horizontal axis. The blue and red arrow-heads correspond to full PIMD using 64 beads.

2.1.3 Combining MTS and RPC

MTS and RPC techniques can be combined fairly straightforwardly, and indeed, such a combination was used already in early classical simulations based on ring-polymer contraction [82]. The idea of accelerating quantum and classical calculations by combining multiple layers evaluated at increasingly accurate levels of theory have been around for some time [111,112]. However, hitherto no attempt has been reported to use RPC (alone or in combination with MTS) in the context of *ab initio* molecular dynamics, despite the fact that this is a scenario in which obtaining an accurate estimate of nuclear quantum effects at a reduced computational cost would be particularly desirable. This is largely due to the considerable implementation overhead connected with combining sophisticated techniques for efficiently solving the electronic-structure problem and the cumbersome formalism that underlies PIMD and its RPC and MTS extensions. We have presented an implementation of MTS and RPC in *i-PI* – a python interface for advanced path integral simulation – that allows one to combine this combination transparently with any software that computes interatomic energies and forces.

2.1.4 Benchmarks on q-TIP4P/f water

As a test of our implementation, we first study liquid water at 300 K modelled using 216 molecules of q-TIP4P/f water at ambient density. The software LAMMPS was used to calculate the interatomic potential; the short-ranged bonded terms (stretches and bends) were computed on 32 replicas and integrated, with $\delta t' = 0.25$ fs, while the long ranged interactions (dispersion and electrostatics) were estimated on a varying number of beads, and integrated with a range of timesteps $\Delta t'$ to study the accuracy of RPC and MTS. As shown in Fig. 2.1, the convergence of observables to quantum expectation values in a simulation of liquid water at room temperature require a large number of replicas, of the order of $P \approx 30$. However, this is mostly due to the internal modes of each water molecule, that are well described by cheap quasi-harmonic terms. The non-bonded dispersion and electrostatic interactions vary on considerably larger length scales, and therefore a RPC scheme that computes the latter on a reduced number of replicas converges very rapidly. A simple range-separation procedure would allow one to reach convergence with a single evaluation of electrostatic interactions^[113]. Here we want however just to verify to what extent one can push a MTS procedure, alone or in combination with RPC, in a PIMD context.

Due to the presence of high-frequency normal modes in the free ring-polymer Hamiltonian one would expect the first occurrence of resonances to appear already at a timestep of about 1fs for a path integral simulation with 32 replicas at 300K^[114], much earlier than the ≈ 2.7 fs limit that would be predicted due to the stretching mode frequency. Figure 2.2 shows that in this particular case the PIMD simulation remains stable up to an outer step of about 2fs, even in a weakly-thermostatted (white-noise Langevin, thermostat correlation time of 2 ps) PIMD simulation. This rather puzzling finding can probably be ascribed to the exceedingly weak coupling between the physical potential and the high-frequency normal modes of the ring polymer, and does not necessarily apply to different systems with stronger anharmonicities. Here, increasing the dynamical masses of high-frequency ring-polymer modes would delay only minimally the resonance barrier, and would not be possible when using approximate quantum dynamics techniques, such as centroid molecular dynamics (CMD)^[115] and thermostatted ring polymer molecular dynamics (TRPMD)^[81,90], that involve specific prescriptions for the magnitude of the dynamical masses. However, a similarly effective solution to deal with ring-polymer

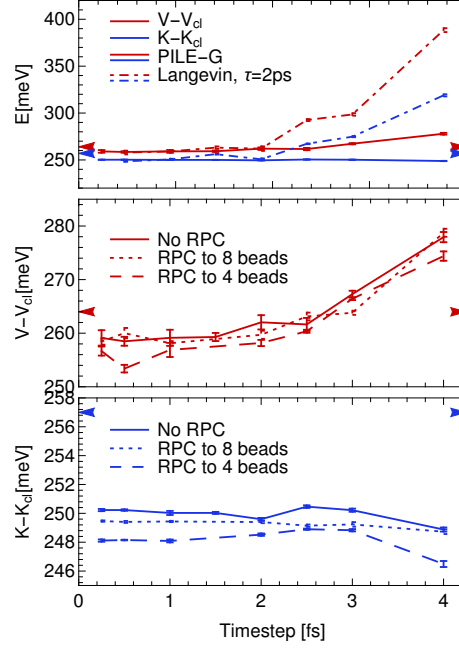


Figure 2.2 – Difference between quantum and classical expectation values of potential (top and middle) and kinetic (top and bottom) energy per molecule for a simulation of q-TIP4P/f liquid water at room temperature, as a function of the outer timestep. The bonded terms were computed with a fixed inner timestep of 0.25 fs, and propagated just outside the free ring-polymer part of the propagator. The non-bonded terms were computed less often, with a time interval as indicated on the horizontal axis. The blue and red curves correspond, respectively, to kinetic and potential energy. Solid curves correspond to 32 beads full PIMD simulation while dotted and dashed curves to simulations in which the non-bonded terms were contracted to 4 and 8 beads respectively. The dot dashed curves in the top panel correspond to simulations using a white noise Langevin thermostat with a time constant of 2 ps while the rest to simulations using a PILE-G thermostat.

resonance barriers is to selectively thermalize the high-frequency vibrations^[116]. Exploiting the approximate knowledge of high-frequency ring polymer modes, one can here simply use optimally-damped Langevin thermostats in the (ring-polymer) normal-modes representation, as explained e.g. in Ref.^[117]. As discussed in Ref.^[118], however, some care must be paid to minimize the effect of normal modes thermostating on the dynamical properties of interest. However, if one is merely interested in structural properties, an efficient thermostating scheme could be the one based on resonance free isokinetic schemes^[119,120], which prevent resonances by limiting the total energy that can be stored in each degree of freedom.

Figure 2.2 shows that the use of an optimally-damped path integral Langevin thermostat^[117]

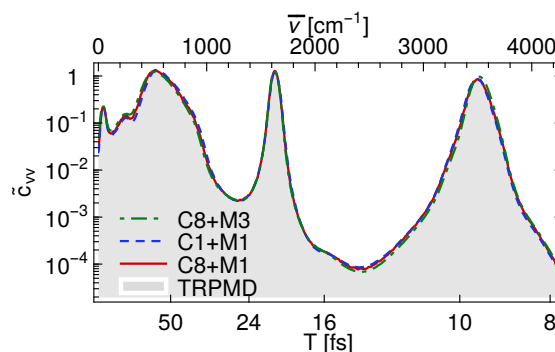


Figure 2.3 – The velocity-velocity autocorrelation function from TRPMD^[90] ($\lambda = 0.5$) runs for q-TIP4P/F water at 300K. A reference run (timestep 0.25fs, 32 beads, shaded area) is compared with a run with 8-beads RPC for the non-bonded interactions, and an outer MTS timestep of 1fs (C8+M1, red continuous line), with a run with 1-bead RPC and 1fs MTS (C1+M1, blue dashed line), and with a run with 8-beads RPC and 3fs MTS (C8+M3, green dot-dashed line).

does actually extend the stability of the method to outer timesteps well above the resonance barrier for the stretches. In fact, the deviations that are observed for the mean potential energy are due to the outer timestep becoming inappropriate to integrate the long-range force (which is relatively quickly varying since we do not use any kind of range separation scheme). This additional stabilization is probably due to the coupling between ring-polymer modes and high-frequency centroid vibrations – which is also the cause for spectral broadening observed in thermostatted RPMD. Dynamical properties appear to be less strongly affected than thermodynamic averages when RPC/MTS techniques are pushed to their limits. Figure 2.3 shows the velocity-velocity autocorrelation function for liquid water computed with TRPMD, and compare the results using full 32-beads PIMD and those obtained with RPC to 8 beads of the non-bonded interactions, and MTS to 1fs outer. Using RPC down to the centroid, or using a 3fs outer timestep, lead to minor deviations from the reference. In summary, these benchmarks demonstrate the effectiveness of our implementation of combined RPC and MTS techniques in *i*-PI, and show that with an appropriate thermostating scheme one can push the stability limit of MTS schemes beyond the resonance barrier of RPMD, for the calculation of both static and dynamical properties (Fig. 2.2 and 2.3).

2.1.5 MTS / RPC combined with electronic structure methods

Having briefly reviewed the main ideas, and demonstrated the stability of RPC+MTS PIMD, let us move on to discuss a few aspects that are specific to a first principles context. To give a practical example, we simulated the quantum distributions of the Zundel cation at the MP2 level of theory, using semi-local DFT as the short-ranged component, and the difference between MP2 and DFT as the long-ranged component. While other choices are possible, this combination is really attractive, as it is general in nature, and the impact of the scheme can be very significant. Indeed, as system size increases, MP2 and semi-local DFT display very different computational costs, since traditional implementations scale as $O(N^5)$ and $O(N^3)$ respectively. For example, three orders of magnitude time difference can be expected for samples of 64 water molecules^[121]. The excellent scalability of the MP2 implementation^[122,123] in CP2K^[124], nevertheless enables an acceptable time to solution. In the MP2+RPC context a large number of cheap DFT calculations must be combined with the expensive MP2 calculations, and the question of load balancing, avoiding idle processes as much as possible, imposes itself. As a simple, yet efficient, solution to this problem, we over-subscribe compute nodes using two processes per core. One process belongs to the P DFT tasks, while one process belongs to the P' MP2 tasks and the contracted DFT tasks, which run at the same time as MP2. While the MP2 tasks compute, the full-path DFT tasks are effectively sleeping, and vice versa. In this way, idle resources are avoided, and only a small fraction of total wall time is spent in the DFT part, which is naturally parallel over the beads, while the computation for each bead is parallelised as well. If the contraction is pushed all the way to the centroid, the cost of performing MP2+DFT with MTS+RPC is basically the same as the cost of doing standard MP2 MD.

The combination of MP2 with semi-local DFT employs the same computational setup as described in Ref.^[121]. In particular, this approach is based on the resolution of identity Gaussian and Plane Waves (RI-GPW) which provides an efficient and scalable approach to perform MP2 based MD in gas and condensed phases.^[126,122,123,127] The Gaussian basis employed for MP2 is of the correlation consistent triple zeta quality^[126] and is parametrized for the pseudopotentials employed.^[128] The optimized density functional (ODF) discussed below is of the GGA family, starting from the PBE1W^[129] functional, for which the small basis and

2.1. Exploiting multiple time / length scales

Simulation	K_H [meV]	K_O [meV]	d_{OO} [Å]
MP2	148.4(5)	58.9(4)	2.413(4)
PBE	143.1(5)	57.5(4)	2.445(4)
PBE+C	143.2(5)	58.5(3)	2.409(5)
PBE+C2+M	144.7(6)	58.3(4)	2.410(1)
ODF	148.9(5)	58.9(4)	2.416(3)
ODF+C	149.3(5)	59.6(4)	2.413(4)
ODF+C+M	149.7(3)	59.3(3)	2.410(3)
ODF+C+M2	149.4(4)	59.4(3)	2.412(5)
ODF+C+M+G	148.0(9)	59.4(5)	2.412(2)
HFX+C+M	155.3(7)	60.6(5)	2.411(3)

Table 2.1 – Expectation values of quantum kinetic energies per H and O atom, and for the O-O distance, in a simulation of a gas-phase Zundel cation at 300K. All simulations were performed using 32 beads and a base timestep of 0.25fs, except that using PIGLET (+G) that used six replicas. PBE refers to use of the standard PBE generalized gradient approximation (GGA) functional^[125], ODF refers to a GGA optimized to MP2^[121], and HFX to Hartree-Fock. Ring-polymer contraction was used to reduce the cost by computing the MP2 forces on the centroid only (+C), or on two beads (+C2). A multiple timestep algorithm was also employed to evaluate such force only once every 1fs (+M) or 2fs (+M2). All simulations were run for 10ps, including 1ps for equilibration. Statistical errors on the last digit are reported in parentheses.

van der Waals D3 parameters^[130] have been specifically refitted in order to reproduce the energetics of bulk liquid water.^[121]

Results for the mean kinetic energy of hydrogen and oxygen atoms, and the mean O-O distance, for different simulations, are reported in Table 2.1. The table caption summarizes the details of the simulations. The standard PBE functional underestimates the H kinetic energy with single-bead contraction, improving as expected with a contraction to two beads. This indicates that PBE differs too much from the reference, requiring a considerable increase in the number of MP2 replicas. Also Hartree-Fock is not very accurate, this time predicting a too large value of the H kinetic energy, despite the RPC correction. This suggests that the MTS strategy proposed in Ref.^[131], which combines HFX and MP2, cannot be used universally for RPC, at least not without a large number of correction beads. Hybrid functionals, which mix GGA exchange with Hartree-Fock exchange, will likely provide a suitable intermediate potential energy surface. However, as a computationally less demanding, but less universal alternative, an optimized semi-local DFT (ODF) shows excellent performance. The results reported in

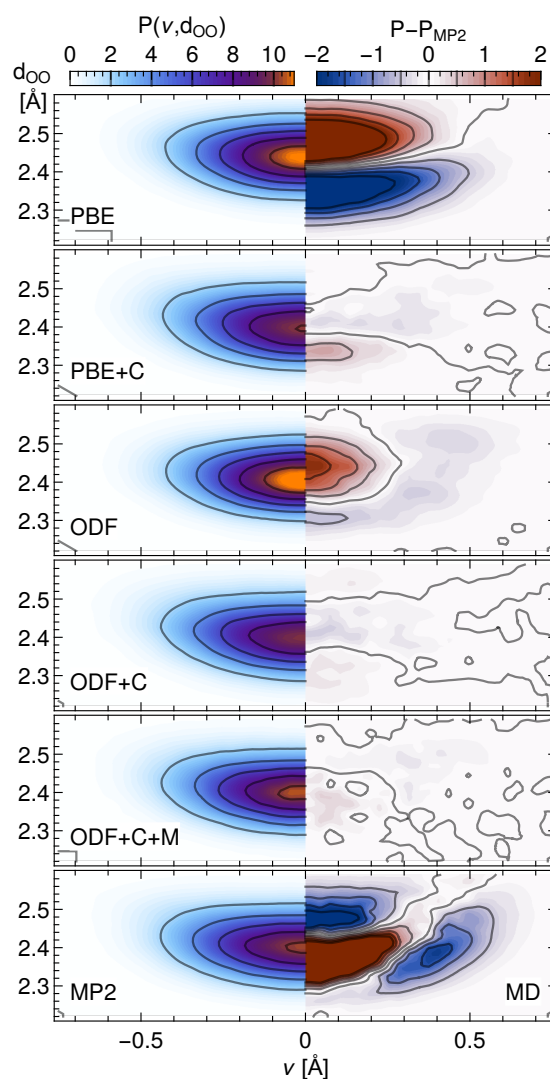


Figure 2.4 – Joint probability distribution for the proton transfer coordinate and the O-O distance in the gas-phase Zundel cation at 300K, modeled with nuclear quantum effects and different electronic structure methods (left panels). Right-hand panels show the difference with respect to the MP2 reference. The bottom-right panel shows the difference between a classical MP2 MD simulation and the (quantum) MP2 reference.

Table 2.1 are within the statistical error bars from the MP2 reference already without any RPC correction, and as a consequence it is not possible to detect statistically significant effects on these averages due to RPC and/or MTS. However, also in this case the RPC correction does have a noticeable effect. This is shown in Figure 2.4 using the joint probability distribution of the O-O distance and the proton transfer coordinate. The difference between the MP2 and the ODF+C is almost zero and smaller than ODF only. The remaining error is largely due to the statistical uncertainty in the probability distributions, which is relatively costly to reduce for the reference MP2 simulation that employs 32 MP2 beads.

Let us finally note that RPC and MTS methods can be seamlessly combined with other strategies to reduce the cost of a PIMD calculation: as demonstrated in Table 2.1, the use of a colored-noise PIGLET thermostat^[132] makes it possible to reduce the number of baseline beads to six, that would be advantageous in cases where the cost of the GGA calculations is not negligible. The implementation we have introduced in *i-PI* to perform these calculations is completely general, and multiple levels of MTS and RPC can be easily combined if one wanted to reach the ultimate level of cost/accuracy optimization.

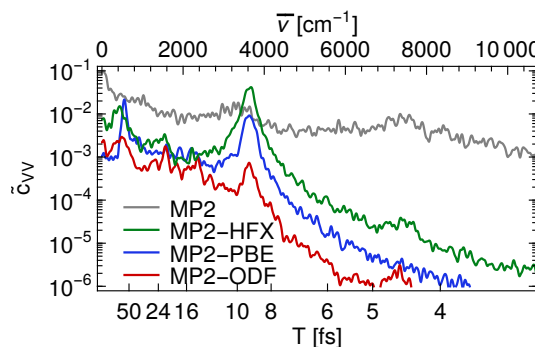


Figure 2.5 – Fourier-transform of the potential-potential correlation function for different RPC simulations. The curve for MP2 (in grey) is taken as a reference; being evaluated on 32 replicas, it also contains the high-frequency non-centroid modes of the ring polymer. The other two curves correspond to the correction potential (MP2 minus GGA) evaluated on the centroid for the HFX (green), PBE (blue) and optimized DFT (red) simulations. The horizontal scale indicates the period of different vibrational modes, for ease of reference. Since these simulations were heavily thermostatted for sampling and stability, the spectrum has no physical meaning, and is just a tool to assess the time scales that are relevant for different potential components.

A simple way to assess the viability of accelerating a simulation by means of RPC and/or MTS is to verify the magnitude and time scale for the variation of the correction potential. Fig. 2.5 shows the Fourier transform of the correction potential correlation function $\langle V^{\text{short}}(t) V^{\text{short}}(0) \rangle$. Clearly, the difference between MP2 and HFX contains strong high-frequency components, which would require a high number of beads in the contracted V^{short} . Using the PBE GGA shows only a marginal advantage, as the spectrum still has a pronounced peak at the stretching frequencies. On the contrary, the ODF exhibits a smaller difference, and in particular a density of states smaller by an order of magnitude for the stretching region - which explains the excellent performance in reproducing quantum kinetic energies even without RPC.

2.2 High-order path integrals

In standard (second-order) path integral schemes, an approximate factorization of the high-temperature Boltzmann operator is introduced, that leads to an error that decreases as $\mathcal{O}(1/P^2)$. High-order techniques use an alternative splitting of the Boltzmann operator^[133,94], leading to an alternative ring polymer Hamiltonian with a faster, $\mathcal{O}(1/P^4)$ convergence to the exact quantum limit. This makes it possible to reduce the number of replicas and hence the computational cost.^[134] While many high-order schemes exist^[133,135], here we focus on the specific case^[2] of a fourth-order Suzuki-Chin (SC) splitting

$$\begin{aligned} Z(N, \mathcal{V}, \beta) &= \text{Tr} \left[\exp(-\beta_P \hat{H})^P \right] \\ &= \text{Tr} \left[\exp(-\beta_P \hat{V}_e/3) \exp(-\beta_P \hat{T}) \exp(-4\beta_P \hat{V}_e/3) \right. \\ &\quad \left. \exp(-\beta_P \hat{T}) \exp(-\beta_P \hat{V}_e/3) \right]^{\frac{P}{2}} + \mathcal{O}(P^{-4}), \end{aligned} \quad (2.4)$$

where $\hat{V}_e = \hat{V} + \frac{\alpha}{6} \beta_P^2 [\hat{V}, [\hat{T}, \hat{V}]]$, $\hat{V}_o = \hat{V} + \frac{1-\alpha}{12} \beta_P^2 [\hat{V}, [\hat{T}, \hat{V}]]$ and $\alpha \in [0, 1]$ is a dimensionless parameter that can be adjusted to improve the convergence for a given problem. It seems, however, that no generally-applicable prescription for its choice can be obtained. In the present study, for reasons that will become apparent later on, we always used $\alpha = 0$. We proceed as we did in the case of second-ordered PIMD, and compute the trace in Eq. 2.4 in the position basis $|\mathbf{q}^{(j)}\rangle$ and insert $P - 1$ completeness of identities $\mathbf{1} = \int d\mathbf{q}^{(j)} |\mathbf{q}^{(j)}\rangle \langle \mathbf{q}^{(j)}|$ between

the potential (\hat{V}_\square) and kinetic energy operators. It is easy to show that for $\alpha = 0$, the fourth ordered canonical partition function reads

$$Z_P^{(4)}(N, \mathcal{V}, \beta) = \left(\frac{1}{2\pi\hbar} \right)^{3NP} \int d\{\mathbf{p}^{(j)}\} d\{\mathbf{q}^{(j)}\} \exp\left(-\beta_P H_P^{(4)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\})\right), \quad (2.5)$$

the classical canonical partition function of the fourth-ordered ring-polymer Hamiltonian

$$H_P^{(4)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\}) = \sum_{j=1}^P \left[\sum_{i=1}^N \frac{1}{2} m_i^{-1} |\mathbf{p}_i^{(j)}|^2 + \sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 (\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)})^2 \right] + \sum_{j=1}^{P/2} \left[\frac{2}{3} V(\mathbf{q}^{(2j-1)}, \mathbf{h}) + \frac{4}{3} V(\mathbf{q}^{(2j)}, \mathbf{h}) + \frac{1}{9} \tilde{V}(\mathbf{q}^{(2j)}, \mathbf{h}) \right]. \quad (2.6)$$

Note that the odd and even replicas feel the physical potential $V(\mathbf{q}^{(j)}, \mathbf{h})$ scaled by factors of $2/3$ and $4/3$ respectively and that the high-order term $\tilde{V}(\mathbf{q}^{(j)}, \mathbf{h}) = \omega_P^{-2} \sum_{i=1}^N m_i^{-1} |\mathbf{f}_i^{(j)}|^2$, that depends on the modulus of the force only acts on the even replicas.

The thermodynamic average of any position dependent observable \hat{A} can either be estimated as $\langle A(\mathbf{q}) \rangle_P^{(4)} = Z_P^{(4)}(N, \mathcal{V}, \beta)^{-1} \text{Tr}[\exp(-\beta \hat{H}) \hat{A}]$, or by deriving the partition function as in Eq. 1.7. The former are called operators (OP) estimators, while the latter are referred to as thermodynamic (TD) estimators. The OP estimator of \hat{A}

$$\langle A(\mathbf{q}) \rangle_P^{OP(4)} = Z_P^{(4)}(N, \mathcal{V}, \beta)^{-1} \int d\{\mathbf{p}^{(j)}\} d\{\mathbf{q}^{(j)}\} \left[\frac{2}{P} \sum_{j=1}^{P/2} A(\mathbf{q}^{(2j-1)}) \right] \exp\left(-\beta_P H_P^{(4)}(\{\mathbf{p}^{(j)}\}, \{\mathbf{q}^{(j)}\})\right) \quad (2.7)$$

only depends on the instantaneous average of the observable over odd replicas. The thermodynamic estimators of observables are generally much more complex and in general require *ad hoc* expressions that depend on higher derivatives of the potential (arising trivially from the partition function).

2.2.1 Challenges associated with high-order path integrals

The difficulty in applying high-order schemes can be understood by considering the fact that the force and virial contain derivatives of \tilde{V} with respect to the atomic positions and the cell

parameters,

$$\tilde{\mathbf{f}}_i^{(j)} \equiv -\frac{\partial \tilde{V}(\mathbf{q}^{(j)}, \mathbf{h})}{\partial \mathbf{q}^{(j)}}, \quad \tilde{\Xi}^{(j)} \equiv \frac{d\tilde{V}(\mathbf{q}^{(j)}, \mathbf{h})}{d\mathbf{h}} \mathbf{h}^T. \quad (2.8)$$

Given that $\tilde{V}(\mathbf{q}^{(j)}, \mathbf{h})$ already contains first-order derivatives of the physical potential, the computation of the forces and the virial, required to sample the canonical and isothermal-isobaric ensembles by means of path integral dynamics, also demands the calculation of high-order derivatives of the potential, which is often cumbersome and computationally prohibitive. Much of the work on the practical implementation of high-order path integrals has therefore focused on avoiding the calculation of these terms, by sampling the standard path integral Hamiltonian and introducing fourth-order statistics by re-weighting^[134,95]. Unfortunately, re-weighting schemes have poor statistical performance for large systems^[134,136], so the application of the SC scheme has until now been limited to small systems and to constant-volume sampling. An alternative way of estimating high-order observables is by using perturbation theory^[137], which can be perceived as a truncated cumulant expansion of the ratio of second and fourth-order partition functions. Although this scheme remains efficient, it requires complicated *ad hoc* estimators for computing thermodynamic averages.

2.2.2 Estimating high-order forces and virials

One should notice, however, that the expression for $\tilde{\mathbf{f}}$ involves the second derivative of V *projected on the mass-scaled force*. As it has been recognized in the context of high-order path integral Monte Carlo^[138,139], and similarly to what has been done for instance in electronic structure theory^[140], it is possible to evaluate this kind of projected second derivatives by finite differences (FD),

$$\tilde{\mathbf{f}}_i^{(2j)} = \omega_P^{-2} m^{-1} \frac{\partial}{\partial \mathbf{q}_i^{(2j)}} |\mathbf{f}_i^{(2j)}|^2 = -2\omega_P^{-2} \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon \delta} \left(\mathbf{f}_i^{(2j)} \Big|_{\mathbf{q}_i^{(2j)} + \epsilon \delta^{(2j)} \mathbf{u}_i^{(2j)}} - \mathbf{f}_i^{(2j)} \right), \quad (2.9)$$

where $\mathbf{u}_i^{(2j)} = \mathbf{f}_i^{(2j)} / m_i$ and $\delta^{(2j)} = \left[(3NP)^{-1} \sum_{i=1}^N \mathbf{u}_i^{(2j)} \cdot \mathbf{u}_i^{(2j)} \right]^{-\frac{1}{2}}$ is a normalization factor, so that ϵ represents the root mean square displacement applied to each atom. This avoids the explicit calculation of the Hessian and allows for the direct sampling of the Suzuki-Chin canonical ensemble. Following a similar strategy, we show that the high-order component of

the virial can be estimated as:

$$\begin{aligned}\tilde{\Xi}^{(2j)} &= \left[-\sum_{i=1}^N \mathbf{q}_i^{(2j)} \otimes \tilde{\mathbf{f}}_i^{(2j)} + \frac{\partial \tilde{V}(\mathbf{q}_i^{(2j)}, \mathbf{h})}{\partial \mathbf{h}} \mathbf{h}^T \right] \\ &= -2\omega_P^{-2} \left[\sum_{i=1}^N \mathbf{f}_i^{(2j)} \otimes \mathbf{f}_i^{(2j)} / m_i + \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon \delta^{(2j)}} \left(\Xi^{(2j)} \Big|_{\mathbf{q}^{(2j)} + \epsilon \delta^{(2j)} \mathbf{u}^{(2j)}} - \Xi^{(2j)} \right) \right].\end{aligned}\quad (2.10)$$

These expression can be used seamlessly to propagate the equations of motion and to evaluate the estimators for thermodynamic and structural properties. The crucial aspect that makes this procedure viable is that the derivative yields a rigorously time-reversible and symplectic integrator, when combined with a velocity-Verlet integrator. As a result, the scheme is stable even for relatively large values of the finite-difference step, which is advantageous e.g. when evaluating the forces in *ab initio* calculations, where residual errors in the convergence of the self-consistent solution to the electronic structure problem inevitably lead to noisy forces.

2.2.3 MTS integrators for high-order path integrals

Note that using the finite difference estimators, the SC scheme is a factor of 1.5 more expensive than standard PIMD. This relatively small overhead can be reduced even further by using a MTS integrator for the high-order forces and virials. We note that the prefactor of $\tilde{\mathbf{f}}^{(j)}$ and $\Xi^{(j)}$ is ω_P^2 , which for realistic values of P , leads to small and slowly varying forces and virials. The time propagator for the fourth-order Hamiltonian over the timestep $\Delta t'$ can be written as

$$\begin{aligned}\exp(iL^{\text{SCPI-LE}} \Delta t') &= \exp(iL_{\gamma}^{\text{PI-LE}} \Delta t') \exp(i\tilde{L}_{\mathbf{p}}^{\text{SCPI}} \Delta t') \\ &\quad \left[\exp(iL_{\mathbf{p}}^{\text{SCPI}} \delta t') \exp(iL_{\mathbf{q}}^{\text{PI}} \delta t') \exp(iL_{\mathbf{p}}^{\text{SCPI}} \delta t') \right]^M \\ &\quad \exp(i\tilde{L}_{\mathbf{p}}^{\text{SCPI}} \Delta t') \exp(iL_{\gamma}^{\text{PI-LE}} \Delta t'),\end{aligned}\quad (2.11)$$

which implies the thermostat step, followed by the update of the momenta using the high-order forces for half the outer timestep $\mathbf{p}^{(j)} \rightarrow \mathbf{p}^{(j)} + \tilde{\mathbf{f}}^{(j)} \Delta t' / 2$, followed by velocity-Verlet for the pseudo second-ordered Hamiltonian with the potential scaled by weighs 2/3 and 4/3 for the odd and even replicas, respectively, for $M = \Delta t' / \delta t'$ steps using a small timestep $\delta t'$, followed by the update of the momenta using the high-ordered forces for half the outer timestep, concluded by the thermostat step. In the context of constant pressure simulations, MTS can

be implemented by using the propagator

$$\begin{aligned} \exp(iL^{\text{SCPI-BZP}}\Delta t') &= \exp(iL_{\gamma}^{\text{PI-BZP}}\Delta t') \exp(i\tilde{L}_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\Delta t') \\ &\quad \left[\exp(iL_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\delta t') \exp\left(\left(\sum_{j=2}^P iL_{\mathbf{p},\mathbf{q},j}^{\text{PI}} + iL_{\mathbf{p},\mathbf{q},\mathcal{V}}^{\text{PI-BZP}}\right)\delta t'\right) \right. \\ &\quad \left. \exp(iL_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\delta t') \right]^M \exp(i\tilde{L}_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\Delta t') \exp(iL_{\gamma}^{\text{PI-BZP}}\Delta t'), \end{aligned} \quad (2.12)$$

where $\exp(i\tilde{L}_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\Delta t')$ implies

$$\begin{aligned} \alpha &\rightarrow \alpha + (\Delta t'/2) [3\mathcal{V}(\tilde{\mathcal{P}}_P^{\text{int}} - \mathcal{P}_P^{\text{ext}})] \\ \tilde{\mathbf{p}}_i^{(j)} &\rightarrow \tilde{\mathbf{p}}_i^{(j)} + \tilde{\mathbf{f}}_i^{(j)} \Delta t'/2, \end{aligned} \quad (2.13)$$

with $\tilde{\mathcal{P}}_P^{\text{int}} = \frac{1}{3\mathcal{V}} \text{Tr} \left[-\sum_{j=1}^P \sum_{i=1}^N (\mathbf{q}_i^{(j)} - \bar{\mathbf{q}}_i) \otimes \tilde{\mathbf{f}}_i^{(j)} - \sum_{j=1}^P \tilde{\Xi}^{(j)} \right]$ the high-order component of the pressure, and $\exp(iL_{\mathbf{p},\alpha}^{\text{SCPI-BZP}}\Delta t')$ corresponds to

$$\begin{aligned} \alpha &\rightarrow \alpha + (\delta t'/2) [3\mathcal{V}(\mathcal{P}_P^{\text{int}} - \mathcal{P}_P^{\text{ext}}) + \beta_P^{-1}] \\ &\quad + (\delta t'/2)^2 \sum_{i=1}^N m_i^{-1} \tilde{\mathbf{p}}_i^{(1)} \cdot \tilde{\mathbf{f}}_i^{(1)} + \frac{1}{3} (\delta t'/2)^3 \sum_{i=1}^N m_i^{-1} \tilde{\mathbf{f}}_i^{(1)} \cdot \tilde{\mathbf{f}}_i^{(1)}, \\ \tilde{\mathbf{p}}_i^{(j)} &\rightarrow \tilde{\mathbf{p}}_i^{(j)} + \tilde{\mathbf{f}}_i^{(j)} \delta t'/2, \end{aligned} \quad (2.14)$$

such that $\mathcal{P}_P^{\text{int}}$ and $\tilde{\mathbf{f}}_i^{(j)}$ are respectively the pressure and forces associated with the pseudo second-ordered Hamiltonian with the potential scaled by factors of 2/3 and 4/3 for the even and odd replicas, respectively.

The MTS scheme reduces the cost of the SC scheme to $1 + \frac{1}{2M}$ times that of standard PIMD, leading to a negligible overhead for realistic values of $M = 4, 8$. This makes the SC scheme already advantageous at room temperature, and its lead substantial for low-temperature or high-accuracy studies. A similar scheme can be easily derived in the context of other high-order factorizations such as that introduced by Takahashi and Imada^[133]. It is also possible to implement perturbed path estimators^[135] on top of a full fourth-order path integral Hamiltonian, providing even faster convergence to quantum expectation values. Finally, further dramatic speed-ups can be obtained whenever one can apply range-separation techniques such as ring-polymer contraction^[91], since we have made sure that our implementation in

i-PI^[141] is fully compatible with that of conventional real and imaginary-time multiple time stepping^[142].

2.2.4 A GLE thermostat for high-order path integrals

Having access to direct sampling of the SC Hamiltonian opens up the possibility of combining high-order path integrals with a generalized Langevin equation acceleration. In the Trotter case, the normal mode (NM) eigenvectors of the Hamiltonian for a harmonic potential $V(q) = m\omega^2 q^2/2$ do not depend on the frequency ω itself, that only leads to a shift to the NM frequencies. This makes it possible to apply sophisticated thermostating strategies, with different GLEs applied to individual NM coordinates – all without the need of knowing the NM decomposition of the physical potential^[132]. Unfortunately, this is not the case for the Suzuki-Chin Hamiltonian. However, since the NM transformation remains an orthogonal transformation, it is possible to apply a single GLE to all Cartesian (or Trotter NM) coordinates which gives the same effect as applying such GLE onto the proper SC NMs.

In the harmonic limit, for a physical potential of frequency ω , the frequencies $\omega_k(\omega)$, and the eigenvectors $\mathbf{u}^{(k)}(\omega)$ of the SC Hamiltonian can be obtained by diagonalizing the dynamical matrix $D_{jj'}$, that reads

$$D_{jj'} = \begin{cases} 2\omega_P^2 + \frac{2}{3}\omega^2 & j = j' \in \text{Odd} \\ 2\omega_P^2 + \frac{4}{3}\omega^2 \left(1 + \frac{1}{6} \left(\frac{\omega}{\omega_P}\right)^2\right) & j = j' \in \text{Even} \\ -\omega_P^2 & j = j' \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

where the cyclic boundary conditions $j + P \equiv j$ are implied. A further complication stems from the fact that the estimator for the fluctuations of q is not just an average over the coordinates of all beads. In fact, one has to choose whether to design the GLE so as to speed up the convergence of either the TD or the OP estimator for $\langle q^2 \rangle$. Given its simplicity, and the direct connection with all structural observables, we opted for the latter choice, that gives

$$\langle q^2 \rangle = \frac{2}{P} \sum_{j \in \text{odd}} \langle [q^{(j)}]^2 \rangle = \frac{1}{P} \sum_k U_k \langle [\tilde{q}^{(k)}]^2 \rangle \quad (2.15)$$

Chapter 2. Accelerated path integral methods

where $U_k = 2 \sum_{j \in \text{odd}} |u_j^{(k)}|^2$ gives the weight of the k -th mode on the displacement of the even beads.

The design of the effective-temperature curve $T^*(\omega)$ can then proceed in a similar way to what was done in Refs. ^[143,132]: considering that for a classical oscillator $\langle q^2 \rangle = \frac{1}{m\beta\omega^2}$ one can write the functional equation

$$\frac{\hbar}{2\omega k_B} \coth \frac{\hbar\omega\beta}{2} = \frac{1}{P} \sum_k U_k(\omega) \frac{T^*(\omega_k(\omega))}{\omega_k(\omega)^2}. \quad (2.16)$$

Here we made explicit the dependence of the eigenvector coefficients and of the normal modes frequencies on the physical frequency of the underlying potential. Solution of Eq. (2.16) can be obtained by singling out the lowest-lying NM, obtaining the iteration

$$T^*(\omega_0) = \frac{\omega_0^2}{U_0} \left[\frac{P\hbar}{2\omega(\omega_0) k_B} \coth \frac{\hbar\omega(\omega_0)\beta}{2} - \sum_{k>0} U_k(\omega_0) \frac{T^*(\omega_k(\omega_0))}{\omega_k(\omega_0)^2} \right] \quad (2.17)$$

that can be made to converge with an appropriate mixing scheme ^[143] and with the starting condition

$$T^*(\omega_0) = \frac{\hbar\omega_0}{\sqrt{6}k_B} \coth \left(\frac{\beta\hbar\omega_0}{\sqrt{6}P} \right) \quad (2.18)$$

Yet another complication associated with using a GLE in connection with SC path integrals is that the lowest normal-mode frequency is not equal to the physical frequency, as in Trotter PIMD. For this reason, one needs to invert the $\omega_0(\omega)$ relation to find what is the physical frequency that corresponds to the argument of T^* we are solving for. In a similar way, one can then obtain the frequencies of the higher NMs as a function of the lower frequency ω_0 , which eventually makes it possible to solve numerically the iteration in Eq. 2.17. In fact, it is possible to give a closed (albeit cumbersome) expression for such inverse relation ^[144]

$$\omega(\omega_0) = \frac{2^{2/3} A^{2/3} + \sqrt[3]{A} (4x_0^2 - 6P^2) + 4\sqrt[3]{2} x_0^2 (3P^2 + 2x_0^2)}{8\beta\hbar\sqrt[3]{A}} \quad (2.19)$$

$$A = 27P^6 - 72P^2 x_0^4 + 16x_0^6 + 3 [81P^{12} - 432P^8 x_0^4 + 384P^4 x_0^8 - 384P^2 x_0^{10}]^{1/2}$$

$$x_0 = \beta\hbar\omega_0.$$

This thermostating scheme enforces different fluctuations on different ring-polymer vibrations so as to obtain converged quantum expectation values for *any* OP-method estimator of position-dependent properties in the harmonic limit, and for any number of beads. Contrary to the Trotter case, where one can further tune ring-polymer fluctuations to accelerate the convergence of the centroid-virial kinetic energy estimator, this is not possible here, so we can expect that the convergence of the quantum kinetic energy will be less efficient than with the PIGLET approach^[132]. GLE parameters enforcing the desired temperature curve for this Suzuki-Chin GLE (SC+GLE) approach have been obtained following the fitting protocol discussed in Ref.^[145], and are available for download from an on-line repository^[146].

2.2.5 Benchmarks on water using a dissociable potential

For a comprehensive benchmark study of the methods discussed in the previous section we will use simulations of water, a prototypical system for the investigation of nuclear quantum effects. For this purpose we use a neural network (NN) potential fitted to ab initio calculations performed with the B3LYP hybrid density functional^[147] and the D3 dispersion corrections by Grimme^[148], as implemented in CP2K^[149]. The potential is fully reactive, i.e. it allows for the possibility of bond breaking and formation, and has recently been shown to provide an excellent description of nuclear quantum effects in water, as probed by isotope fractionation and the nuclear quantum kinetic energy^[150], at the same time allowing us to obtain thorough sampling. The potential was evaluated using a NN implementation^[151] for LAMMPS^[152]. Unless otherwise specified, each result we report involved a trajectory of at least 200 ps for a supercell containing 128 molecules at the experimental density. We enforced constant-temperature sampling at $T = 300$ K using a PILE scheme^[117] with $\gamma_k = \omega_k/2$, and a weak, global thermostat on the centroid – so that effectively canonical-sampling runs correspond to the thermostated ring-polymer molecular dynamics (TRPMD) protocol^[90] suitable to discuss quantum dynamical properties. In order to probe the behavior of our approach in a lower-temperature regime, we also performed simulations of a 96-molecules box of hexagonal ice at $T = 100$ K. For colored-noise simulations we used the PIGLET thermostat^[132] for Trotter PI, and the SC+GLE strategy discussed above for SC PIMD. These calculations will be a challenging test case for our techniques, because the reactive nature of the NN potential allows for quantum

fluctuations of the hydrogen bond probing the strongly anharmonic regions in the potential energy surface of water^[58].

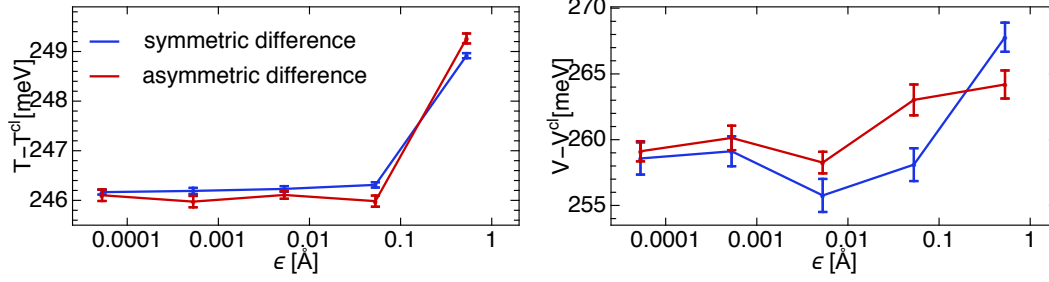


Figure 2.6 – Expectation values of the quantum contributions per molecule to the potential energy V as well as to the kinetic energy T , as a function of the finite-difference displacement ϵ , for a SC-PIMD simulation of liquid water at 300 K performed with 16 beads. The two sets of points correspond to the symmetric (blue) and the asymmetric (red) finite-difference integrators. Error bars indicate the statistical error, which is of the order of 1% for the potential and of the order of 0.1% for the kinetic energy. The superscript “cl” indicates the classical component. Note that given the definition of the displacement vector \mathbf{u} , the finite-difference step ϵ indicates the root-mean-squared displacement of an atom during the evaluation of the derivative.

Stability of the finite-difference scheme

A possible problem that one has to be aware of when using force evaluation schemes based on discrete approximations of the derivatives is that in many cases – most notably for *ab initio* simulations – imperfect convergence of self-consistency schemes can introduce numerical noise. In particular, when using a small displacement in a finite difference scheme, the signal-to-noise ratio degrades, which can lead to instabilities in the integration of the equations of motion. Therefore it is important to test how sensitive are the results to the specific value of the atomic displacement. Figure 2.6 shows that the SC integrators we introduce here, due to their time-reversibility and symplectic properties, show exceptionally good stability, with no appreciable effect of the root mean squared atomic displacement on the quantum expectation values of the potential and kinetic energies for $\epsilon \lesssim 0.1\text{\AA}$.

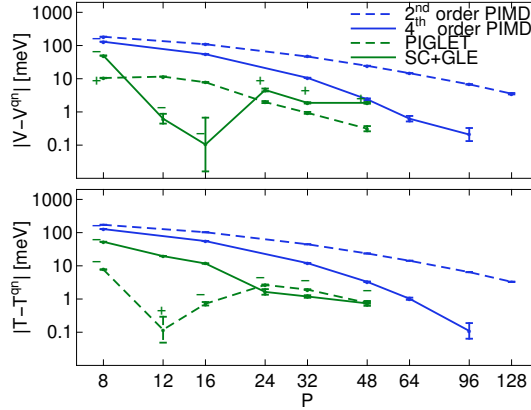


Figure 2.7 – Error per molecule on the value of the potential energy V as well as the kinetic energy T , as a function of the number of beads P , for a simulation of liquid water at 300 K performed with second and fourth-order PIMD, and with the corresponding colored-noise methods, namely PIGLET (second-order) and SC+GLE (fourth-order). We only report here the OP-method estimators. The fully-converged value is taken to be SC PIMD with $P = 48$, and errors are plotted on a log-log scale to highlight the faster convergence of fourth-order methods.

Convergence of energy estimators

The most straightforward measure for the convergence of a PIMD method to the quantum limit is given by the potential and kinetic energy estimators. Figure 2.7 shows such convergence tests, comparing Trotter and SC path integrals with and without colored noise. Results are in line with the expectations. Fourth-order PIMD gives a much improved asymptotic convergence, without the statistical instabilities observed in re-weighting strategies^[136] and giving with $P = 16$ results that are superior to Trotter PI with $P = 32$. The number of evaluation of $\tilde{\mathbf{f}}_i^{(j)}$ can be reduced with a MTS scheme. Even by computing the SC force as often as every $M = 2$ steps, it can be clearly seen that also at room temperature our finite-differences implementation of high-order path integrals provides higher accuracy at a smaller cost than Trotter PIMD. As shown in figure 2.8, the improvement becomes even more significant as the temperature is lowered. In a simulation of hexagonal ice at $T = 100\text{K}$, SC PIMD reaches an error of a few meV per molecule when $P = 48$. When using Trotter PIMD, one would need to use more than 128 beads to obtain a similar accuracy.

GLEs improve significantly the convergence of both standard and fourth-order PIMD, giving potential energies that are within a few percent of the converged results with as few as 4-6 beads

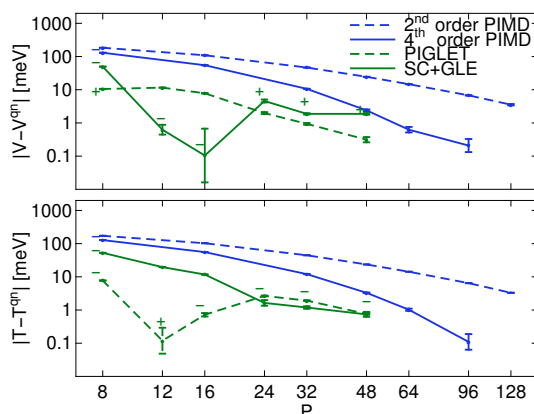


Figure 2.8 – Error per molecule on the value of the potential energy V as well as the kinetic energy T , as a function of the number of beads P , for a simulation of ice at 100 K performed with second and fourth-order PIMD, and with the corresponding colored-noise methods, namely PIGLET (second-order) and SC+GLE (fourth-order). We only report here the OP-method estimators. The fully-converged value is taken to be SC PIMD with $P = 128$, and errors are plotted on a log-log scale to highlight the faster convergence of fourth-order methods.

for water, and 16-24 beads for ice. Although the GLE-thermostatted results are better than the canonically sampled PI simulations for all values of P , we observe that the convergence of GLE techniques is non-monotonic, similar to what was observed in simulations of small molecules at ultra-low temperature^[153]. It appears that the convergence of SC+GLE is not better than that obtained by PIGLET, which underscores the fact that the limiting factor for convergence of GLE schemes has more to do with zero-point energy leakage between different modes than with the asymptotic convergence of the PI section of the method. SC+GLE results are more sensitive to the coupling strength of the colored noise than in the case of PIGLET, probably due to the more complex form of the full path integral Hessian in the harmonic limit.

Although the possibility of combining high-order path integrals with correlated noise sampling might be beneficial in some specific cases – for instance when computing structural properties at ultra-low temperature – it seems that the best course of action should be to use PIGLET whenever an accuracy of a few percent is sufficient, and resort to SC PIMD with conventional thermostating whenever one wants (a) to reach the ultimate level of convergence, (b) to use sampling techniques (e.g. replica exchange) for which it is necessary to have a well-defined functional form for the phase-space density, or (c) to compute complicated estimators whose convergence is not accelerated by GLEs.

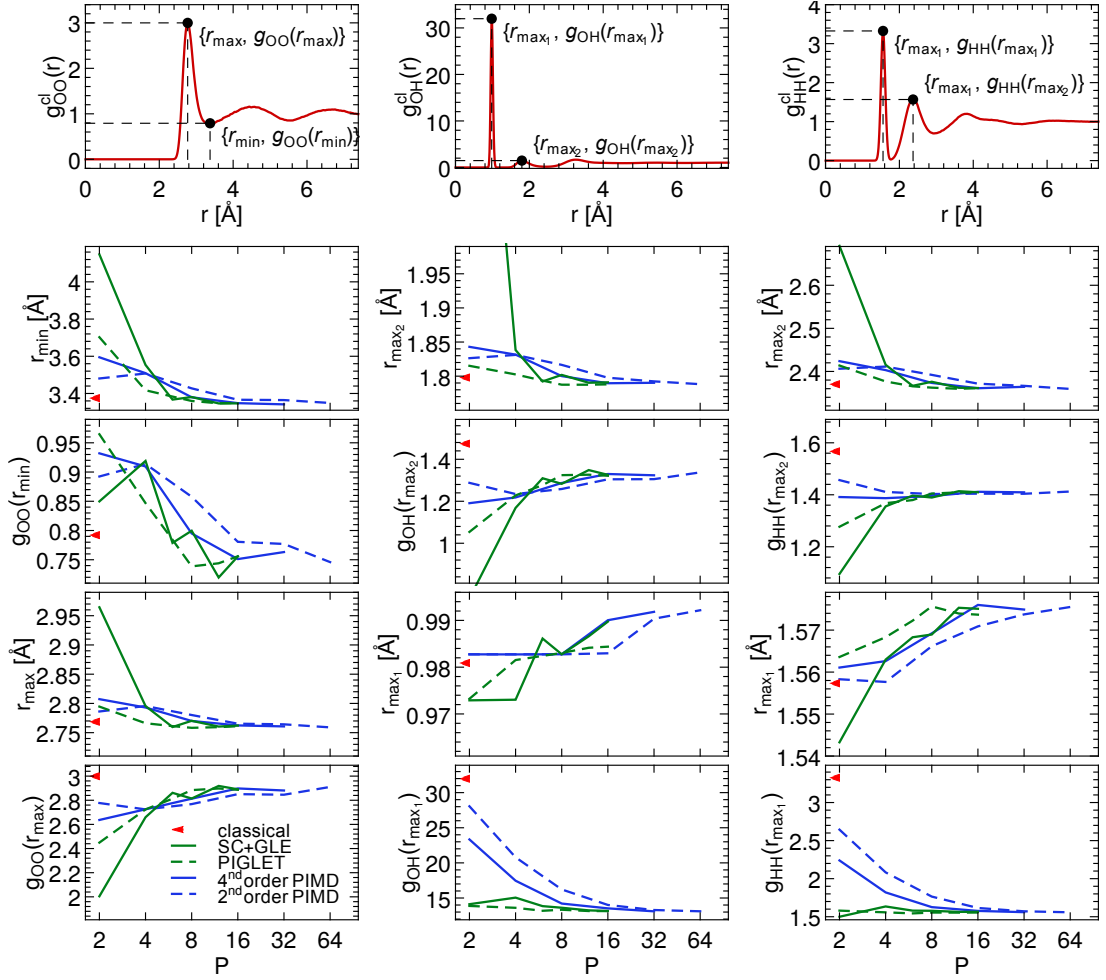


Figure 2.9 – The top panels show the radial distribution functions for O-O (left), O-H (center), and H-H (right) for the classical simulations along with some characteristic points whose convergence as a function of the number of replicas P is plotted in the four lower panels for each case. The red arrows show the corresponding values for a purely classical simulation.

Radial distribution functions

The radial pair correlation functions $g(r)$ represent the most frequently used indicators of the structure of water. Figure 2.9 shows the convergence of a few key features in the O-O, O-H and H-H correlation functions. As it has already been noted^[154], for Trotter PIMD there is an interesting non-monotonic convergence behavior of the g_{OO} distribution function, that gets less structured when going from classical to 2 and 4 beads, and then becomes more structured when it approaches convergence. Such a trend can be seen as a manifestation of the competition between quantum effects in different vibrational modes, that progressively converge as the number of replicas is increased. Overall, the convergence of the radial distribution functions with P is fully consistent with what is observed for the energy estimators. The more strongly quantized degrees of freedom – such as the O-H stretch – show the slowest convergence, and the most dramatic improvements with SC path integrals and colored-noise techniques. GLE methods give very good agreement with 6-8 beads, but if a very high accuracy is required – as it is often needed in case of radial distribution functions, for which changes in the peak shapes of a few percent can be significant – SC PIMD with 16 beads gives the best performance/cost ratio.

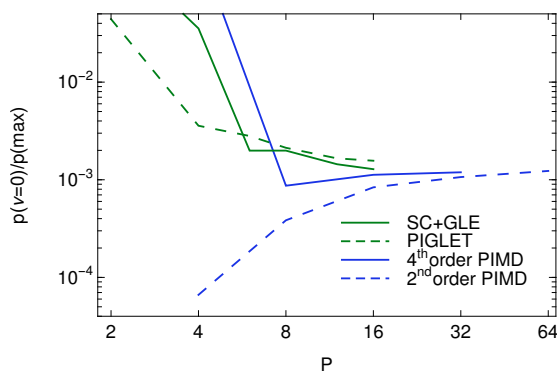


Figure 2.10 – Relative probability for observing proton delocalization over a H-bond. $p(v)$ is the probability density relative to the proton-transfer coordinate v , and the plot reports the ratio between $p(0)$ and the most likely H-bond configuration $p(\max)$, as a function of the number of replicas P . Note the convergence is slower than for the energy in Fig. 2.7.

H-bond fluctuations

One of the most remarkable effects of quantum fluctuations in room-temperature water is the occurrence of transient self-dissociation events, in which a quantum fluctuation momentarily brings a proton closer to the acceptor oxygen atom than to the oxygen it is covalently bound to^[58]. The extent of these fluctuations is a particularly challenging quantity to compute, because of the small fraction of particles that undergo such broad excursions at any given time, the strong anharmonicity of the potential in this region, and the dependence on the level of electronic structure theory^[155]. Figure 2.10 shows the probability of having a proton mid-way between the donor and acceptor oxygen relative to the probability of the most common value of the proton transfer coordinate v . The convergence is very slow for all methods, with the exception of SC PIMD - although for $P \leq 4$ fourth-order methods give dramatic over-estimation of these fluctuations. Colored-noise methods do accelerate convergence, but tend to yield too high fluctuations. For $P = 6$, PIMD would underestimate the fluctuations by a factor of 5, whereas PIGLET provides a too high value by a factor of 2. SC+GLE improves the convergence relative to PIGLET – an advantage that is however less significant when one considers the increase in computational cost. Even when predicting strongly anharmonic fluctuations, GLE techniques make it possible to reach semi-quantitative accuracy quickly, and fourth-order path integrals are useful to reach full convergence in the asymptotic regime.

Vibrational density of states

This far we have focused exclusively on static, time-independent properties. The path integral formalism is of a statistical mechanical nature, and strictly speaking no dynamical observable can be inferred. That said, several methods inspired by PIMD (such as centroid molecular dynamics, CMD^[156,115] and ring polymer molecular dynamics, RPMD^[81,157]) have been proposed to approximately estimate diffusion coefficients, vibrational spectra and other time-dependent quantities. For these benchmarks we will focus on thermostatted RPMD (TRPMD)^[90], a simple approach that can be seen as combining elements of CMD and RPMD, alleviating some of their most severe artifacts^[158], at the price however of a broadening of high-frequency peaks^[118]. The idea is just to attach a Langevin thermostat to ring-polymer modes, with a damping coefficient adjusted to be proportional to the frequency of the mode

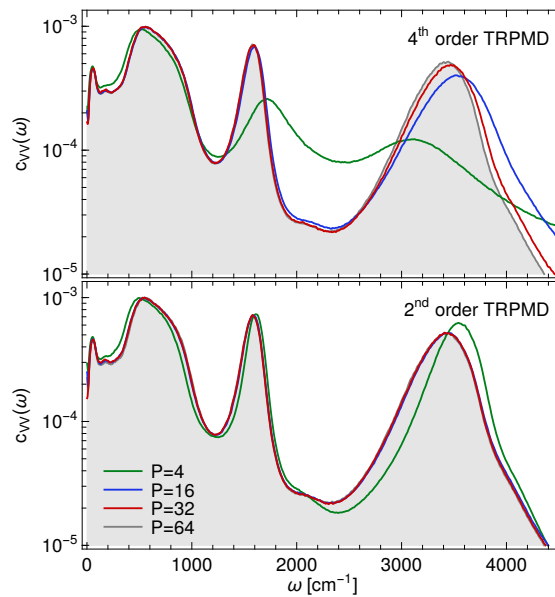


Figure 2.11 – Vibrational densities of states (Fourier-transforms of the velocity-velocity correlation functions) for TRPMD simulations of liquid water at 300 K, using a second-order Hamiltonian (lower panel) and a fourth-order Hamiltonian (upper panel). Simulations with $P = 4, 16, 32$ (green, blue, red) are compared with a fully-converged Trotter TRPMD simulation (gray, shaded).

in the free-particle limit. For $V = 0$ there is no difference between the second and fourth-order Hamiltonians, and consequently the TRPMD approach can be applied in exactly the same way to a fourth-order simulation.

Generally, one performs (T)RPMD using a number of replicas that is sufficient to converge satisfactorily the static properties to their quantum values. In Figure 2.11 we investigate the convergence of the vibrational density of states (velocity-velocity correlation spectrum) of water with increasing numbers of replicas. Interestingly, the $c_{vv}(\omega)$ converges faster than structural properties. When using $\mathcal{H}_P^{\text{tr}}$, $P = 16$ is sufficient to obtain a vibrational spectrum that is indistinguishable from the fully-converged limit. On the other hand, convergence for $\mathcal{H}_P^{\text{sc}}$ is dramatically slowed down. This is consistent with what observed in Ref. ^[95] for a harmonic potential and the closely-related case of Takahashi-Imada path integrals: the physical vibration is shifted to higher values, and the discrepancy decays slowly, as $1/P^2$. Even at $P = 32$ one can observe a significant blue shift and broadening of the OH stretch peak relative to fully converged Trotter TRPMD. Although TRPMD based on fourth-order path integrals eventually converges to the same spectrum as conventional Trotter TRPMD, it

does so very slowly, and so there is no advantage in applying fourth-order factorizations to approximate quantum dynamics. The SC scheme could however be used to accelerate the convergence of the mean field centroid force in the case of fully adiabatic centroid molecular dynamics.^[134]

3 Efficient calculation of complicated observables

3.1 Heat capacity

The heat capacity, *i.e.*, the amount of energy required to increase a material's temperature is a fundamental thermodynamic property. It is used to monitor phase transitions^[159], to calculate heat exchange during thermodynamic cycles^[160], and as a starting point for many models used to understand gas adsorption^[161]. The heat capacity is affected significantly by the quantum nature of nuclei, as it is a measure of the change in the population of the quantized energy levels of a system. For instance, the constant volume heat capacity (C_v) of an ideal crystal, described by the Debye model deviates significantly from its classical Dulong-Petit limit at low temperatures. The calculation of the quantum heat capacity for general anharmonic systems requires path integral molecular dynamics simulations^[138,162,163,92]. Since the number of replicas required to converge the heat capacity is generally much higher than that required to converge the energy^[92], path integral heat capacity calculations for

The following chapter is partially based on the journal articles:

1. V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. Van Speybroeck, and M. Ceriotti, "Modeling the Structural and Thermal Properties of Loaded Metal–Organic Frameworks. An Interplay of Quantum and Anharmonic Fluctuations," *Journal of Chemical Theory and Computation*, vol. 15, pp. 3237–3249, May 2019. VK was involved in the design of the research, in the development and implementation of the method, in running some of the calculations, in performing analysis and in the writing of the manuscript.
2. V. Kapil, A. Cuzzocrea, and M. Ceriotti, "Anisotropy of the Proton Momentum Distribution in Water," *The Journal of Physical Chemistry B*, vol. 122, pp. 6048–6054, June 2018. VK was involved in the design of the research, in the development and implementation of the method, in running the calculations, in performing analysis and in the writing of the manuscript.

most molecular systems containing hydrogen are generally considered expensive at moderate temperatures, and prohibitive at low temperature. More so, naive heat capacity estimators have poor statistical properties, and therefore require long simulations^[162]. It is therefore essential to develop methodologies for efficient computation of the heat capacity. Here, we derive low variance constant volume and constant pressure heat capacity estimators, and demonstrate their improved performance with respect to existing estimators.

3.1.1 Estimators of heat capacity using standard PIMD

The most naive way of estimating the internal energy is by averaging the primitive estimator

$$E^{\text{PR}} = \frac{3NP}{2\beta} - \frac{1}{P} \sum_{j=1}^P \sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 (\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)})^2 + \frac{1}{P} \sum_{j=1}^P V(\mathbf{q}^{(j)}, \mathbf{h}), \quad (3.1)$$

which is obtained by taking the derivative of the partition function with respect to β . This estimator, however, is inefficient as its variance depends linearly on P , meaning that length of simulation required to converge its average increases linearly with P . An efficient way of computing the internal energy, such that the variance of the estimator only depends weakly on P , is using the virial theorem, which replaces the spring term by an inner product between the positions and the forces. The resulting centroid-virial estimator^[164] has the form

$$E^{\text{CV}} = \frac{3N}{2\beta} + \frac{1}{P} \sum_{j=1}^P (\mathbf{q}^{(j)} - \bar{\mathbf{q}}) \cdot \mathbf{f}^{(j)} + \frac{1}{P} \sum_{j=1}^P V(\mathbf{q}^{(j)}, \mathbf{h}), \quad (3.2)$$

where $\bar{\mathbf{q}}$ is the centroid of the beads. Direct temperature derivative of the internal energy estimators, leads to primitive and centroid virial heat capacity estimators^[162]

$$\langle C_V^\square \rangle_P^{(2)} = k_B \beta^2 \left[\langle E^\square E^{\text{PR}} \rangle_P^{(2)} - \langle E^\square \rangle_P^{(2)} \langle E^{\text{PR}} \rangle_P^{(2)} - \left\langle \frac{\partial E^\square}{\partial \beta} \right\rangle_P^{(2)} \right] \quad (3.3)$$

where $\square = \text{TR}, \text{PR}$. Due the presence of the spring term in E^{PR} , the variance of these estimators unfortunately also depends explicitly on P , rendering their computation inefficient for realistic values of P . As in the case of internal energy, it is useful to use the virial theorem, and derive

the so called “double-centroid-virial” estimator^[162]

$$\langle C_V^{\text{DCV}} \rangle_P^{(2)} = k_B \beta^2 \left[\langle E^{\text{CV}} E^{\text{CV}} \rangle_P^{(2)} - \langle E^{\text{CV}} \rangle_P^{(2)} \langle E^{\text{CV}} \rangle_P^{(2)} + \langle E' \rangle_P^{(2)} \right] \quad (3.4)$$

where

$$E' = \frac{3N}{2\beta^2} + \frac{1}{2P\beta} \left[\frac{3}{2} \sum_{j=1}^P (\mathbf{q}^{(j)} - \bar{\mathbf{q}}) \cdot \mathbf{f}^{(j)} - \frac{1}{2} \sum_{j=1}^P (\mathbf{q}^{(j)} - \bar{\mathbf{q}}) \frac{\partial^2 V(\mathbf{q}^{(j)}, \mathbf{h})}{\partial \mathbf{q}^{(j)2}} (\mathbf{q}^{(j)} - \bar{\mathbf{q}}) \right] \quad (3.5)$$

The direct evaluation of the E' term is complicated by the presence of terms that depend explicitly on the Hessian of the potential. A practical way of estimating E' , as prescribed by Yamamoto^[165], is through the temperature derivative

$$E' = \frac{3N}{2\beta^2} - \frac{1}{P} \sum_{j=1}^P \frac{\partial^2}{\partial \beta'^2} \left[\beta' V(\bar{\mathbf{q}} + \sqrt{\frac{\beta'}{\beta}} (\mathbf{q}^{(j)} - \bar{\mathbf{q}}), \mathbf{h}) \right]_{\beta'=\beta} \quad (3.6)$$

where the second derivative is estimated using the centered finite difference formula $\frac{\partial^2 f}{\partial x^2} \approx \frac{1}{\delta x^2} [f(x + \delta x) + f(x - \delta x) - 2f(x)]$. The methodology proposed by Yamamoto therefore requires $2P$ additional force evaluation at positions $\mathbf{q} + \sqrt{\frac{\beta'}{\beta}} (\mathbf{q}^{(j)} - \bar{\mathbf{q}})$ for $\beta' = \beta \pm \delta\beta$.

While the approach proposed by Yamamoto is both elegant and practical, it is actually possible to estimate E' directly with a smaller computational effort. One can note that the second derivative in Eq. 3.5 is the Hessian projected on the vector joining the beads with the centroid. Taking inspiration from the finite difference formula for computing the projected Hessian in the SC forces^[2], we have found a simple formula $E' = \frac{3N}{2\beta^2} + \frac{1}{2P\beta} \left[\sum_{j=1}^P (\mathbf{q}^{(j)} - \bar{\mathbf{q}}) \cdot \mathbf{f}'^{(j)} \right]$, where

$$\mathbf{f}'^{(j)} = \frac{3}{2} \mathbf{f}^{(j)} + \frac{1}{2\epsilon} \left(\mathbf{f}^{(j)} \Big|_{\mathbf{q}^{(j)} + \epsilon(\mathbf{q}^{(j)} - \bar{\mathbf{q}})} - \mathbf{f}^{(j)} \right) \quad (3.7)$$

This expression gives the result as the Yamamoto estimator, assuming that the forces are calculated accurately, but requires just P additional force evaluations while displacing the replicas along the position vector that connects them with the centroid. The advantage of this formula will become more evident in the context of high-order PIMD.

Estimators of the constant pressure heat capacity C_P can be computed similarly, by starting from the β derivative of the isothermal-isobaric partition function. The estimators thus

obtained are very similar to those of C_V ; the estimators of energy are replaced by those of enthalpy $H^\square = E^\square - \mathcal{P}^{\text{ext}}\mathcal{V}$, and terms such as E' and $\frac{\partial E^\square}{\partial \beta}$ are retained.

3.1.2 Estimators of heat capacity using high-order PIMD

As discussed in Chapter 2, there exist two kinds of estimators in the context of SC PIMD. Those which are estimated from the partition function using thermodynamic (TD) identities are called thermodynamic estimators, while those which are obtained as expectation values of operators are called operator (OP) estimators. We begin by discussing the TD estimators of energy and heat capacity. Direct differentiation of the partition function with respect to β yields the TD primitive estimator of energy,

$$E^{\text{TD;PR}} = \frac{3NP}{2\beta} - \frac{1}{P} \sum_{j=1}^P \sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 \left(\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)} \right)^2 + \frac{1}{P} \sum_{j=1}^{P/2} \left[\frac{2}{3} V(\mathbf{q}^{(2j-1)}) + \frac{4}{3} V(\mathbf{q}^{(2j)}) + \frac{1}{3} \tilde{V}(\mathbf{q}^{(2j)}, \mathbf{h}) \right], \quad (3.8)$$

and a subsequent use of the virial theorem, as in the case of standard PIMD, leads to the TD centroid virial estimator of energy

$$E^{\text{TD;CV}} = \frac{3N}{2\beta} + \frac{1}{P} \sum_{j=1}^P \left(\mathbf{q}^{(j)} - \bar{\mathbf{q}} \right) \cdot \left(\mathbf{f}^{\text{sc}(j)} \right) + \sum_{j=1}^{P/2} \left[\frac{2}{3} V(\mathbf{q}^{(2j-1)}) + \frac{4}{3} V(\mathbf{q}^{(2j)}) + \frac{1}{3} \tilde{V}(\mathbf{q}^{(2j)}, \mathbf{h}) \right], \quad (3.9)$$

where $\mathbf{f}^{\text{sc}(j)}$, the weighted physical force plus the high-order term acting on the j^{th} replica, is readily evaluated using the finite difference formula^[2] during the course of molecular dynamics. The corresponding TD heat capacity estimators obtained by deriving the averages of these estimators with respect to β , and the TD centroid virial heat capacity estimators obtained by using the virial theorem are

$$\begin{aligned} \langle C_V^{\text{TD};\square} \rangle_P^{(2)} &= k_B \beta^2 \left[\langle E^{\text{TD};\square} E^{\text{TD;CV}} \rangle_P^{(2)} - \langle E^{\text{TD};\square} \rangle_P^{(2)} \langle E^{\text{TD;PR}} \rangle_P^{(2)} - \left\langle \frac{\partial E^{\text{TD};\square}}{\partial \beta} \right\rangle_P^{(2)} \right], \\ \langle C_V^{\text{TD;DCV}} \rangle_P^{(2)} &= k_B \beta^2 \left[\langle E^{\text{TD;CV}} E^{\text{TD;CV}} \rangle_P^{(2)} - \langle E^{\text{TD;CV}} \rangle_P^{(2)} \langle \text{TD}; E^{\text{CV}} \rangle_P^{(2)} + \langle E'^{\text{TD}} \rangle_P^{(2)} \right] \end{aligned} \quad (3.10)$$

where, E'^{TD} is a highly complicated expression that depends on the third derivatives of the potential energy. Fortunately, the prescription of Yamamoto^[163], allows one to express it as the temperature derivative

$$E'^{\text{TD}} = \frac{3N}{2\beta^2} - \frac{1}{P} \sum_{j=1}^P \frac{\partial^2}{\partial \beta'^2} \left[\beta' V^{\text{sc}}(\bar{\mathbf{q}} + \sqrt{\frac{\beta'}{\beta}} (\mathbf{q}^{(j)} - \bar{\mathbf{q}}), \mathbf{h}) \right]_{\beta'=\beta}, \quad (3.11)$$

where $V^{\text{sc}}(\mathbf{q}^{(j)}, \mathbf{h})$ is the scaled physical potential plus the high of term of the j^{th} replica. Practical evaluation of this term requires requires $2P$ additional force evaluations as in the case of standard PIMD.

An elegant way of approaching the computation of heat capacity within the SC scheme, and subsequently reducing the associated computational cost, is by using OP estimators of energy. These estimators take the same form as in the case of standard PIMD, but are only evaluated on odd replicas. For instance the OP centroid virial estimator of energy is

$$E^{\text{OP;CV}} = \frac{3N}{2\beta} + \frac{2}{P} \sum_{j=1}^{P/2} (\mathbf{q}^{(2j-1)} - \bar{\mathbf{q}}) \cdot \mathbf{f}^{(2j-1)} + \frac{2}{P} \sum_{j=1}^{P/2} V(\mathbf{q}^{(2j-1)}, \mathbf{h}). \quad (3.12)$$

Not only are OP estimators simpler to express, but they also reduce the order of the derivative of the potential by one, required for the computation of the energy and the heat capacity (*vide infra*). We have shown that by deriving the average of the OP centroid virial estimator of energy, and using the virial theorem, it is possible to obtain the double centroid virial heat capacity OP estimator

$$\langle C_V^{\text{OP;DCV}} \rangle_P^{(4)} = k_B \beta^2 \left[\langle E^{\text{OP;CV}} E^{\text{OP;CV}} \rangle_P^{(4)} - \langle E^{\text{OP;CV}} \rangle_P^{(4)} \langle E^{\text{OP;CV}} \rangle_P^{(4)} + \langle E'^{\text{OP}} \rangle_P^{(4)} \right] \quad (3.13)$$

where, $E^{\text{OP-CV}}$ has the same expression as in the case of standard PIMD, except that it is evaluated on even replicas. We show that it can also be computed as the finite difference $E' = \frac{3N}{2\beta^2} + \frac{1}{P\beta} \left[\sum_{j=1}^{P/2} (\mathbf{q}^{(2j-1)} - \bar{\mathbf{q}}) \cdot \mathbf{f}^{(2j-1)} \right]$, where

$$\mathbf{f}^{(2j-1)} = \frac{3}{2} \mathbf{f}^{(2j-1)} + \frac{1}{2\epsilon} \left(\mathbf{f}^{(2j-1)} \Big|_{\mathbf{q}^{(2j-1)} + \epsilon(\mathbf{q}^{(2j-1)} - \bar{\mathbf{q}})} - \mathbf{f}^{(2j-1)} \right) \quad (3.14)$$

can be computed by displacing the odd replicas along the position vector that connects them

with the centroid. This estimator only required $P/2$ additional force evaluations, i.e. it has one-fourth the cost of the Yamamoto's estimator of E'^{TD} .

3.1.3 Benchmarks on liquid water

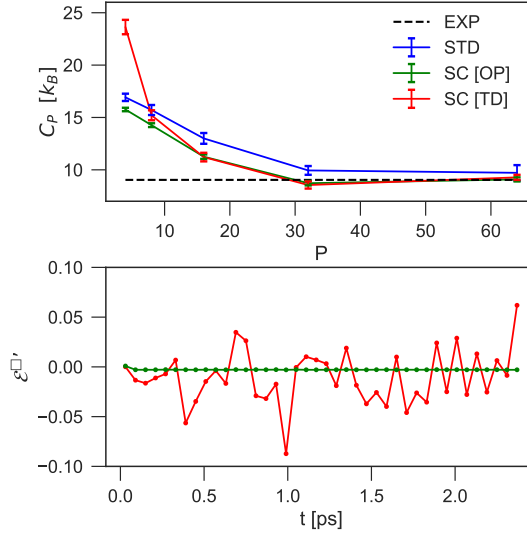


Figure 3.1 – The top panel shows the convergence of the isobaric heat capacity of liquid water at 300 K, modeled by the q-TIP4P/f potential, as a function of the number of replicas. The blue curve was obtained from standard PIMD using the Yamamoto estimator while the red and green curves were obtained from SC PIMD using the Yamamoto (TD) and OP estimator respectively. The dashed black line represents the experimental result. Error bars indicate statistical uncertainty. The bottom panel shows that instantaneous values of the the computationally expensive E'^{\square} term when computed with the OP and the Yamamoto (TD) estimator.

We tested the SC estimators on the well-known isobaric heat capacity of liquid water which is $1 \text{ cal mol}^{-1} \text{ K}^{-1}$ or $9 k_B$ per molecule at 300 K. The water molecules are modeled with the q-TIP4P/f force field^[166]. As shown in Figure 3.1, both estimators converge to the same value and are in excellent agreement with the experiments. More importantly, however, the variance of the computationally expensive E'^{\square} term with the OP method is almost two orders of magnitude smaller than with the Yamamoto estimator. The OP estimator also only requires one fourth of the number of force evaluations than its Yamamoto counterpart and should therefore always be preferred.

3.2 Particle momentum distribution

The momentum distribution of particles $n(\mathbf{p})$ can be measured directly using Deep Inelastic Neutron Scattering (DINS) experiments^[167,25], and provide sensitive information on the chemical environment or the local structure of target particles. For instance, the subtle features obtained from the momentum distributions can be used to understand phenomena such as the softening of OH bonds in ice^[168]. Moreover, many useful quantities such as the quantum kinetic energies^[25], as well as the kinetic energy tensor^[169] can be estimated while measuring particle momentum distributions. DINS experiments provide highly accurate estimates of the quantum kinetic energies, therefore they can be used to study the accuracy of *ab initio* potential energy surfaces obtained from electronic structure methods.

Since the classical PMD of all particles is a Maxwell-Boltzmann distribution, any deviation of the PMD from the classical limit is a direct probe of the quantum nature of nuclei, and therefore can be simulated *only* if the nuclei are considered quantum particles. Here, we discuss how the particle momentum distribution can be estimated within the imaginary time path integral method by opening the paths of target atoms. We present a theoretical framework that greatly simplifies the implementation of the “open” path integral method. We also discuss the poor statistical performance of the standard momentum distribution estimator, and present a novel virial like estimator that reduces the computational effort by over an order of magnitude.

3.2.1 Open paths for calculating the particle momentum distribution

We take as starting point a well-established expression for the canonical particle momentum distribution of the i' th atom in a system composed of N atoms at inverse temperature β , in terms of the Fourier transform of the off-diagonal components of the one-particle density matrix: $n(\mathbf{p}) = (2\pi\hbar)^{-3} \int d\Delta e^{-i\hbar^{-1}\mathbf{p}\cdot\Delta} \rho_{i'}(\Delta)$ with

$$\rho_{i'}(\Delta) = Z^{-1} \int d\{\hat{\mathbf{q}}\} \langle \hat{\mathbf{q}}_1, \dots, \hat{\mathbf{q}}_{i'}, \dots, \hat{\mathbf{q}}_N | \exp(-\beta \hat{H}) | \hat{\mathbf{q}}_1, \dots, \hat{\mathbf{q}}_{i'} - \Delta, \dots, \hat{\mathbf{q}}_N \rangle. \quad (3.15)$$

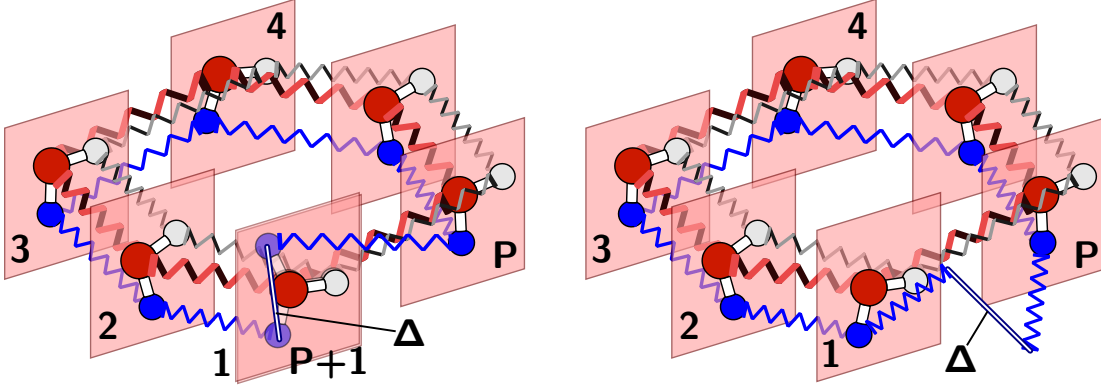


Figure 3.2 – The discretization of the imaginary time path integral using a uniform grid starting from $\tau = 0$ to $\tau = \beta\hbar$ (left) gives the standard open PIMD Hamiltonian while the alternate discretization (right) makes it possible to integrate out the ends, yielding a Hamiltonian which describes an open paths with P replicas.

$n(\mathbf{p})$ can be computed performing a second-order Trotter splitting of the Boltzmann operator into P factors, $e^{-\beta\hat{H}} \approx \left[e^{-\beta_P \frac{\hat{V}}{2}} e^{-\beta_P \hat{T}} e^{-\beta_P \frac{\hat{V}}{2}} \right]^P$ and inserting complete sets of position states between each pair of factors. This yields

$$n(\mathbf{p}) = (2\pi\hbar)^{-3} \int d\Delta e^{-i\hbar^{-1}\mathbf{p}\cdot\Delta} N(\Delta) \quad (3.16)$$

where, $N(\Delta) = \langle \delta(\Delta - \mathbf{q}_{i'}^{(1)} + \mathbf{q}_{i'}^{(P+1)}) \rangle_p^{(2)}$ is the end-to-end distribution of the i' th atom in the isomorphic open polymer Hamiltonian

$$H_P^{(o)}(\{\mathbf{p}\}, \{\mathbf{q}\}) = \sum_{j=1}^P \left[\sum_{i=1 \neq i'}^N \frac{1}{2} m_i^{-1} |\mathbf{p}_i^{(j)}|^2 + \sum_{i=1 \neq i'}^N \frac{1}{2} m_i \omega_P^2 (\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)})^2 \right] + \sum_{j=1}^P \left[\frac{1}{2} V(\mathbf{q}^{(j)}, \mathbf{h}) + \frac{1}{2} V(\mathbf{q}^{(j+1)}, \mathbf{h}) \right], \quad (3.17)$$

with $\mathbf{q}_i^{(j)} \equiv \mathbf{q}_i^{(j+P)}$ implied for $i \neq i'$. Note that the single particle density matrix of the i' th atoms is equivalent to the end to end distribution of the open path associated with the i' th atom of the isomorphic Hamiltonian. Since the end points of all but one ring polymer coincide, one has to introduce an additional replica just to evaluate the potential for two configurations that differ only by the positions of the i -th particle, and that are each weighted by a factor 1/2 (Figure 3.2). Besides the slight computational overhead, the presence of the extra replica for the target species and the scaling of the potential makes the implementation of this Hamiltonian

in an existing PIMD code somewhat cumbersome.

3.2.2 A convenient algorithm for implementing open PIMD

To avoid this inconvenience, we use in the implementation of PIMD for open and closed ring polymers – also popularly known as open PIMD – in the i-PI code^[141] an alternative second-order splitting $e^{-\beta\hat{H}} \approx \left[e^{-\beta_P \frac{\hat{T}}{2}} e^{-\beta_P \frac{\hat{V}}{2}} \right] \left[e^{-\beta_P \frac{\hat{V}}{2}} e^{-\beta_P \hat{T}} e^{-\beta_P \frac{\hat{V}}{2}} \right]^{(P-1)} \left[e^{-\beta_P \frac{\hat{V}}{2}} e^{-\beta_P \frac{\hat{T}}{2}} \right]$. We show in Ref.^[4], that it is possible to compute the end-to-end distribution $N(\Delta)$ by sampling the Hamiltonian of a truncated ring polymer

$$H_P^{(o)}(\{\mathbf{p}\}, \{\mathbf{q}\}) = \sum_{j=1}^P \left[\sum_{i=1}^N \frac{1}{2} m_i^{-1} [\mathbf{p}_i^{(j)}]^2 + \sum_{i=1}^N \frac{1}{2} m_i \omega_P^2 [\mathbf{q}_i^{(j)} - \mathbf{q}_i^{(j+1)}]^2 + \right. \quad (3.18)$$

$$\left. V(\mathbf{q}_i^{(j)}, \mathbf{h}) \right] - \frac{1}{2} m_i' \omega_P^2 [\mathbf{q}_{i'}^{(1)} - \mathbf{q}_{i'}^{(P)}]^2 \quad (3.19)$$

in which all the replicas experience the full physical potential and both open and closed replicas are represented by P beads. Implementing Molecular Dynamics or Monte Carlo sampling for this Hamiltonian in an existing closed-path code is trivial since it just requires the modification of the normal mode (or staging) transformation for the target species^[117].

The attentive reader will have noticed that the alternative Trotter splitting contains two additional free-particle propagators, that describe the fluctuations of the end-points of the path around the first and the P -th bead. These fluctuations can be evaluated analytically, and correspond to a Gaussian convolution of the 1-st-to- P -th bead distribution. In practice, in order to estimate $N(\Delta)$ one simply needs to compute the histogram of Δ using a kernel function that depends parametrically on β , m and P :

$$N(\Delta) = \langle G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \rangle_P^{(o)}, \quad (3.20)$$

with $G^{3D}(\mathbf{x}, \mathbf{x}') = (\sqrt{2\pi}\sigma_P)^{-3} e^{-\frac{(\mathbf{x}-\mathbf{x}')^2}{2\sigma_P^2}}$, and $\sigma_P = \sqrt{m_i^{-1} \beta_P \hbar^2}$.

When computing $n(\mathbf{p})$ in an isotropic system such as liquid water, one is generally interested in the spherical average of the end-to-end distribution, $N(\Delta)$, because the spherically-averaged

momentum distribution can be obtained from it as

$$n(p) = \int d\Delta \, 4\pi\Delta^2 N(\Delta) \frac{\text{Sin}(p\Delta)}{p\Delta}. \quad (3.21)$$

$N(\Delta)$ can be obtained from $N(\mathbf{\Delta})$ by simply writing $\mathbf{\Delta}$, in spherical coordinates and integrating over the polar and azimuthal angles

$$N(\Delta) = \int_0^{2\pi} d\phi \int_{-1}^1 d(\text{Cos}(\theta)) \, \langle G^{3D}(\mathbf{\Delta} - (\mathbf{q}^1 - \mathbf{q}^P)) \rangle_P^{(o)} = \langle G^r(\Delta, |\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}|) \rangle_{H_P} \quad (3.22)$$

where $G^r(x, x') = (2\pi\sigma_P^2)^{-1/2} \frac{1}{xx'} \left[e^{-\frac{1}{2\sigma_P^2}(x-x')^2} - e^{-\frac{1}{2\sigma_P^2}(x+x')^2} \right]$ is the appropriate Kernel used to bin the end-to-end distance.

3.2.3 An efficient estimator for the momentum distribution

The major problem associated with open path simulations is that the convergence of the end-to-end distribution is very slow with respect to simulation time, especially for $\Delta \sim 0$. This translates into a large error in the momentum distribution, and could even lead to additional spurious peaks arising from the noise in $N(\Delta)$. The solution to this problem proposed by Lin *et. al*, is a displaced path estimator^[101] which essentially computes the thermodynamic force associated with the end-to-end distance. This estimator can be computed using free energy perturbation (FEP) by sampling from a closed ring polymer, but has a computational overhead arising from the evaluation of the potential energy of the replicas along a open polymer (obtained by displacing the closed ring polymer). Unfortunately, the FEP method is applicable only to small systems, as its efficiency degrades exponentially with increasing system size. In the strong quantum regime, or in the limit of large system size, it is more efficient to compute the mean force using the thermodynamic integration method that bears the computational cost of multiple simulations with constraints/restraints at different fixed values of Δ . Other methods that can be used to yield efficient estimates of the particle momentum distribution are enhanced sampling techniques, such as metadynamics^[170] or variationally enhanced sampling^[171], that accelerate convergence of the free energy surface of the end-to-end vector.

We have proposed an elegant solution to this problem, by deriving a virial-like estimator that

computes the derivative of $N(\Delta)$. We begin by noting the following properties of the derivative of the 3D kernel:

$$\nabla_{\Delta} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) = -\nabla_{\mathbf{q}_i^{(1)}} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) = \nabla_{\mathbf{q}_i^{(P)}} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}), \quad (3.23)$$

$$\nabla_{\mathbf{q}_i^{(j \neq 1, P)}} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) = 0 \quad (3.24)$$

which allow us to express the gradient of the end-to-end distribution as a linearly scaled sum of derivative $\nabla_{\Delta} N(\Delta) = \langle \sum_{j=1}^P \lambda_j \nabla_{\mathbf{q}_i^{(j)}} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \rangle_P^{(o)}$, where λ_j is an arbitrary sequence with fixed boundary conditions $\lambda_1 = -\frac{1}{2}$ and $\lambda_P = \frac{1}{2}$. Of the P gradients, only those w.r.t the end beads contribute to the average; The rest of them only contribute to the fluctuations of the estimator, which is advantageous since an appropriate choice of λ_k should allow us to minimize the variance of the estimator. However, before we proceed to that, we cast the "scaled gradient" in a virial like form, by performing an integral by parts

$$\nabla_{\Delta} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) = \sum_{j=1}^P \langle \lambda_j \nabla_{\mathbf{q}_i^{(j)}} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \rangle_P^{(o)} \quad (3.25)$$

$$= \sum_{j=1}^P \langle \lambda_j G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \beta_P \nabla_{\mathbf{q}_i^{(j)}} H_P \rangle_P^{(o)} \quad (3.26)$$

While it would be possible to optimize the λ_j based on the computed bead-bead correlations from a reference calculation, we find that results close to the optimum can be obtained taking the λ_j is in such a way that the contribution from the spring force coming from intermediate beads vanishes. This condition can be met by imposing that λ_j is an arithmetic progression with a common difference of $\frac{1}{P-1}$ for which the fluctuations in the spring term becomes independent of P and the estimator takes the simplified form

$$\nabla_{\Delta} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) = -\langle \mathbf{g} G^{3D}(\Delta, \mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \rangle_P^{(o)}, \quad (3.27)$$

where $\mathbf{g} = \mathbf{g} = \frac{\sigma_P^{-2}}{1-P^{-1}} \left[(\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}) \right] + \sum_{j=1}^P \beta_P \lambda_j \mathbf{f}_i^{(j)}$ is the scaled gradient. We have also derived the gradient of the radial distribution function of Δ

$$\frac{dN(\Delta)}{d\Delta} = -\lim_{P \rightarrow \infty} \langle \mathbf{g} \cdot \left[\frac{\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}}{|\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}|} \right] G^{\text{dr}}(\Delta, |\mathbf{q}_i^{(1)} - \mathbf{q}_i^{(P)}|) \rangle_P^{(o)} \quad (3.28)$$

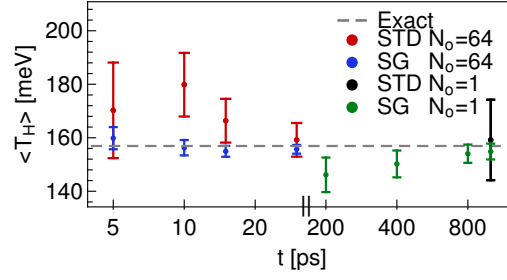


Figure 3.3 – The convergence of the kinetic energy of a proton in supercooled q-TIP4P/f water at 271K from open PIMD simulations using 64 beads calculated from the momentum distribution using the scaled gradient (SG) estimator (green). As a reference for comparison the value calculated from the standard estimator (black) with 1ns of trajectory and the exact result (dashed) from closed PIMD are also plotted. The red and the blue points are obtained using the standard and the scaled gradient estimators from simulations where one proton per molecule is represented by an open path.

where $G^{\text{dr}}(x, x') = (2\pi\sigma_p^2)^{-1/2} (x'x)^{-2} \left[e^{-\frac{(x'+x)^2}{2\sigma_p^2}} (xx' + \sigma_p^2) + e^{-\frac{m_i(x'-x)^2}{2\sigma_p^2}} (xx' - \sigma_p^2) \right]$ is the appropriate kernel.

3.2.4 Benchmarks on different phases of water

To test the convergence of the regular and the scaled gradient estimators we ran simulations of 64 molecules of q-TIP4P/f water at 271K using 64 imaginary time slices. The distribution of momentum of a proton in water can essentially be described as an anisotropic Gaussian^[172,132,173], so we can use as a convergence check the comparison between the quantum kinetic energy computed from a closed-path simulation and that obtained from the second moment of $n(p)$. As shown in figure 3.3, the scaled gradient estimator (which can be obtained from an open path simulation at virtually no computational overhead) shows an error which is approximately 5 times smaller than that associated with the standard estimator. For simulations with a single open path, 1ns-long simulations are needed to approach a statistical error sufficient to discriminate between different water molecules, whereas more than 20 would be needed without using the derivative estimator. Fig. 3.3 also demonstrates that the approximation introduced in Ref.^[69] – that is, opening one path per water molecule – leads to negligible systematic error, and once combined with the derivative estimator allows reaching a statistical error of about 2 meV with trajectories that are only 10 ps long.

4 Accelerated methods for vibrational spectra

Vibrational spectroscopic techniques – from conventional infrared (IR) and Raman to advanced femtosecond pump-probe^[174,175], sum-frequency generation, second harmonic scattering^[176,177], and multi-dimensional vibrational spectroscopy^[178] – are a cornerstone of chemistry^[174]. These techniques have a multitude of applications such as the characterization of functional groups in chemical systems^[179], the determination of the atomistic mechanisms of phase transitions through insight into chemical environments^[180], and the identification of unique structural fingerprints of molecular crystals^[181]. The use of atomistic simulations for the computation of spectroscopic properties facilitates the interpretation of these experiments and provides support to the characterization of novel materials.

Even neglecting effects that go beyond the Born-Oppenheimer (BO) decoupling of electronic and nuclear degrees of freedom, accurate calculations of the vibrational spectra of materials require an explicit treatment of the quantum dynamics of the nuclear degrees of freedom^[66]

The following chapter is partially based on the journal articles:

1. M. Rossi, V. Kapil, and M. Ceriotti, “Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation,” *The Journal of Chemical Physics*, vol. 148, p. 102301, July 2017 VK was involved in the implementation of the method, in running calculations, in performing analysis and in the writing of the manuscript.
2. V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, “Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats,” *The Journal of Chemical Physics*, vol. 152, p. 124104, Mar. 2020 VK was involved in the design of the research, in the development and implementation of the method, in running calculations, in performing analysis and in the writing of the manuscript.

on the electronic ground state potential energy surface. Quantum dynamics is in principle exactly obtained from the solution of the time dependent Schrödinger equation for the nuclei, but this is only practical for systems containing a handful of degrees of freedom^[182,183]. Condensed phase systems can be studied^[184] either by an exact treatment of the quantum dynamics of a subset of the nuclear degrees of freedom^[185], or through classical dynamics on the quantum free energy surface of the nuclei^[79,104]. Among the methods in the latter class, several of the most popular ones are based on the imaginary time path integral framework – such as (thermostatted) ring polymer molecular dynamics^[88,89] ((T)RPMD), centroid molecular dynamics^[186,87] (CMD), path integral Liouville dynamics^[105] and the recently developed quasi-centroid molecular dynamics^[107] (QCMD). These methods ignore real time coherence but include effects arising from equilibrium quantum fluctuations and have been validated on several model systems and small molecules for which exact or highly accurate results are available^[79,89,85,89,87]. While these methods show great promise for accurate determination of spectroscopic properties^[187,188,189], their cost remains high when combined with a potential energy surface computed by *ab initio* electronic structure methods. Methods such as ring polymer contraction^[91] and multiple time stepping^[40] as well as their combination^[1], as discussed in Chapter 2, can be used to reduce the cost of these simulations, however, they rely on range separation of the potential energy landscape which is not universally given.

Among the many methods that have been introduced in the past decade to accelerate the convergence of path integral calculations^[190], those that combine (path integral) molecular dynamics with a generalized Langevin equation^[96,71,191] (GLE), such as the quantum thermostat (QT) and PIGLET, can be applied transparently to empirical, machine learning or first principles simulations. The PIGLET thermostat has been used to evaluate all sorts of thermodynamic properties, including structural observables^[192], free energies^[193], momentum distributions^[71], and quantum kinetic energies^[59] with a reduction in computational effort varying between a factor of 5 at ambient conditions to a factor of 100 at cryogenic temperatures^[194]. The aggressive thermostating used to impose quantum fluctuations, however, significantly disrupts the dynamics of the system, and common wisdom is that the calculation of dynamical properties using PIGLET is impossible.

Here we present a simple post-processing strategy that makes it possible to reconstruct dynamical properties from classical and quantum GLE thermostatted simulations. We apply this method to study trajectories generated using PIGLET, leading to a dramatic reduction in the cost of including quantum effects in spectroscopic quantities. We show that when applied to model systems the accuracy of the scheme is on par with that of conventional path integral dynamical schemes, aside from a small residual broadening of the spectra. We then demonstrate the usefulness of our approach by computing the IR and Raman spectra of solid and liquid water, using state-of-the-art machine-learning interatomic potentials, dipole moment, and polarizability surfaces.

4.1 A Generalized Langevin Equation Thermostat

The generalized Langevin equation for a particle with unit mass in one dimension, subject to a potential $V(q)$, is given by the non-Markovian process

$$\begin{aligned}\dot{q} &= p \\ \dot{p} &= -V'(q) - \int_{-\infty}^t K(t-s)p(s)ds + \zeta(t)\end{aligned}\tag{4.1}$$

where $K(t)$, is the memory kernel that describes dissipation, and $\zeta(t)$ is a Gaussian random process with a time correlation function $H(t) = \langle \zeta(t)\zeta(0) \rangle$. Throughout this chapter, we consider unit mass in all equations. The numerical integration of this equation is computationally challenging since it requires the knowledge of the entire history of the particle's trajectory. However, exploiting the equivalence between the non-Markovian dynamics of Eq. 4.1 and Markovian dynamics in an extended space, n auxiliary degrees of freedom \mathbf{s} can be coupled linearly to the physical momentum, which results in the Markovian Langevin equation

$$\begin{aligned}\dot{q} &= p \\ \begin{pmatrix} \dot{p} \\ \dot{\mathbf{s}} \end{pmatrix} &= \begin{pmatrix} -V'(q) \\ \mathbf{0} \end{pmatrix} - \begin{pmatrix} a_{pp} & \mathbf{a}_p^T \\ \bar{\mathbf{a}}_p & \mathbf{A} \end{pmatrix} \begin{pmatrix} p \\ \mathbf{s} \end{pmatrix} + \begin{pmatrix} b_{pp} & \mathbf{b}_p^T \\ \bar{\mathbf{b}}_p & \mathbf{B} \end{pmatrix} \begin{pmatrix} \xi \\ \boldsymbol{\xi} \end{pmatrix}.\end{aligned}\tag{4.2}$$

Here $\boldsymbol{\xi}$ is a $n + 1$ dimensional vector of uncorrelated Gaussian numbers. In order to label the portions of the matrices that describe the coupling between the different components of the extended state vector $\mathbf{x} \equiv (q, p, \mathbf{s})^T$, we use the following notation:

$$\begin{array}{c}
 \begin{array}{ccc}
 & q & p & \mathbf{s} \\
 q & m_{qq} & m_{qp} & \mathbf{m}_q^T \\
 p & \bar{m}_{qp} & m_{pp} & \mathbf{m}_p^T \\
 \mathbf{s} & \bar{\mathbf{m}}_q & \bar{\mathbf{m}}_p & \mathbf{M}
 \end{array}
 \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} \mathbf{M}_p \left. \begin{array}{c} \\ \\ \\ \end{array} \right\} \mathbf{M}_{qp}
 \end{array} \quad (4.3)$$

Upon integrating out the auxiliary degrees of freedom, equation 4.1 is recovered with

$$\begin{aligned}
 K(t) &= 2a_{pp}\delta(t) - \mathbf{a}_p^T e^{-|t|\mathbf{A}} \bar{\mathbf{a}}_p \\
 H(t) &= d_{pp}\delta(t) - \mathbf{a}_p^T e^{-|t|\mathbf{A}} [\mathbf{Z}\mathbf{a}_p - \mathbf{d}_p]
 \end{aligned} \quad (4.4)$$

where $\mathbf{Z} = \int_0^\infty e^{-\mathbf{A}t} \mathbf{D} e^{-\mathbf{A}^T t} dt$ and $\mathbf{D}_p = \mathbf{B}_p \mathbf{B}_p^T$. This implies that by tuning the elements of the matrices \mathbf{A}_p and \mathbf{B}_p , a Generalized Langevin equation with the desired friction kernel and noise correlation can be approximated within a Markovian framework. Note that although we focused on a one-dimensional case to simplify the notation, it is also possible to apply Eqn. (4.2) to each Cartesian coordinate of an atomistic system. Since the overall dynamics is invariant to a unitary transformation of the coordinates, the response of the system would be the same as if the GLEs had been applied in e.g. the normal modes coordinates.

4.1.1 Controlling Classical Dynamics

Let us consider a particle subject to a harmonic potential $V(q) = \frac{1}{2}\omega_0^2$, and coupled to a GLE. The time evolution of its state vector $\mathbf{x} = (q, p, \mathbf{s})^T$ can be expressed as:

$$\begin{pmatrix} \dot{q} \\ \dot{p} \\ \dot{\mathbf{s}} \end{pmatrix} = - \begin{pmatrix} 0 & -1 & \mathbf{0} \\ \omega_0^2 & a_{pp} & \mathbf{a}_p^T \\ \mathbf{0} & \bar{\mathbf{a}}_p & \mathbf{A} \end{pmatrix} \begin{pmatrix} q \\ p \\ \mathbf{s} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \mathbf{0} \\ 0 & & \mathbf{B}_p \\ \mathbf{0} & & \end{pmatrix} \begin{pmatrix} 0 \\ \boldsymbol{\xi} \end{pmatrix}. \quad (4.5)$$

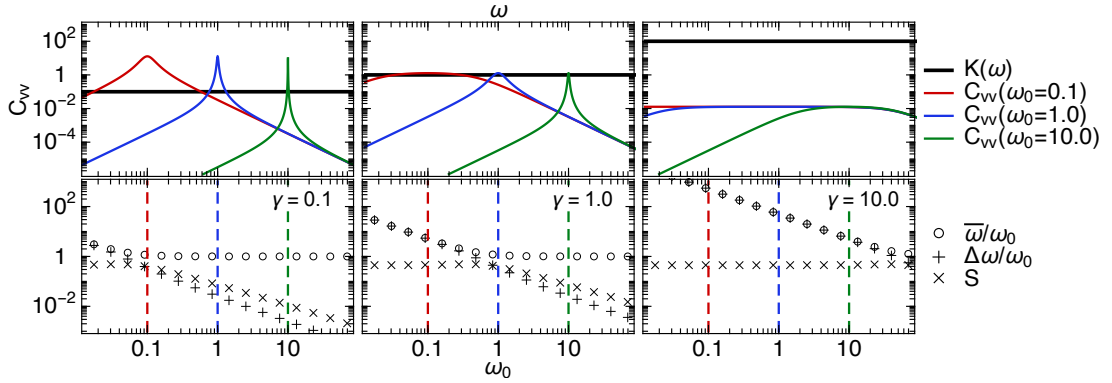


Figure 4.1 – Various regimes of the white noise thermostat acting on the harmonic oscillator. The left, central and right panels, respectively representing the under-damped ($\gamma = 0.1$), optimally-damped ($\gamma = 1$) and over-damped limits ($\gamma = 10$), show the velocity auto-correlation functions (top) and GLE metrics (bottom) for various values of the physical frequency. We choose three values for the physical mode ω_0 , labeled in the picture and shown with different colors. The GLE metrics as defined by equations 4.11, 4.12 and 4.13 are represented by circular, plus shaped and cross shaped markers respectively.

Since the force is linear in q , equation 4.5 takes the form of an Ornstein-Uhlenbeck process, that can be written concisely as

$$\dot{\mathbf{x}} = -\mathbf{A}_{qp}\mathbf{x} + \mathbf{B}_{qp}\xi. \quad (4.6)$$

Since its finite-time propagator is known analytically^[195], it is possible to compute any time correlation function in terms of the drift and diffusion matrices \mathbf{A}_p and \mathbf{B}_p . For instance, the vibrational density of states can be computed exactly by taking the Fourier transform of the velocity-velocity correlation function, and reads:

$$C_{pp}(\omega, \omega_0) = \frac{1}{[\mathbf{C}_{qp}(\omega_0)]_{pp}} \left[\frac{\mathbf{A}_{qp}(\omega_0)}{\mathbf{A}_{qp}^2(\omega_0) + \omega^2} \mathbf{C}_{qp}(\omega_0) \right]_{pp}, \quad (4.7)$$

where the stationary covariance matrix can be obtained by solving the Riccati equation $\mathbf{A}_{qp}\mathbf{C}_{qp} + \mathbf{C}_{qp}\mathbf{A}_{qp}^T = \mathbf{B}_{qp}\mathbf{B}_{qp}^T$.

It is useful to perform a spectral decomposition of Eq. 4.7 in order to gain more insight into the spectrum of a GLE-thermostatted oscillator. It is straightforward to show that by writing $\mathbf{A}_{qp}(\omega_0) = \mathbf{O} \text{diag}(\Omega) \mathbf{O}^{-1}$ where \mathbf{O} is the matrix of eigenvectors and Ω a vector containing the corresponding eigenvalues, the expression for the velocity-velocity correlation function can

be written as

$$C_{pp}(\omega, \omega_0) = \sum_{jr} O_{pj} \frac{\Omega_j(\omega_0)}{\Omega_j^2(\omega_0) + \omega^2} O_{jr}^{-1} \frac{[\mathbf{C}_{qp}(\omega_0)]_{rp}}{[\mathbf{C}_{qp}(\omega_0)]_{pp}}. \quad (4.8)$$

For example, in the case of white-noise Langevin with friction γ , the correlation function reads

$$C_{pp}^{(\gamma)}(\omega, \omega_0) = \frac{2\gamma\omega^2}{\pi(\gamma^2\omega^2 + \omega^4 - 2\omega_0^2\omega^2 + \omega_0^4)}. \quad (4.9)$$

The spectrum in Eq. 4.8 corresponds to a sum of Lorentzian functions, with the peaks positions and lineshapes determined by the poles at $\omega = \pm i\Omega_j$. Motivated by this spectral decomposition, we define several quantities that give a concise description of the shape of the spectrum. After having introduced the integral function of the spectrum

$$\begin{aligned} W(\omega_a, \omega_b) &= \frac{2}{\pi} \int_{\omega_a}^{\omega_b} C_{pp}(\omega, \omega_0) d\omega = \\ &= \left\{ \left[\tan^{-1} \left(\frac{\omega}{\mathbf{A}_{qp}(\omega_0)} \right) \right]_{\omega_b}^{\omega_a} \frac{2\mathbf{C}_{qp}(\omega_0)}{\pi[\mathbf{C}_{qp}(\omega_0)]_{pp}} \right\}_{pp}, \end{aligned} \quad (4.10)$$

which can be computed easily based on the same eigendecomposition of \mathbf{A}_{qp} , we define the median

$$\bar{\omega}(\omega_0) \rightarrow W(0, \bar{\omega}) = 0.5, \quad (4.11)$$

that characterizes the position of the peak, and the interquartile distance

$$\begin{aligned} \Delta\omega(\omega_0) &= \frac{1}{2}(\omega_{0.75} - \omega_{0.25}) \\ &\rightarrow W(0, \omega_{0.25}) = 0.25 \\ &\quad W(0, \omega_{0.75}) = 0.75 \end{aligned} \quad (4.12)$$

that characterizes its width. Together, these two indicators are sufficient to determine fully a

Lorentzian lineshape

$$L(\omega, \omega_0) = \frac{1}{\pi} \frac{\Delta\omega(\omega_0)}{(\omega - \bar{\omega}(\omega_0))^2 + [\Delta\omega(\omega_0)]^2}. \quad (4.13)$$

In order to quantify the presence of multiple poles or other sources of asymmetry in the lineshape that are not captured by $\bar{\omega}$ and $\Delta\omega$, we introduce a “non-Lorentzian-shape” factor S ,

$$S(\omega_0) = \left| \int_0^\infty [C_{pp}(\omega, \omega_0) - L(\omega, \omega_0)]^2 d\omega \right|^{0.5}. \quad (4.14)$$

According to the definitions above, a perfect δ -like Lorentzian spectrum would have $\bar{\omega}/\omega_0 = 1$, $\Delta\omega/\omega_0 = 0$, and $S = 0$. In order to exemplify how these measures behave in the case of a simple white noise thermostat attached to the harmonic oscillator, we show in Fig. 4.1 how the velocity-velocity spectrum of oscillators of different frequency ω_0 changes with different regimes of white noise, and how the measures defined in Eqs. 4.11 to 4.14 relate to the magnitude of the perturbation induced in a δ -like spectrum shape.

Analyzing Fig. 4.1, we can see that, as expected, the regime that introduces the least disturbance to the VDOS is the underdamped regime (the limit where $\mathbf{A}_p = 0$ is microcanonical dynamics) – and that for a given γ the modes with lower frequency suffer the most pronounced relative disturbance. Focusing on the underdamped case, the measures $\bar{\omega}/\omega$ and $\Delta\omega$ predict the shift and broadening of the peaks at low frequencies, as well as the lack of disturbance at high frequencies. Going to the optimally damped and the overdamped case, the disturbances to the spectra get more pronounced through the whole range of frequencies, and it is easy to follow how the different indicators we introduced quantify this change. The S measure is always relatively small, indicating that a simple white-noise thermostat does not affect significantly the Lorentzian character of the peaks.

In the same spirit as the fitting procedure introduced in Ref. ^[145], we define figures of merit that target these measures, and complement the indicators of sampling efficiency that were previously introduced. By giving different weights to different targets and to different frequency ranges, it is possible to generate GLE thermostats that are *designed* to have a prescribed

effect when applied to a given system. As we will show below, even in cases for which the GLE thermostat disturbs classical molecular dynamics in quite extreme ways, based on the analytical prediction of such disturbance one can recover the true dynamics of the underlying system.

4.1.2 Predicting and correcting the dynamical disturbance of a GLE

Equation 4.7 predicts the velocity-velocity correlation function for a harmonic oscillator of frequency ω_0 subject to a given GLE. If one considers an assembly of independent oscillators of different frequencies, the total correlation function of the system can be written as $\sum_i C(\omega, \omega_i)$. Taking the limit of a continuum distribution corresponding to the density of states $g(\omega)$, one can write

$$C_{vv}^{\text{GLE}}(\omega) = \int d\omega' g(\omega') C(\omega, \omega'). \quad (4.15)$$

Note that if rather than the total velocity correlation function one were computing a linear combination of correlation functions (e.g. a dipole spectrum to which each oscillator contributes with its own transition dipole moment), Eq. (4.15) would still hold, with $g(\omega)$ representing a combination of the density of states and the weight of each mode. The question, of course, is how well this relation would hold in a real, anharmonic system, and how well the indicators of dynamical disturbance can be used to tune the behavior of the GLE dynamics - given that the kernel $C(\omega, \omega')$ was derived under the assumption of harmonic dynamics. To benchmark this framework in a realistic scenario, we performed simulations of NN liquid water at 300K and experimental density. We computed the vibrational density of states from a reference NVE simulation of the same model, and then compared it with the Fourier transform of the velocity-velocity correlation function resulting from different kinds of GLE. Figure 4.2 shows the results for white-noise Langevin dynamics using different values of the friction, and two GLE matrices. GLE(A) was designed to dramatically disturb all low-frequency modes, whereas GLE(B) was optimized to only affect modes within a narrow range of frequencies between 3000 and 4000 cm^{-1} . One can see that the GLE spectrum is qualitatively distorted in accordance with the three indicators $\bar{\omega}$, $\Delta\omega$ and $S(\omega)$, but also that convoluting the NVE density of states according to Eq. (4.15) yields a near-perfect quantitative prediction of the GLE dynamics. These results

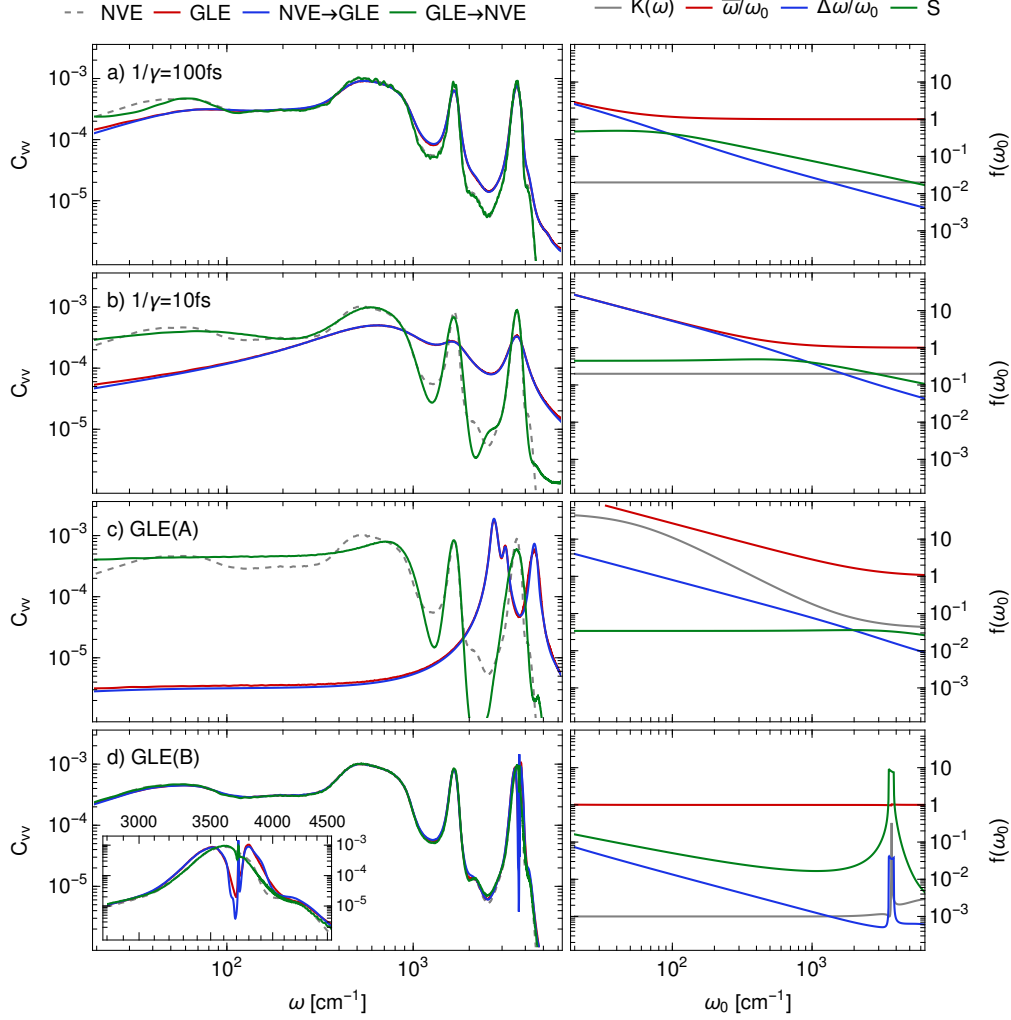


Figure 4.2 – Each row reports the velocity-velocity correlation spectrum for a thermostatted simulation of liquid water at 300K and experimental density (left) and the predicted measures of spectral disturbance ($\bar{\omega}/\omega_0$, $\Delta\omega/\omega_0$, $S(\omega)$) as a function of frequency, together with the GLE friction kernel $K(\omega)$ (right). The spectrum from the thermostatted trajectory (GLE) is compared with the density of states obtained from microcanonical runs (NVE), as well as with the spectrum predicted by convoluting the density of states with $C(\omega, \omega_0)$ (NVE→GLE) and the density of states reconstructed by deconvoluting the thermostatted spectrum (GLE→NVE). The simulations were performed with a strong white-noise thermostat (a), a very-strong white-noise thermostat (b), a GLE designed to distort dramatically the whole spectrum (c), and a GLE designed to only affect the stretching peak (d with inset).

open a path to the design of thermostats that only affect a portion of the frequencies while leaving the others untouched, as is the case for GLE(B).

Given the remarkable accuracy of the analytical prediction of the GLE dynamical disturbance, the possibility of performing the inverse operation arises – that is to analytically predict the NVE density of states given the velocity-velocity correlation function obtained from a thermostatted run. This operation corresponds to a deconvolution of the GLE spectrum using $C(\omega, \omega')$ as a convolution kernel. It is well-known that this class of inverse problems is very unstable, and that an appropriate regularization is crucial to obtain sensible results that are not dominated by noise. Direct inversion using Tikhonov regularization with a Laplacian operator led to promising but unsatisfactory results. In particular, we found a tendency to obtain large spurious oscillations in the low-density parts of the spectrum, often leading to unphysical negative-valued curves.

We therefore used the Iterative Image Space Reconstruction Algorithm (ISRA), that enforces positive-definiteness of the solution^[197,198]. Initializing the iteration with the GLE-computed velocity correlation spectrum, $f_0(\omega) = c_{vv}^{\text{GLE}}(\omega)$, the ISRA amounts to repeated application of the iteration

$$f_{n+1}(\omega) = \frac{f_n(\omega)h(\omega)}{\int dx D(\omega, x)f_n(x)} \quad (4.16)$$

where we have defined

$$\begin{aligned} h(\omega) &= \int dx C(x, \omega) c_{vv}^{\text{GLE}}(x) \\ D(\omega, x) &= \int dy C(y, \omega) C(y, x). \end{aligned} \quad (4.17)$$

The ISRA converges to a local solution satisfying $\int dx C(\omega, x)f_\infty(x) = c_{vv}^{\text{GLE}}(\omega)$. We found that a convenient way to monitor the convergence is to compute at each step the residual, and the Laplacian of f_n ,

$$\begin{aligned} r_n &= \int d\omega \left| \int dx C(\omega, x)f_n(x) - c_{vv}^{\text{GLE}}(\omega) \right|^2 \\ l_n &= \int d\omega |f_n''(\omega)|^2. \end{aligned} \quad (4.18)$$

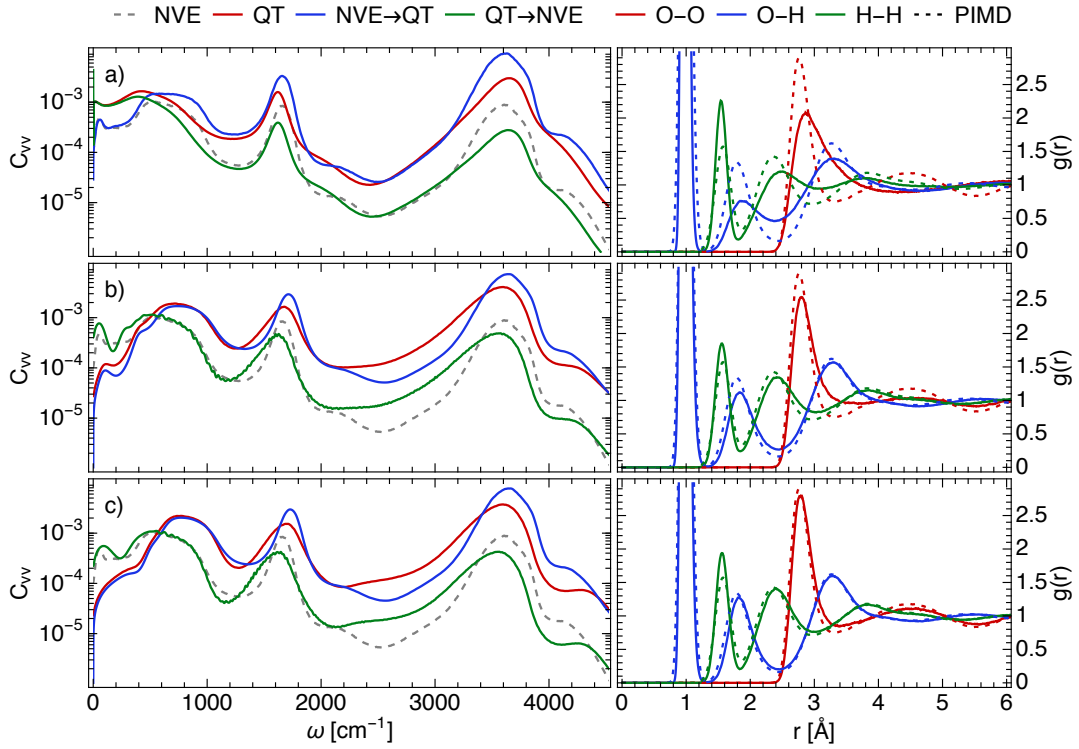


Figure 4.3 – The panels on the left report the velocity-velocity correlation functions, obtained from quantum-thermostatted simulations of liquid water at 300K and constant experimental density. As in Fig. 4.2, the spectrum from a GLE simulation is compared with the NVE density of states, as well as with the transformed and reconstructed spectra. Panels on the right depict the radial O-O, H-H and O-H distribution functions from the QT runs, compared with those from a converged PIMD calculation^[196] (dashed lines). The topmost panels correspond to a weakly-coupled GLE, the middle and bottom panels correspond to strongly coupled GLEs fitted independently

Plotting (r_n, l_n) on a log-log scale reveals a behavior resembling a L-curve plot, that can be used as a guide to avoid over-fitting – although in practice we find that the well-known slow asymptotic convergence of the ISRA effectively prevents reaching a situation in which f_n becomes too noisy. As can be seen from Fig. 4.2, this approach provides an excellent reconstruction of the true density of states even in cases in which the GLE dynamics distorts the spectrum of water beyond recognition. There are of course discrepancies, particularly in the low-frequency region that is both strongly anharmonic and harder to statistically converge. Nevertheless, the possibility of correcting for the disturbance induced by a GLE on the dynamics of complex atomistic systems opens up opportunities to obtain more accurate estimates of dynamical properties from simulations that use Langevin equations to stabilize trajectories,^[116] or that contain intrinsic stochastic terms^[199,200,201,202].

4.1.3 Dynamical properties from a quantum thermostat

Besides correcting dynamical properties in classical thermostatted simulations, this iterative reconstruction of the unperturbed DOS could be particularly helpful in another scenario. As mentioned in the Introduction, GLEs have been successfully applied as a tool to sample a non-equilibrium distribution in which different vibrational modes reach a stationary frequency-dependent effective temperature $T^*(\omega)$. In particular, the so-called “quantum thermostat”^[203] and “quantum thermal bath”^[204] try to enforce a temperature curve that mimics a quantum-mechanical distribution of energy in the normal modes of the system. Trying to maintain this temperature imbalance in an anharmonic system inevitably leads to zero-point energy leakage^[205], i.e. cross-talk between different normal modes that lead to deviations from the desired $T^*(\omega)$. This problem can be addressed by using a strongly-coupled GLE^[145], that results however in a pronounced disturbance of the system’s motion – making any inference on quantum effects on dynamical properties little more than guesswork. Being able to compensate for the dynamical disturbance induced by a GLE can make this approach somewhat more credible, and less dependent on the details of the thermostat.

Figure 4.3 gives a demonstration of this idea – as well as a clear warning to the dangers of using the results of a quantum GLE without careful validation. Let us start by discussing the accuracy of the QT in terms of structural properties, for which we can obtain a reliable

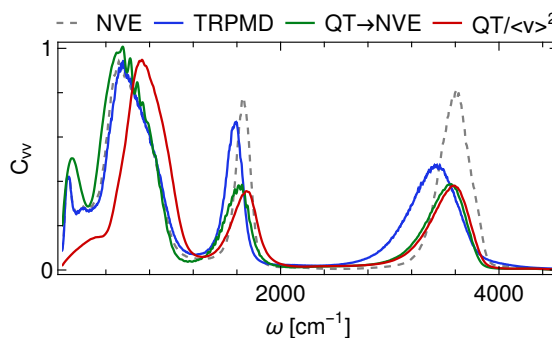


Figure 4.4 – A comparison between the classical vibrational density of states for a NN model of room-temperature water (NVE), that estimated from critically-damped TRPMD (TRPMD), with the QT velocity-velocity correlation function scaled by $C_{pp}(\omega_0)$ ($QT/\langle v \rangle^2$) and finally the dynamically-corrected QT (QT \rightarrow NVE). The QT parameters are those used for panel (c) in Figure 4.3.

benchmark from a fully converged^[196] PIMD simulation of the same NN model. As seen from the radial distribution functions, using a weakly coupled quantum thermostat (panel a) leads to significant zero-point energy leakage. The stretching modes show narrower fluctuations compared to PIMD, and the O-O distribution demonstrates a dramatic loss of structure, which is compatible with a much higher effective temperature of librational and translational modes. Increasing the coupling to the thermostat (panels b and c) improves significantly the structure of water, that becomes very close to that from the PIMD simulation. This comes however at the price of a very pronounced disturbance of the dynamical properties, that is most apparent in the low-frequency part of c_{vv} .

Moving on to dynamical properties, let us now discuss the relations between the (classical) density of states, the GLE spectrum and the curves obtained by convolution and deconvolution through the kernel¹ $C^*(\omega, \omega_0) = m\beta C_{pp}(\omega_0)C(\omega, \omega_0)$. The deconvolution process corrects at the same time for dynamical disturbances and the frequency-dependent occupations of different normal modes, so any deviation between the reconstructed spectrum and the classical DOS is an indication of anharmonic effects, and/or zero-point energy leakage that induces deviations from the target $T^*(\omega)$. As shown in the lower panel of Fig. 4.3, the iteratively-reconstructed DOS displays the qualitative features one would expect from a quantum spectrum of water: the low-frequency modes are effectively unchanged relative to a

¹It is useful to use a non-normalized kernel, as it automatically corrects for the different occupations of normal modes of different frequency when converting between the density of states and the power spectrum.

classical DOS, whereas stretches and bends show a considerable red shift and broadening. The reconstructed spectra from panels b and c – that correspond to different but strongly coupled GLEs – are qualitatively very similar, particularly when contrasted with the weakly-coupled GLE in panel a. In the latter case, the low-frequency modes are overheated, leading to an overestimation of the DOS relative to the classical limit, and the stretching peak shows a blue shift, consistent with the fact that H-bonds are broken and stretch modes are underpopulated compared to the true quantum distribution.

While there is no absolute benchmark for quantum effects on dynamical properties of water, it is useful to compare the results from the “dynamically-corrected” QT simulations with those from a TRPMD simulation. As shown in Figure 4.4, the dynamical corrections do much more than rescaling frequencies by the QT occupations $C_{pp}(\omega_0)$. The heavily-distorted low-frequency part of the spectrum becomes very close to the classical DOS, and small corrections are also applied to stretches and bending. While there is a considerable difference between the TRPMD spectrum and the corrected QT spectrum in the bending and stretching region, one should note that a similar discrepancy can be seen between TRPMD, CMD and other approximate quantum dynamical techniques^[118].

4.1.4 Dynamical properties from a PIGLET thermostat

While the results of the quantum thermostat display qualitative features associated with nuclear quantum effects such as the red shift of the stretching and bending modes, and the increase in the diffusion coefficient, they suffer from the fact that the quantum thermostat is not systematically improvable, and offers only a qualitative description of the quantum Boltzmann distribution.

In order to overcome these limitations, we show that a deconvolution scheme can also be applied to the case of spectra obtained with the PIGLET technique. Given that in PIGLET simulations the centroid of the ring polymers is subject to a classical Boltzmann sampling thermostat, we correct the velocity autocorrelation function of the centroid using the deconvolution kernel defined by the thermostat that acts on it, assuming that the (non-equilibrium) GLE thermostats that act on the internal modes of the ring-polymer do not affect the dynamics

4.1. A Generalized Langevin Equation Thermostat

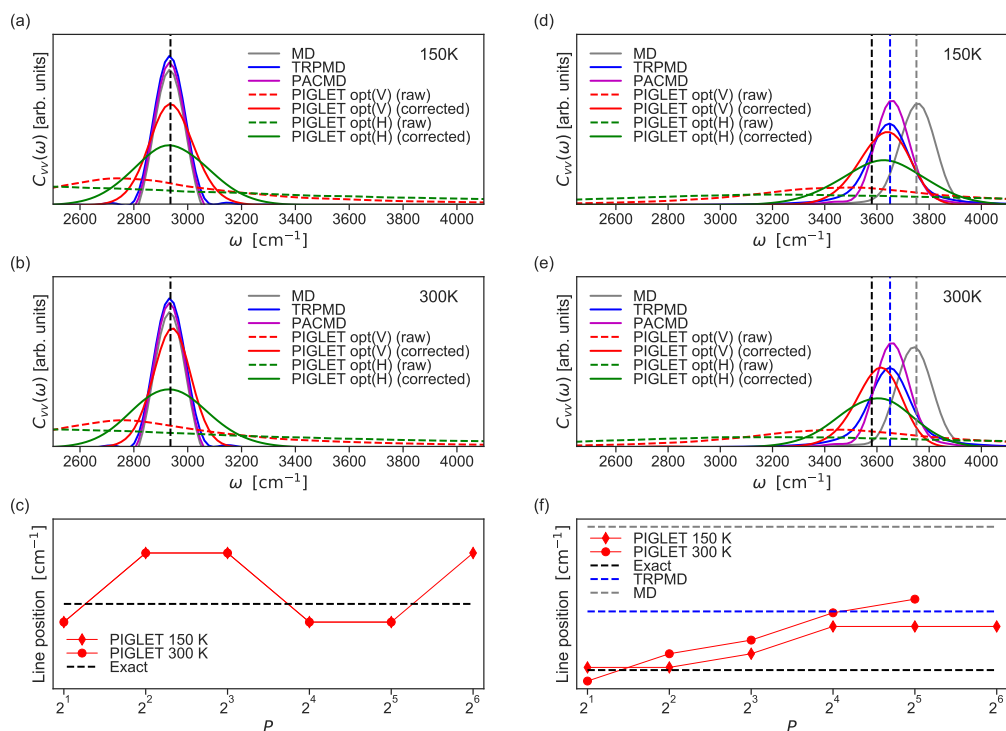


Figure 4.5 – Vibrational density of states as calculated by the velocity autocorrelation $C_{vv}(\omega)$ for a harmonic oscillator parametrized to model a C–H bond ((a), (b) and (c)) and a Morse oscillator parametrized to model an O–H bond ((d), (e) and (f)). Simulations were performed at 150 K ((a) and (d)) and 300 K ((b) and (e)) using MD (grey), TRPMD (blue), PA-CMD (purple), and using the opt(H) (green) and opt(V) (red) variants of the PIGLET thermostat. In the case of the harmonic oscillator, the curves obtained using MD and PA-CMD have been vertically shifted so that they do not coincide with that of TRPMD. The dashed and solid lines show the spectra obtained using the PIGLET thermostat (green and red), and the corrected spectra using the deconvolution scheme described in this work, respectively. Panels (c) and (f) show the convergence of the line position using the opt(V) variant of PIGLET as a function of number of replicas P at 150 K (diamond markers) and 300 K (circular markers). The dashed black line shows the 0 \rightarrow 1 transition frequency.

of the centroid. This assumption is exact in the harmonic limit for any number of beads, as the time evolution of the centroid is decoupled from the other ring polymer modes. In the limit of large number of beads PIGLET tends to standard canonical sampling, and the (purified) centroid dynamics should converge to that of TRPMD.

We begin illustrating the merits and the limitations of this approach by studying the VDOS of simple and analytic model systems for which exact results are known. We consider a one-dimensional harmonic $\hat{H} = \frac{\hat{p}^2}{2\mu} + \frac{1}{2}K(\hat{q} - q_0)^2$, with $\mu = 1694.9533$, $q_0 = 2.0598$ and $K = 0.3035$ a.u. tuned to reproduce a typical C–H stretching vibration^[206]. In the harmonic limit methods such as TRPMD, CMD and even classical MD are known to deliver exact line positions. We have used two variants of the PIGLET thermostat – that we will refer to as opt(V) and opt(H) – with different parameterizations of the centroid thermostat. The parameters of the opt(V) and opt(H) thermostats have been tailored, respectively, to minimize the autocorrelation time of the potential and the total energy, and the latter is roughly speaking twice as aggressive as the former^[74]. As shown in panels (a) and (b) of Fig. 4.5, the PIGLET thermostats have a detrimental effect on the VDOS, broadening and shifting the peaks to an extreme extent, which is the reason why they are usually considered unsuitable for the calculation of dynamical properties. The “corrected spectra” obtained by applying the deconvolution procedure, instead, have line positions that are in good agreement with the exact results at both 150 K and 300 K². The line widths of the corrected spectra are slightly larger than those obtained with other methods, even after deconvolution. Between the two variants of PIGLET, opt(H) leads to higher Lorentzian broadening, which indicates that contrary to the case of a classical trajectory the deconvolution procedure cannot eliminate completely the effect of the thermostat, making the lineshape dependent on the parameterization of the GLE. As shown in panel (c) of Fig. 4.5, the line positions obtained by “correcting” the opt(V) spectra agree with the exact results to within 15 cm^{-1} , which is the frequency resolution of the computed spectra.

Encouraged by the observation that – particularly for the weaker opt(V) centroid thermostat – deconvoluted PIGLET spectra show harmonic vibrations at the correct frequency, and only a modest peak broadening, we consider the case of a one-dimensional Morse oscillator.

²It can be shown that in the harmonic limit, for exact integration, results ought to be independent of temperature, so the agreement between the two temperatures indicate that the method is robust to integration and sampling errors.

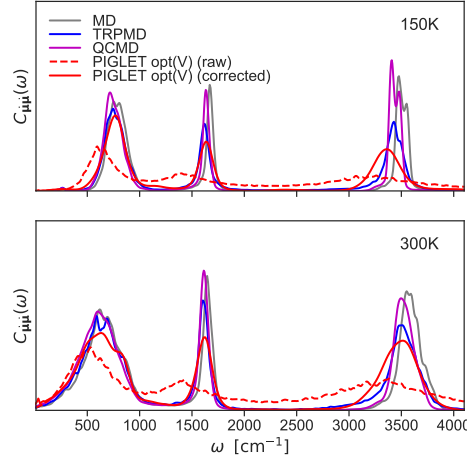


Figure 4.6 – Infra-red $C_{\mu\mu}(\omega)$ spectra as calculated by the autocorrelation of the time derivative of the dipole moment for hexagonal ice at 150 K (top) and liquid water (bottom) at 300 K, described with the q-TIP4P/f potential. Simulations have been performed using MD (grey), TRPMD (blue), QCMD (purple), and using opt(V) (red) variant of the PIGLET thermostat. The dashed and solid lines show the spectra obtained using the PIGLET thermostat (red), and the corrected spectra using the deconvolution scheme described in this work, respectively.

The parameters of the Hamiltonian $\hat{H} = \frac{\hat{p}_\mu^2}{2\mu} + D \left(1 - e^{-\alpha(\hat{q}-q_0)} \right)^2$ are tuned to describe the anharmonic O–H stretching mode in water ($\mu = 1741.0519$, $q_0 = 2.0598$ and $\alpha = 1.1605$ a.u., as in Ref. [206]), which displays a large nuclear quantum effect in the frequency of the $0 \rightarrow 1$ transition [89]. As shown in panels (e) and (f) of Fig. 4.5, the line positions obtained using all methods are blue-shifted with respect to the exact $0 \rightarrow 1$ transition frequency. Classical MD blue-shifts the line position by over 100 cm^{-1} , while CMD and TRPMD, which both require 64 and 32 replicas at 150 K and 300 K, respectively, are shifted by around 60 cm^{-1} . As shown in panel (g), the line positions of the deconvoluted PIGLET spectra remain close to the exact result for a small number of replicas and systematically converge to line positions of TRPMD as the number of replicas is increased. For the number of beads that typically converge structural observables to their quantum limit i.e. 16 replicas at 150 K and 6 replicas at 300 K [2], the results obtained using the proposed scheme are within 50 cm^{-1} of the exact result and in good agreement with TRPMD and CMD. As in the case of the harmonic oscillator the spectra obtained using the opt(H) variant of PIGLET are considerably broadened even after deconvolution. On the other hand, the spectra obtained using opt(V) are only slightly broader than those obtained with TRPMD, and give accurate line positions with a computational cost

that is 3-4 times lower.

Finally, we test the approach on condensed-phase systems. We begin by studying the IR spectrum of hexagonal ice at 150 K and liquid water at 300 K using the q-TIP4P/f water model^[207] and a linear dipole moment surface, as used in a number of prior investigations^[208,107,86]. The IR spectrum $C_{\dot{\mu}\dot{\mu}}(\omega)$ is calculated using the autocorrelation of the time derivative of the instantaneous dipole moment of the system $\mu(t)$ and is normalized to integrate to unity. Since the opt(V) variant of PIGLET has been shown to consistently provide more resolved spectra, we report results only with this variant using 6 replicas at 300 K and 16 replicas at 150 K. Furthermore, we have not used PA-CMD as it is known to exhibit a curvature problem^[209] at low temperatures. We instead present results from QCMD simulations taken from Ref.^[107]. As shown in Fig. 4.6, the thermostatted IR spectra are severely affected but show three distinct bands that correspond to the librational, bending and stretching motion. All the bands are well resolved by applying the deconvolution procedure. At 300 K, all three bands are in good agreement with TRPMD and QCMD, aside from a slight broadening of the stretching band for the deconvoluted PIGLET spectra. At 150 K, instead, the librational and bending bands remain in good agreement with TRPMD and QCMD while the stretching band is broadened and red shifted by around 60 cm^{-1} with respect to these methods. Given that in the case of gas phase water, for which exact quantum dynamical results are available, both QCMD and TRPMD yield consistently a blue-shift of 60 cm^{-1} for the stretching band^[107], the discrepancy is comparable with the typical accuracy of much more demanding methods.

Applications **Part II**

5 Anisotropy of the Proton Momentum Distribution in Water

Aqueous systems exhibit many distinctive properties^[210], many of which are affected by the quantum nature of protons, even at room temperature and above^[211]. One of the most evident signatures of the quantum fluctuations of nuclei is the deviation of the momentum distribution of protons $n(\mathbf{p})$ from the classical Maxwell-Boltzmann limit. This is essentially a consequence of the fact that position and momentum do not commute, making the distribution of momentum depend on the local potential felt by the proton. This quantity can be measured directly through Deep Inelastic Neutron Scattering (DINS) experiments^[167,25,172,212,213]. However, the interpretation of these experiments is not straightforward because the dependence of the momentum on the potential is not a trivial one, and because the spherically-averaged momentum distribution contains relatively little information, and so high quantitative accuracy of measurements and simulations is necessary to reach a compelling comparison. Early reports of large anomalies in the temperature dependence of proton kinetic energy in supercooled water^[214,215] have not been reproduced in path integral simulations using empirical forcefields^[216], and have been considerably reassessed in subsequent measurements^[169]. An early study of this discrepancy based on vibrational self-consistent field calculations came to the conclusion that softening of the OH stretch due to electrostatic interactions could not reproduce the experimental momentum distribution at room temperature and below^[217].

The following chapter is partially based on the journal article V. Kapil, A. Cuzzocrea, and M. Ceriotti, "Anisotropy of the Proton Momentum Distribution in Water," *The Journal of Physical Chemistry B*, vol. 122, pp. 6048–6054, June 2018. VK was involved in the design of the research, in the development and implementation of the method, in running the calculations, in performing analysis and in the writing of the manuscript.

The model, however, was giving a generally poor description of quantum fluctuations, and even failed to account for the sign of the change in kinetic energy upon condensation. More recently, a concerted effort between experiment^[169] and modelling^[218] has shown that widely different interatomic water potentials reproduce to a very good accuracy the equilibrium isotope fractionation (that is directly related to the quantum kinetic energy^[218]), and that their proton kinetic energy differs less than 2meV/atom among each other. Experiments close to the melting point agree very well with these estimates. Disagreement still exists, however, for room-temperature water, for which experiments consistently measure a kinetic energy which is about 7% lower than predicted by modelling. Furthermore, the quantum kinetic energy is not sufficient to characterize fully the quantum momentum distribution, that results from the combination of several vibrational modes with different zero-point energies.

A more complete comparison requires the evaluation of the full, anisotropic particle momentum statistics. The standard methods for the computation of $n(\mathbf{p})$ are “open”^[69,219] and “displaced”^[101] Path Integral Molecular Dynamics (PIMD) simulations which include exactly the quantum effects on the motion of nuclei on an *ab initio* potential energy surface. However, these simulations have a particularly high computational cost. In the case of open PIMD, the estimator used to calculate the momentum distribution allows to accumulate statistics for a single particle out of the entire simulation, and as a consequence requires very long trajectories to obtain converged averages. Displaced PIMD computes the distribution by estimators based on free energy perturbation – that can be affected by large statistical artifacts^[136] – or on thermodynamic integration – that require multiple trajectories and allows to sample one single particle. A few *ab initio* simulations exist that show qualitative agreement with experiments for the full momentum distribution^[100,220,173]. However, as discussed above, clear discrepancies still exist^[219,169,173]. The integrated nature of the signal, and the subtle dependence on thermodynamic conditions, call for a quantitative benchmark to be established.

5.1 Computational setup

Having developed and implemented an efficient sampling scheme, and demonstrated the potential of the scaled gradient estimator (see Chapter 3) in reducing the cost of computing the momentum distribution by over an order of magnitude, we begin the investigation of the momentum distribution of various phases of water at room temperature and below. Experiments performed in this regime have observed a non monotonic dependence of the kinetic energy across the temperature of maximum density of water^[215]. They indicate that supercooled water possesses an excess proton kinetic energy^[169,221], about 7% higher than that of room temperature water. In order to investigate the source of this anomaly, we use three very successful water models- q-TIP4P/f^[222], NN-B3LYP+D3^[196] and MBPOL^[223] - to estimate the proton momentum distribution and the mean proton kinetic energy respectively using open and closed PIMD simulation for ice and supercooled water at 271 K and liquid water at 296 K and 300 K using 64 replicas. Since proton momentum distribution and quantum kinetic energy are dominated by short range interactions, we used small simulations cells of water and ice containing 64 molecules and 96 molecules, respectively at experimental densities. Four proton disordered structures were used for simulating ice. To avoid the huge computational cost of *ab initio* path integral simulations, we use a neural network^[224,225] fit to reference density functional theory calculations using the B3LYP hybrid functional^[226] and D3 empirical Van der Waals corrections^[130], that was shown in previous publications to provide very accurate estimates of the quantum properties of the first-principles reference at a fraction of the cost^[196,218]. Since the MBPOL potential also has a cost that is not negligible, we fit a short-range NN potential^[224,227] to MBPOL energy and forces (parameters of the fit provided in the SI) and use it together with the i-PI implementation^[142] of multiple timestep molecular dynamics^[40] and ring polymer contraction^[113], performing a full MB-POL evaluation every 4 steps, and on a contracted ring polymer with 8 beads, which are very conservative values that still allows us to achieve a 20× speedup in terms of wall-clock time. The end-to-end vector and the forces acting on the target atoms were sampled every 2 fs. The momentum distributions were calculated using the scaled gradient estimator from 40 ps long simulations. For ice, distributions obtained from different starting configurations were further averaged.

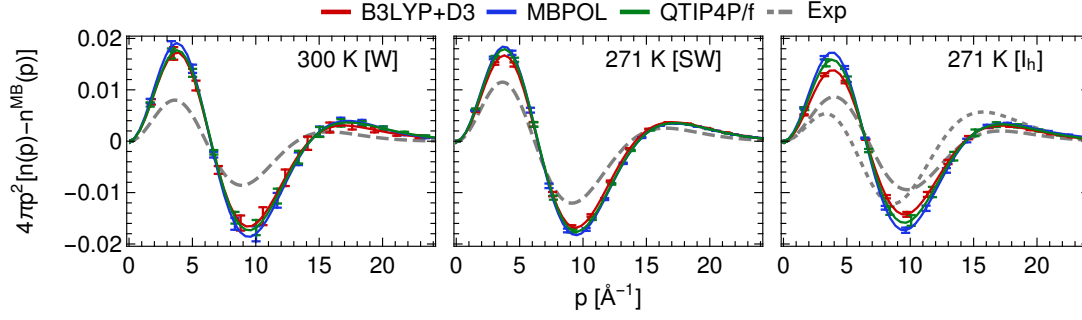


Figure 5.1 – The difference between the radial proton momentum distribution and the Maxwell-Boltzmann distribution with the same second moment for B3LYP+D3 (red), MBPOL (blue), q-TIP4P/f (green) and experimental water (grey). The dashed lines correspond to a recent DINS experiment^[169] on room temperature, supercooled and hexagonal ice. The dotted line corresponds to a distribution constructed from a Inelastic Neutron Scattering (INS) experiment for hexagonal ice at 271 K^[228].

Phase	B3LYP+D3 [\AA^{-1}]	MBPOL [\AA^{-1}]	q-TIP4P/f [\AA^{-1}]	Exp [\AA^{-1}]
271 K [SW]	(3.1, 3.9, 7.0)	(3.0, 4.0, 7.1)	(3.0, 4.0, 7.0)	(3.7, 4.3, 6.6)
271 K [I _h]	(3.4, 3.9, 6.9)	(3.1, 3.9, 7.1)	(3.1, 4.0, 7.0)	(2.9, 5.0, 6.5)
300 K [W]	(3.0, 4.1, 6.9)	(3.0, 3.9, 7.2)	(3.0, 4.0, 7.1)	(3.1, 5.3, 5.8)

Table 5.1 – Mean components of the proton kinetic energy along the x, y, z directions for various phases of B3LYP+D3, MBPOL and q-TIP4P/f water, obtained by fitting the spherical average of an anisotropic Gaussian to the momentum distribution calculated from open PIMD simulations with 64 imaginary time slices. The x direction is along the hydrogen bond and y, z lie on the plane perpendicular to it. The residuals of the fit are of the order of 10^{-3} . Experimental results for ice and supercooled water are from a recent DINS measurement^[169] while those for room temperature water were obtained from a private communication^[229].

5.2 Quantum kinetic energies

We first consider the mean proton kinetic energies listed in Table 5.1. Our results for supercooled water and ice are in excellent agreement with the experiment^[169] and the values calculated at the triple point in an earlier work^[218]. Contrary to the experimental observation, however, the increase in kinetic energy between the supercooled liquid and room-temperature conditions suggests that the proton doesn't experience any softening of free energy along the proton transfer coordinate, and that the change in kinetic energy is consistent with a simple increase in thermal excitations of the low-frequency degrees of freedom.

5.3 Anisotropic proton momentum distributions

In order to provide a more complete picture that can help reconcile simulations and experiments, we then proceed to evaluate and compare the full proton momentum distribution. Given that the shape of $n(p)$ contains an overwhelming dependence from the mean kinetic energy, we decided to show in Fig. 5.1 the difference between the momentum distribution and the Maxwell-Boltzmann distribution corresponding to the quantum kinetic energy reported in table 5.1. This particular representation highlights the anisotropy introduced by the zero point energy of the collective modes. We find that distributions calculated by the three models are very close to each other. This is in line with the results of an earlier work^[150] which showed that the kinetic energy is distributed along the stretch, bend and hindered rotation in the ratio $\sim 4 : 1.5 : 1$ for all the models, and is consistent with the main vibrational frequencies of water and with inelastic neutron scattering measurements^[221]. DINS experiments^[169] are also in good qualitative agreement with the models, although they consistently show a less-pronounced degree of anisotropy. There is also a small spread in the curves measured at different temperatures and between DINS experiments – that obtains directly $n(p)$ – and inelastic neutron scattering data^[221], that also give qualitative agreement with the computed anisotropy even though it obtains $n(p)$ from an indirect analysis. It should be stressed that the nature of the spherical averaging is such that the anisotropic momentum distribution depends rather weakly on the anisotropy of $n(p)$, leading to large error bars on the anisotropy coefficients, and to difficulties in identifying inhomogeneities in the sample^[230]. The computed $n(p)$ can be fitted to an anisotropic Gaussian lineshape^[132,173] with essentially no error, which strongly suggests that (at least for room-temperature water) the spherically-averaged distribution does not contain enough information to infer the presence of deviations from a quasi-harmonic description. Table 5.1 shows that the three models are in excellent, quantitative agreement with each other, and small changes as a function of temperature, whereas experimental values are somewhat more erratic, probably due to the difficulty in separating the anisotropic contributions, that are strongly correlated in the fit.

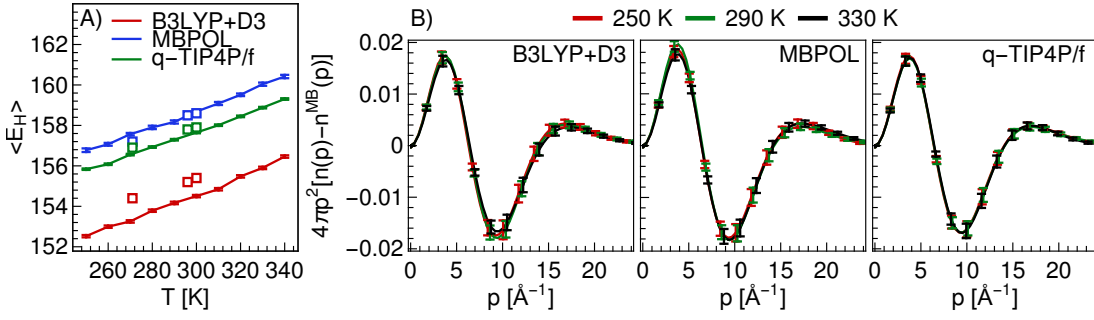


Figure 5.2 – Panel A shows the temperature dependence of the proton kinetic energy calculated using closed PIMD simulations that sample the NPT ensemble using B3LYP + D3 (red), MBPOL (blue) and QTIP4P/f (green) water potentials. The square markers correspond to the values obtained from NVT simulations at experimental densities. Panel (B) shows the temperature dependence of the difference between the proton momentum distribution calculated using NPT open PIMD simulations and the Maxwell-Boltzmann distribution with the same mean kinetic energy for B3LYP + D3 (left), MBPOL (middle) and QTIP4P/f (right) water potentials.

5.4 Constant pressure ensemble

In order to rule out subtle sources of error in the simulations, such as the fact that melting points of different models can vary substantially, and that the equilibrium molar volumes at constant pressure do not always correspond to the experimental density that we used in our NVT simulations, we also performed a scan of the temperature range from 250 K to 340 K, running for each water model 10 additional NpT simulations using both closed paths (to estimate the kinetic energy) and open paths (to estimate the anisotropy). After successfully equilibrating the densities, we study the temperature dependence of the kinetic energy and find that all the models display a monotonic increase in the kinetic energy with essentially the same slope (figure 5.2) confirming that the disagreement with experiments at room temperature is not an artifact of our simulation protocol. Density effects are also small, as evidenced by the minute differences between NVT and NpT results. The anisotropic components of $n(p)$ are even more stable across different temperatures, being essentially constant for each model over a temperature range of almost 100K – which is small compared to the scale of the quantum kinetic energy of protons.

5.5 Discussion

In summary, we find that the particle momentum distribution of water does not show any trace of anomalous behavior when going from the supercooled limit to temperatures well above room temperature, even when considering the full, anisotropic form of $n(p)$. Our analysis is particularly thorough, including three very different water models: a fixed point-charge model, a dissociable machine-learned potential fitted to hybrid density functional theory data, and a state-of-the-art many-body forcefield that has shown consistently quantitative accuracy in reproducing many of the classical and quantum properties of water^[231,232,233]. We repeat our simulations at different thermodynamic conditions, spanning a range of 100 K around the experimental melting point, to exclude the possibility that anomalous effects may be shifted to a different temperature due to the inaccuracies of the models. The three potentials are in excellent agreement with each other, despite being based on widely different models and physical approximations. The ≈ 2 meV accuracy of DINS measurements is not sufficient to discriminate between them. This should elicit new efforts to improve the accuracy and reliability of DINS experiments, that despite the technical challenges offer a rather unique approach to probe quantum nuclear fluctuations in aqueous environments, complementing other kinds of neutron spectroscopies^[221]. We suggest that ice Ih constitutes a very promising system to be used for benchmarking. Measurements on single crystals should give a better handle of the anisotropy of the distribution, and allow for a more stringent benchmark of both atomic-scale models and of measurements of the integrated, radial distribution, that is necessary before a comparison in more challenging conditions^[210] can be meaningfully attempted. Meanwhile, the scaled-gradient estimator we introduce here, together with the efficient and comprehensive implementation in an open-source code that can be interfaced with many electronic-structure packages, and the use of short-range machine-learning potentials that reduce the cost of performing high-end ab initio path integral calculations, should make it possible to further test the accuracy of the atomistic models for this difficult problem. Efforts in this direction should focus on the remaining disagreement between experiments and models at room temperature. As we have shown, the computed particle momentum distribution is essentially the same for increasingly accurate potentials, that also exhibit a kinetic energy that is consistent with equilibrium fractionation data of liquid and gaseous

water. If this last discrepancy cannot be traced to an experimental problem, it might be the signal for an extraordinarily elusive anomalous behavior of this deceptively simple substance.

6 Quantum spectroscopy of Water

Complementing the study of the anisotropic proton momentum distributions of different phases of water, we calculate quantum spectroscopic quantities for solid and liquid phases of water, that shed light on the chemical environment of atoms by looking at their responses to the electric field of neighbouring atoms. Quantitatively accurate spectroscopic predictions of water are of utmost important, as they not only enrich our understanding of the hydrogen bond network across different phases of water^[168], but also lay the foundation for reliable modelling of more complex systems such as aqueous solutions of acids and bases^[56], and phenomena such as aqueous reactions^[234], and dissolution of salts and molecular crystals in water^[235].

Simulating IR and Raman spectra of water in an accurate manner presents several challenges, such as a high accuracy in modelling the Born Oppenheimer (BO) potential energy landscape (PES), and dielectric response properties such as polarization and polarizabilities, as well as an explicit treatment of quantum dynamics of the nuclear degrees of freedom. Here, we use our deconvolution scheme (see Chapter 4), that has been shown to recover quantum dynamical properties from path integral simulations accelerated by Generalized Langevin equations (GLE) thermostats, to inexpensively model quantum nuclear dynamics of water in

The following chapter is partially based on the journal article V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, "Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats," *The Journal of Chemical Physics*, vol. 152, p. 124104, Mar. 2020. VK was involved in the design of the research, in the development and implementation of the method, in running the calculations, in performing analysis and in the writing of the manuscript.

combination with first principles level predictions of the PES and dielectric responses using state of the art machine learning techniques.

6.1 Machine learning models

In order to inexpensively evaluate the dipole and polarizability surfaces, we trained a new model based on the recently developed symmetry-adapted Gaussian process regression (SA-GPR) ^[236] framework, which has proven to be capable of generating highly accurate models of the polarizability of organic molecules ^[237] and of molecular crystals ^[238]. We trained models for the polarization $\boldsymbol{\mu}$ of pure water and for the polarizability $\boldsymbol{\alpha}$, using 1000 boxes of pure water, each containing 32 molecules, generated by replica exchange molecular dynamics (REMD) simulations with the q-TIP4P/F ^[207] water model ¹. The polarization, $\boldsymbol{\mu}$ and polarizability, $\boldsymbol{\alpha}$ for each frame was calculated at the DFT/PBE-USPP level using the modern theory of polarizability ^[239], with a plane-wave cutoff of 55 Ry and a $5 \times 5 \times 5$ k -point grid.

We generated models for the polarization (a vector quantity) and the polarizability (a rank-2 symmetric tensor) using symmetry-adapted Gaussian-process regression (SA-GPR), along with λ -SOAP kernels ^[236]. We will refer to the models as $\mu\text{H}_2\text{O}$ for polarization and $\alpha\text{H}_2\text{O}$ for polarizability. The $\mu\text{H}_2\text{O}$ model predicts the polarization of a bulk water system \mathcal{A} as a spherical tensor of order $\lambda = 1$ $\boldsymbol{\mu}(\mathcal{A})$ with components $m \in \{-1, 0, 1\}$,

$$\mu_m(\mathcal{A}) = \sum_{i \in \text{env}} \sum_{m'} k_{m,m'}^{1,\zeta}(\mathcal{A}, \mathcal{X}_i) w_{i,m'}, \quad (6.1)$$

where i is a local environment extracted from the training set, $k_{m,m'}^{1,\zeta}(\mathcal{A}, \mathcal{X}_i)$ is a $\lambda = 1$ spherical kernel between molecule \mathcal{A} and environment \mathcal{X}_i , and $w_{i,m'}$ is a weight; the vector of weights \boldsymbol{w} is obtained by a regularized matrix overlap.

The subscript ζ in the definition of the kernel indicates that it is taken by multiplying a pure

The work on training machine learning models was performed by Dr. David M. Wilkins

¹These data first appeared in Ref. ^[236] and can be found at <https://archive.materialscloud.org/2018.0009/v1>

spherical kernel by a scalar kernel raised to some power,

$$k_{m,m'}^{1,\zeta}(\mathcal{A}, \mathcal{X}_i) = k_{m,m'}^1(\mathcal{A}, \mathcal{X}_i) (k^0(\mathcal{A}, \mathcal{X}_i))^{\zeta-1}. \quad (6.2)$$

The model was optimized with respect to the following hyperparameters:

- n_{\max} , the number of radial functions used to resolve the local density.
- l_{\max} , the highest degree of spherical harmonic used to resolve the local density.
- ζ , the exponent defined by Eq. (6.2).
- n_{comp} , the number of spherical harmonic components retained to represent each environment, selected by a farthest-point sampling (FPS) method^[240].
- n_{env} , the number of environments drawn from the training set^[241,242]. The environments to be used were chosen by furthest point sampling.
- σ_G , the width of Gaussians used to produce the SOAP density field^[243].
- r_0 and m , parameters for radial scaling of the weights associated with each atom in a local environment^[244]. The radial cutoff for an environment was set to $r_c = 5.0 \text{ \AA}$.
- σ^2 , the regularization applied when inverting kernel matrices.

These hyperparameters were optimized with 5-fold cross-validation using 800 of the molecules. Table 6.1 gives the optimal hyperparameters for $\mu\text{H}_2\text{O}$ (the same parameters were used here to build the spherical kernel and the scalar kernel used in applying Eq. (6.2)).

Table 6.1 – Hyperparameters used to build the $\mu\text{H}_2\text{O}$ model.

n_{\max}	l_{\max}	ζ	n_{comp}	n_{env}	$\sigma_G/\text{\AA}$	$r_0/\text{\AA}$	m	η
6	4	2	400	1000	0.20	2.0	5	1×10^{-5}

The $\alpha\text{H}_2\text{O}$ model predicts the $\lambda = 0$ (scalar) and $\lambda = 2$ components of the polarizability

separately as,

$$\alpha^0(\mathcal{A}) = \bar{\alpha}^0 + \sum_{i \in \text{env}} k^{0,\zeta}(\mathcal{A}, \mathcal{X}_i) w_i, \quad (6.3)$$

$$\alpha_m^2(\mathcal{A}) = \sum_{i \in \text{env}} \sum_{m'} k_{m,m'}^{2,\zeta}(\mathcal{A}, \mathcal{X}_i) w_{i,m'}, \quad (6.4)$$

where ζ is defined analogously to Eq. (6.2), and $m, m' \in \{-2, -1, 0, 1, 2\}$. $\bar{\alpha}^0$ is the average of the scalar part of the polarizability in the training set. The two predictions are then combined to give a prediction of the full Cartesian tensor. The definitions of these two spherical components are as given in Ref. [237]

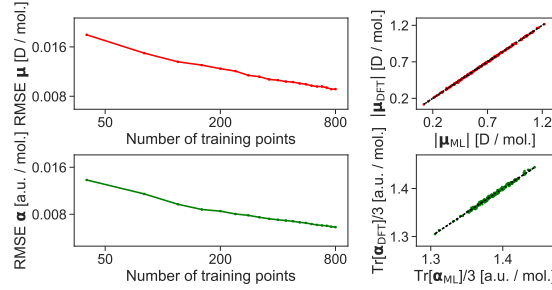


Figure 6.1 – Left: Learning curves for the $\mu\text{H}_2\text{O}$ (top) and $\alpha\text{H}_2\text{O}$ (bottom) models for the dipole moment μ and polarizability α of bulk water, showing the root mean square error (RMSE) in predicting μ and α of a testing set containing 200 frames as a function of the number of frames used to train the model. Right: Parity plots (right) of the predictions of these models with ideal predictions shown in dashed black lines.

The $\alpha\text{H}_2\text{O}$ model is comprised of two models, which are trained separately, and each of which has its own set of hyperparameters. Though the $\lambda = 0$ kernel that is used in building the $\lambda = 2$ model does not need to be the same as that used in building the $\lambda = 0$ model, the same hyperparameters were used for each in building $\alpha\text{H}_2\text{O}$. Table 6.2 gives the hyperparameters used in building this model.

Table 6.2 – Hyperparameters used to build the $\alpha\text{H}_2\text{O}$ model.

	n_{max}	l_{max}	ζ	n_{comp}	n_{env}	$\sigma_G/\text{\AA}$	$r_0/\text{\AA}$	m	η
$\lambda = 0$	8	6	2	400	1000	0.35	2.0	4	1×10^{-6}
$\lambda = 2$	4	2	2	400	1000	0.30	2.0	4	1×10^{-6}

$\mu\text{H}_2\text{O}$ predicts the polarization per molecule with an error of 8.8×10^{-3} D, or $\sim 1\%$ of the intrinsic variation in the training set, and predicts an average molecular dipole moment

of 2.8 D in this set, which is well within experimental error bars^[245]. $\alpha\text{H}_2\text{O}$ predicts the polarizability per molecule with an error of 5.9×10^{-2} a.u., or $\sim 11\%$ of the intrinsic variation in the training set. Fig. 6.1 summarizes the performance of these two models in predicting the dielectric response properties of bulk water.

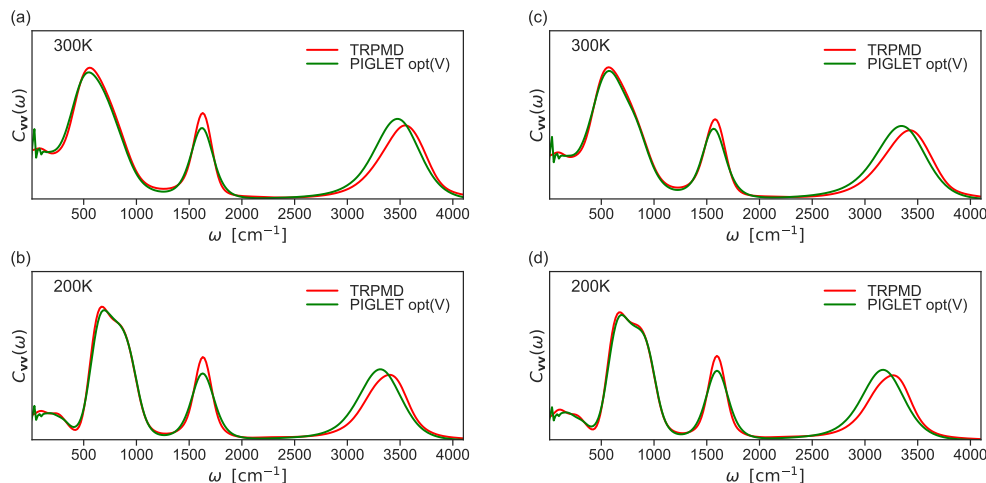


Figure 6.2 – Vibrational density of states as calculated by the velocity autocorrelation $C_{vv}(\omega)$ for liquid water at 300 K ((a) and (c)) and ice ((b) and (d)) at 200 K with PIGLET (green) and TRPMD (red) methods, using neural network potentials trained on revPBE0+D3 ((a) and (b)) and B3LYP+D3 ((c) and (d)) energetics.

6.2 IR and Raman spectra

We first generate 100 ps long trajectories with opt(V) PIGLET parameters, using 12 replicas for hexagonal ice at 200 K and 6 replicas for liquid water at 300 K. In order to assess the comparative importance of the underlying potential energy surface and that of the approximations to quantum dynamics, we ran simulations with two ML potentials based on a Behler-Parrinello neural network framework^[224] – the one introduced in Ref.^[2] based on B3LYP+D3 reference calculations, and the one introduced in Ref.^[246], based on revPB0+D3. As shown in Fig. 6.2, the peaks in the deconvoluted PIGLET velocity-velocity correlation spectra agree well with those from thermostatted ring polymer molecular dynamics (TRPMD), except for a red-shift of the stretching of $\approx 80\text{ cm}^{-1}$, similar to what observed for q-TIP4P/f. For this reason, in what follows we only discuss results for PIGLET, and compare them to experiments.

As shown in panels (a) and (c) of Fig. 6.3, the stretching band of the IR spectrum of liquid

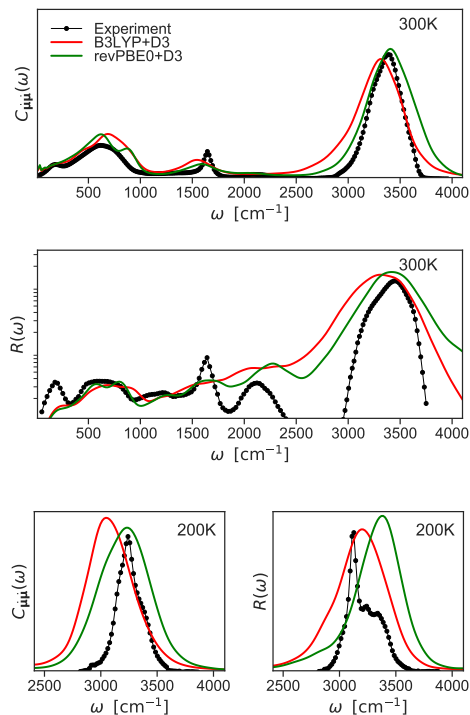


Figure 6.3 – Infra-red $C_{\mu\mu}(\omega)$ and the reduced depolarized Raman $R(\omega)$ spectra calculated using the autocorrelation of the time derivative of the dipole moment and the polarizability for hexagonal ice (bottom) at 200 K and liquid water (top) at 300 K using the PIGLET thermostat. Simulations were performed using neural network potentials trained on B3LYP+D3 (red) and revPBE0+D3 (green) data. Black curves show results obtained from experiments^[247,248,249].

water is well reproduced by the revPBE0+D3 potential, in agreement with Ref.^[189], while the B3LYP+D3 potential red shifts the stretching band with respect to the experimental result^[247]. The librational band is well reproduced by both potentials. The only major discrepancy in the spectra is the over-softening of the bending band, which is an artefact that was already observed in Ref.^[2] and linked to a shortcoming of the neural network potential. As shown in panels (b) and (d) of Fig. 6.3, the position of the stretching band of the reduced depolarized Raman spectrum obtained from revPBE0+D3, computed using the time derivative of the instantaneous polarizability of the system following the procedure of Ref.^[250], is in excellent agreement with both experimental result and previous calculations^[189]. As in the case of IR, B3LYP+D3 systematically red shifts the stretching band with respect to the experimental result^[248]. The bending and libration portion of the spectrum, however, is reproduced only qualitatively. Due to the overwhelming intensity of the stretching band, the deconvolution algorithm introduces levels of noise in the lower-frequency region that are comparable to the intensity of the spectrum. At 200 K the position of the IR stretching band of hexagonal ice calculated using the revPBE0+D3 potential is in excellent agreement with experiments^[249] as shown in the bottom panel of Fig. 6.3. However, the Raman stretching band of ice is blue-shifted by revPBE0+D3 and fortuitously well reproduced by the B3LYP+D3 potential, suggesting that a significant part of the error in modeling the Raman spectrum of ice is due to the inherent limitations of DFT. The agreement of the stretching IR and Raman peaks with the experiments could also arise from a partial error cancellation from the approximations in DFT, and the neglect of the non-centroid components of the spectra. Therefore, it would be interesting to figure out the extent of non-linearity of the dipole and polarizability time correlation functions by comparing their $t \rightarrow 0$ limit, with the results from path integral molecular dynamics. Overall, the accuracy of our results shows that the approximation inherent in applying the deconvolution procedure to the correlation spectra of non-linear dielectric responses does not introduce major artefacts, and that the level of consistency between different approximate quantum dynamics methods is equivalent or better than that between different flavors of (hybrid) density functional theory.

6.3 Discussion

When combined with ML potentials trained on dispersion corrected hybrid DFT and with ML predictions of dielectric responses computed at the GGA level, the deconvolution method provides IR and Raman spectra of water that are in good agreement with experiment, with a discrepancy with respect to TRPMD that is comparable to the difference between two reference electronic-structure methods. The ML response functions include non-linear terms, showing that the dynamical correction – despite having been derived for linear operators – performs very well also in the presence of mild non-linearities. The reduction in computational cost due to GLE thermostats, will make it possible to routinely assess the importance of quantum nuclear effects in the dynamical behavior of gas phase molecules, and complex condensed-phase systems, where the combination of TRPMD or centroid molecular dynamics with first-principles energetics becomes prohibitively expensive. Our results show that ML dielectric predictions, together with dynamical corrections to GLE-accelerated path integral simulations open the way to accurate and computationally efficient modeling of both standard spectroscopic probes such as IR and Raman, and of more advanced spectroscopies such as second harmonic scattering and sum-frequency generation.

7 Modeling structural and thermal properties of loaded metal-organic frameworks

Tailor-made porous materials^[251] like metal-organic frameworks (MOFs)^[252] are at the core of emerging technologies due to their exceptional physical and chemical properties, such as a tunable ultrahigh porosity and an associated enormous gas storage capacity. Therefore, they have been proposed for applications such as adsorbed natural gas (ANG) storage in vehicles^[253,254], adsorption-driven heat pumps^[255,256], and carbon capture and sequestration (CCS)^[257,258]. While a lot of work still needs to be done to optimize the crucial adsorption and storage properties of these porous materials,^[259] studies on other critical requirements such as heat management are gaining interest.^[260] For instance, the heat capacity, *i.e.*, the amount of energy required to increase the material's temperature, is a fundamental thermodynamic property of interest in these applications which involve large thermal fluctuations, as adsorption and desorption processes imply the release or consumption of energy. Moreover, the heat capacity of the MOF affects the energy penalty to regenerate the adsorbent in, for example, CCS.^[261] To date, however, information on the heat capacity is lacking for most MOFs^[161,262] and the influence of adsorbed guest molecules on the heat capacity has not yet been investigated, in contrast to other thermal properties such as the thermal expansion

The following chapter is partially based on the journal article V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. Van Speybroeck, and M. Ceriotti, "Modeling the Structural and Thermal Properties of Loaded Metal–Organic Frameworks. An Interplay of Quantum and Anharmonic Fluctuations," *Journal of Chemical Theory and Computation*, vol. 15, pp. 3237–3249, May 2019. VK was involved in the design of the research, in the development and implementation of the method, in running some of the calculations, in performing analysis and in the writing of the manuscript.

behavior and the thermal conductivity.^[263,264]

Within this context, an efficient and accurate simulation protocol to tackle the structural and thermal properties of MOFs including all the relevant physical effects could facilitate a better understanding of the structure-property relations and suggest design principles for materials with improved properties. Due to the importance of finite-temperature effects, anharmonicity, and nuclear quantum effects (NQEs), the modeling of the thermophysics of MOFs is generally not a trivial exercise. The first two effects have already been the subject of many investigations and were included in our protocol to characterize the thermodynamics of MOFs.^[265,266] Furthermore, very recently, some of the present authors highlighted the necessity of an accurate theoretical framework for the design of thermoresponsive MOFs.^[267] However, the impact of NQEs has so far received far less attention within the MOF community, despite the many light atoms contained in the crystal structure and present inside the pores.^[268,269,270]

In this regard, path integral molecular dynamics (PIMD)^[68] provides an ideal reference framework for the evaluation of thermodynamic averages, as it seamlessly captures both NQEs and the anharmonic motion of nuclei. The statistics of distinguishable quantum particles can be obtained through the equivalence between the thermodynamics of a quantum system and a classical ring polymer containing P replicas of the system.^[271] In the limit of large P values, NQEs can then be systematically accounted for. The major downside of this technique is the associated high computational cost, *i.e.*, P times the corresponding cost of classical molecular dynamics (MD) simulation. However, several methodological advances^[134,272,196,91,142,135] that enable a reduction of the computational cost have made it a mainstream technique for material modeling.^[273]

An additional difficulty arises from the fact that most experiments and practical applications are performed in isothermal-isobaric conditions, while the vast majority of atomistic simulations are performed with a fixed unit cell, corresponding to isochoric conditions. As most solid materials have a very small compressibility, the difference between the two ensembles is often negligible. For MOFs on the contrary – particularly when loaded with a gas – the behavior in isobaric and isochoric conditions can be very different. Some of us emphasized the importance of taking into account the variations of the cell shape to simulate properties of flexible

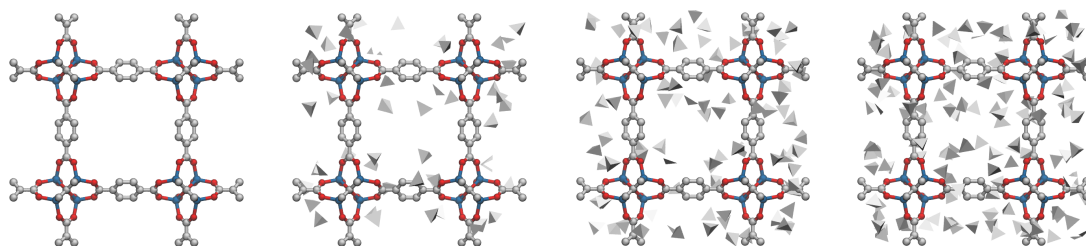


Figure 7.1 – The structure of the MOF-5 with a gas loading x of 0, 50, 100, and 150 molecules of methane (left to right) in the conventional unit cell $(8(\text{Zn}_4\text{O}(\text{CO}_2)_6) \cdot x \text{CH}_4)$. The oxygen, carbon, and zinc atoms are shown in red, silver, and blue respectively. For the sake of aesthetics, the hydrogen atoms are not included. The methane molecules are represented by silver tetrahedra.

MOFs.^[265] While algorithms for performing path integral simulations at constant pressure conditions exist^[72,73], an accurate evaluation of the thermophysical properties requires a very large number of replicas to reach convergence. Here, we use the high-order path integral method that we have developed (see chapter 2), to greatly accelerate the convergence of these simulations.

This method, in combination with a first-principles-based force field^[274,275,266], makes it possible to characterize the structural and thermophysical properties of complex molecular systems such as guest-loaded MOFs. We investigate the archetypical case of the well-known MOF-5^[276,277] in the presence and absence of methane in its pores (see Figure 7.1). Evaluating and understanding the impact of methane adsorption on the properties of MOFs is especially important as they have been proposed as potential adsorbents for natural gas storage applications.^[253,259,254] We demonstrate the crucial role of a complete statistical-mechanical description of the quantum and anharmonic fluctuations in MOFs for a correct description of the structural properties and the heat capacity of guest-loaded MOFs. By meticulously disentangling anharmonic and nuclear quantum effects for both the lattice and the guest molecules, we are able to propose an efficient empirical calculation scheme which may be used to screen MOFs for beneficial thermal properties on a larger scale.

7.1 Materials

The materials that are considered in this theoretical work are pristine and methane-loaded MOF-5 scaffolds.^[276] This framework consists of $\text{Zn}_4\text{O}(\text{CO}_2)_6$ inorganic nodes connected through 1,4-benzenedicarboxylate (bdc) linkers. The unit cell is cubic and contains eight inorganic nodes, as shown in Figure 7.1. We consider three different loadings x of 50, 100, and 150 methane molecules in the conventional unit cell ($8(\text{Zn}_4\text{O}(\text{CO}_2)_6) \cdot x \text{CH}_4$), which encompass both the low- and high-adsorption regime.^[278,253] At 100 bar and room temperature, for example, approximately 120 methane molecules are present per conventional unit cell, as experimentally measured by Mason *et al.*^[253].^[279] It is also important to note that experimentally the chemical potential is controlled and not the number of guest molecules, and that the experimental loading is highly dependent on the temperature. For computational convenience, we performed all our molecular dynamics simulations at a fixed number of particles. The purpose of comparing multiple loads is only to provide qualitative insight into the interplay between temperature, framework, and guest molecules rather than to compare directly with experiment.

7.2 First-principles-derived force fields

The molecular simulations are performed using newly developed force fields for MOF-5 and methane. They are derived with QuickFF^[274,275], a software package developed to derive force fields for MOFs in an easy yet accurate way based on information obtained from first-principles input data. Isolated cluster models were used to generate the required first-principles input data, which includes the equilibrium geometry and Hessian together with the atomic charges. Within the QuickFF protocol, the quantum mechanical potential energy surface (PES) is approximated by a sum of analytical functions of the nuclear coordinates that describe the covalent and noncovalent interactions. The covalent interactions, which mimic the chemical bonds between the atoms, are approximated by different terms as a function of the internal coordinates (bonds, bends, out-of-plane distances, and dihedrals). The noncovalent interactions are composed of electrostatic and van der Waals interactions. The guest-host interactions between MOF-5 and methane only include noncovalent terms.

7.3 Calculation of thermodynamic observables

The simple OP estimators for the total energy (E) and enthalpy (H) are listed in Chapter 3. Since the standard estimators for the heat capacity in path integral methods tend to be very complex and exhibit a large variance, we have derived an OP double-virial estimator for the isobaric (and isochoric) heat capacity $C_P = \frac{\partial \mathcal{H}}{\partial T}$. The derivation is presented in chapter 3, where it is also shown that this estimator has very good statistical properties and outperforms existing heat capacity estimators^[163]. However, in this study we computed thermophysical properties over a broad range of temperatures and found it more convenient to estimate C_P by means of a finite difference approximation to the temperature derivative of the enthalpy:

$$C_P(T) = \frac{\partial H}{\partial T} \approx \frac{H|_{T+\Delta T} - H|_{T-\Delta T}}{2\Delta T}. \quad (7.1)$$

A dedicated estimator will prove useful in simulations that are targeted at a single, specific temperature.

7.4 Computational Details

The required first-principles cluster data for the determination of the covalent terms in the force field are generated with Gaussian 16^[280] using the B3LYP^[281] exchange-correlation functional. A 6-311G(d,p) basis set^[282] is used for the C, O, and H atoms, together with the LanL2DZ basis set for Zn.^[283] The atomic charges are derived with the Minimal Basis Iterative Stockholder (MBIS) partitioning scheme^[284]. The atomic charges of the MOF-5 clusters are obtained from the PBE^[285] electron density computed with GPAW^[286]. For methane, the atomic charges are derived from the B3LYP all-electron density obtained with Gaussian 16. The parameters of the van der Waals interactions are taken from the MM3 force field.^[287,288] The van der Waals interactions are calculated up to a cutoff of 15 Å and a tail correction is added to the potential and its derivatives.^[289] The initial configurations of the methane molecules are generated using RASPA^[290] by inserting methane molecules at random positions, while ensuring that only realistic intermolecular distances are retained. Afterwards a canonical Monte Carlo algorithm was used to equilibrate the positions.

Chapter 7. Modeling structural and thermal properties of loaded metal-organic frameworks

For MOF-5 with and without methane, we perform classical and path integral MD simulations at a mechanical pressure of 1 bar and at different temperatures in the range of 100 K to 500 K. The classical MD simulations of both loaded and pristine MOF-5 are performed using Yaff in the classical isothermal-isobaric ensemble, without constraints on the unit cell.^[265] While the covalent interactions are calculated by Yaff, the expensive long-range interactions are computed by LAMMPS^[291] in a computationally efficient manner. The equations of motion are updated via a Verlet scheme, with a timestep of 0.5 fs. The temperature is controlled via a single Nosé-Hoover chain consisting of three beads, with a relaxation time of 100 fs.^[292,293,294] A Martyna-Tobias-Klein barostat with a relaxation time of 1000 fs is used to control the pressure.^[295,51,265] We performed five independent runs of 500 ps, starting from a different random seed and from different methane positions. For the empty MOF-5, a single trajectory of 500 ps was used. An equilibration time of 100 ps was considered.

The PIMD simulations are performed with the universal force engine i-PI^[5] in the Suzuki-Chin isothermal-isobaric ensemble, in which the cubic symmetry is kept fixed. The evaluation of the forces is carried out by Yaff and lammps, similar to the classical MD simulations, while the time evolution of the nuclei to sample the appropriate thermodynamic ensemble is done with i-PI. To control the temperature, a PILE-L thermostat^[117] is applied to the system and a white noise Langevin thermostat^[42] is applied to the cell. To control the pressure, a path-integral version of the Bussi-Zykova-Parinello (BZP) barostat^[296,73], adapted to the SC scheme, is used. The time constants for the thermostats and the barostats are same as the ones used in the classical simulations. A BAOAB type^[297] MTS scheme^[40] is used to integrate the equations of motion. The computationally cheap short-range terms of the force field are computed on 64 replicas and integrated with a timestep of 0.25 fs. The remainder of the interactions, *i.e.* the expensive long-range interactions, are computed on 8 replicas using RPC and integrated with a timestep of 1 fs. As discussed above, a finite differences strategy is adopted to determine the heat capacity from the enthalpy with a temperature interval of 25 K. We performed thirty independent runs of 50 ps, starting from a different random seed and from different methane positions. For the empty MOF-5, five independent trajectories of 125 ps were used. An equilibration time of 25 ps was considered.

The importance of the inclusion of NQEs and anharmonicities in the modeling of the heat

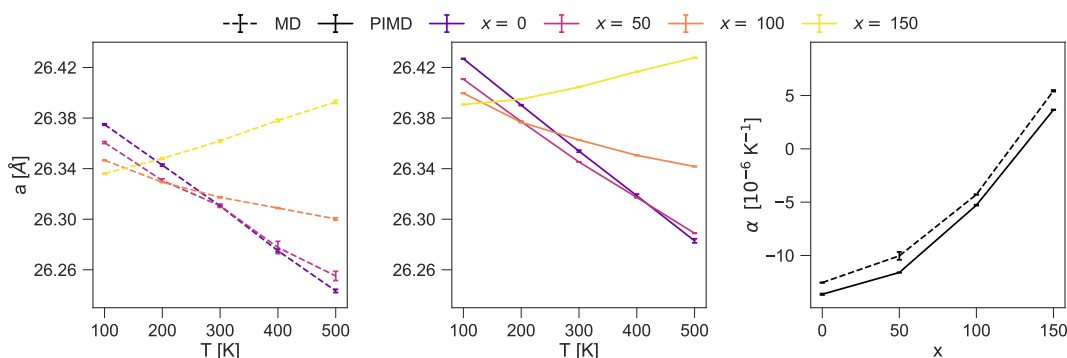


Figure 7.2 – Panels (a) and (b) show the lattice parameter a of MOF-5 with $x = 0, 50, 100, 150$ molecules of methane as a function of temperature (T), obtained from classical MD and PIMD respectively. Panel (c) shows the linear thermal expansion coefficient (α) as a function of x . The classical and quantum estimates are respectively shown with dashed and solid lines. Error bars indicate statistical uncertainty.

capacity is probed by comparing the results with other methods such as classical MD, which neglects NQEs, and the harmonic approximation, which neglects anharmonicity. We discuss the accuracy of these commonly-adopted approximations and provide empirical relations, which might resolve the general lack of knowledge of the heat capacity of this class of materials.

7.5 Structural properties

To unravel the influence of adsorbates on the framework and finally on the heat capacity, we start by investigating the structural response of MOF-5 for various loadings and temperatures. Here, one could expect that a proper inclusion of NQEs already becomes important as zero-point effects were recently found to substantially increase the volume of MOF-5 when comparing classical MD with PIMD.^[270] Additionally, NQEs have previously been observed to change the volume of bulk alkanes by about 10%.^[298,299,300] A comparison of Figures 7.2 (a) and (b) indeed reveals that the inclusion of NQEs increases the volume by almost 1 % for all loadings and temperatures. Horizontally, this shift corresponds to a more substantial temperature reduction of about 100 K.

Interestingly, the qualitative changes in the volume as a function of loading at the different temperatures does not change appreciably with or without the inclusion of NQEs. At low tem-

peratures, the material slightly shrinks in the presence of methane. The observed adsorption-induced deformation can be understood by attractive van der Waals interactions between the framework and the adsorbed methane.^[301,302] At higher temperatures, by contrast, the empty framework has the lowest volume, as entropic and kinetic effects start dominating and the adsorbed molecules increase the internal pressure, which leads to a volumetric expansion when increasing the loading. The main effect of NQEs on the volume of the guest-loaded system is thus an upward volume shift,^[270] which is to a large extent independent of the number of guest molecules and the temperature.

Varying the concentration and the type of adsorbates in the framework was suggested by Calero and co-workers^[264] as a way to control and tune the thermal expansion of a system based on classical MD simulations. We confirm that with methane it is possible to go from the well-known negative thermal expansion behavior of the empty framework^[303,304,305] towards positive thermal expansion. A proper inclusion of NQEs in our molecular dynamics simulations does not influence the predictions in this temperature window^[270] and this conclusion still holds with methane in the pores, as shown in Figure 7.2 (c).

A more surprising picture emerges when looking at the distribution of methane inside MOF-5. Recent PIMD simulations^[299] of bulk methane (at 110 K) have shown that NQEs lead to significant changes in the structure of methane at low temperature, corresponding to an overall softening of the structure and an increase in the intermolecular distance by about 0.1 Å. In contrast, in our study of methane confined in the pores of MOF-5, even at 100 K – where NQEs are expected to be the greatest – there is no appreciable difference between the shape of the classical and quantum distribution functions of methane, as shown in Figure 7.3. This can be understood by the fact that the change in the structure of bulk methane comes entirely from the isotropic expansion of the gas¹. In the case of methane molecules confined in the pores, the low compressibility of the framework makes the expansion as observed in bulk methane when including NQEs impossible.

This discussion shows that the structural response of MOF-5 to a varying number of methane

¹A 10 % increase in volume due to NQEs, assuming the molecules to be spherical is associated with a $(1.10)^{\frac{1}{3}}$ times increase in the effective radius. The classical "radius" is close to 4 Å. And thus the increase in the intermolecular distance due to NQEs coming from the isotropic expansion of the gas is $(1.10)^{\frac{1}{3}} * 4 - 4 \approx 0.13$ Å.

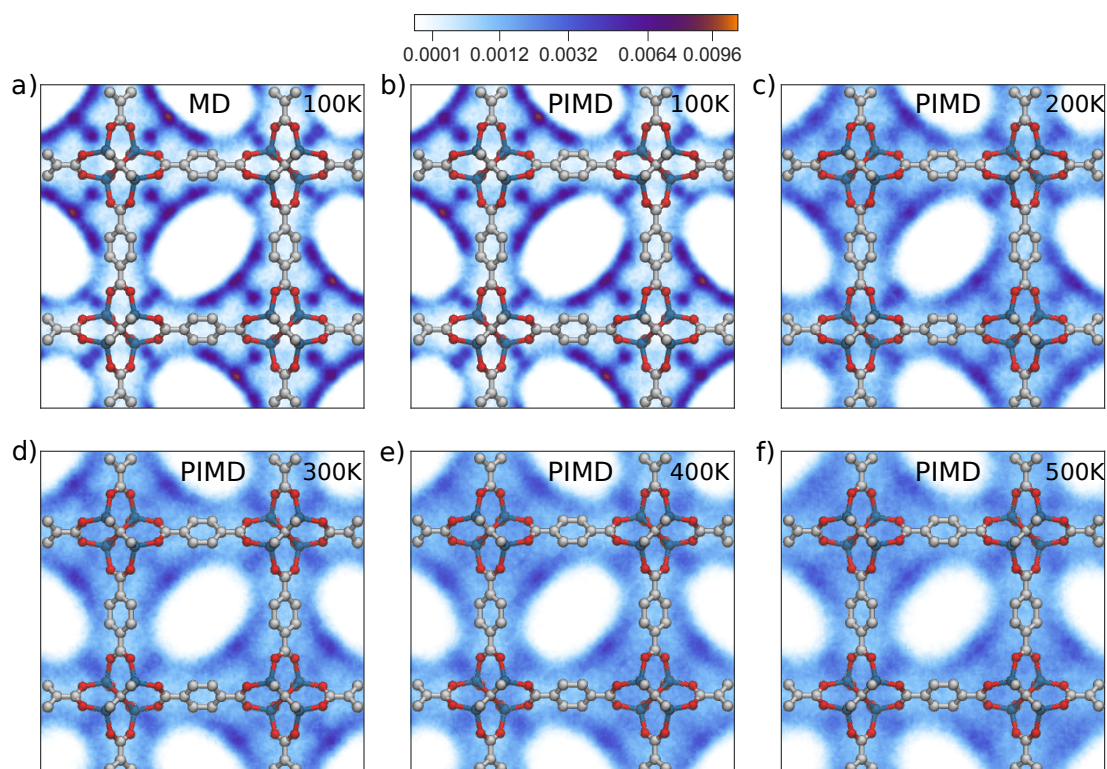


Figure 7.3 – The methane distribution in the pores of MOF-5 at different temperatures as obtained from PIMD simulations. Orange spots indicate high probability adsorption sites. Other colors show the distribution of the low probability methane positions in the conventional unit cell and represent the probability representation (from very high (orange), to high (dark blue), to low (white) probability). The results are displayed for a loading x of 100 methane molecules per conventional unit cell ($8(\text{Zn}_4\text{O}(\text{CO}_2)_6)$).

molecules and temperature is largely unaffected by NQEs, except for the zero-point lattice fluctuations. Our observations also corroborate the common practice of ignoring NQEs when studying the loading of porous materials by Grand Canonical Monte Carlo simulations^[290]. Nevertheless, this conclusion cannot be generalized to other adsorbates, especially those possessing stronger intermolecular interactions such as hydrogen bonding.

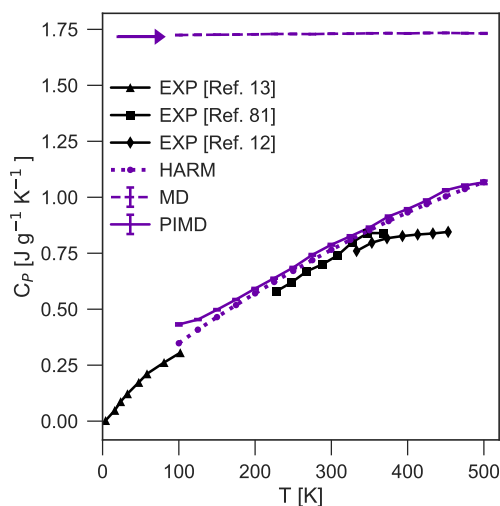


Figure 7.4 – Heat capacity C_P of the empty MOF-5 as a function of temperature (T) computed using classical MD (dashed), PIMD (solid) and the harmonic approximation (dotted). The right pointing arrow shows the Dulong-Petit limit. Different experimental results are shown in black using triangular^[262], square^[306], and diamond^[161] markers. Error bars indicate statistical uncertainty.

7.5.1 Heat Capacity

MOF-5 has been the subject of a few experimental heat capacity studies^[161,307,306,262] which have shown that the material has a low specific (or molar) heat capacity, about 0.7 J/g·K at room temperature, even when compared to other MOFs. Depending on the type of the application, a large (*e.g.* for ANG to limit temperature fluctuations) or a small (*e.g.* for CCS to limit the energy penalty) heat capacity is sought after. It is thus important to understand how this property changes at different levels of loading and temperature, and to determine the factors influencing the heat capacity, which is now possible using our high-order PIMD scheme.

In the previous section, it has been shown that classical MD can – at least qualitatively – be used to model the structural response of MOF-5 in the presence of methane at various temperatures. This approach is however expected to fail for the description of the heat capacity since the heat capacity of many systems is dominated by NQEs at room temperature, as evidenced by experimental deviations from the classical Dulong-Petit law. The most common way of including NQEs for solids is the static harmonic approximation, using Einstein or Debye harmonic model for solids, which is able to reproduce the heat capacity of many solids and will therefore also be used for comparison.

We begin by presenting the estimates of the temperature dependence of the isobaric heat capacity of the empty MOF-5 framework. As shown in Figure 7.4, the classical MD estimates (dashed line) are in agreement with the Dulong-Petit law. The simulations yield an almost constant value of $3 k_B$ per degree of freedom, which indeed results in large deviations from the experimental values.^[161,307,306,262] Upon inclusion of NQEs with our PIMD scheme (solid line), we find that the results follow the experimental measurements reasonably well up to almost 400 K. This agreement is remarkable, as these measurements are typically carried out on the as-synthesized sample, which possibly includes solvents^[161] and differs from the perfect crystal that we have simulated. Figure 7.4 also reveals that the results obtained using the

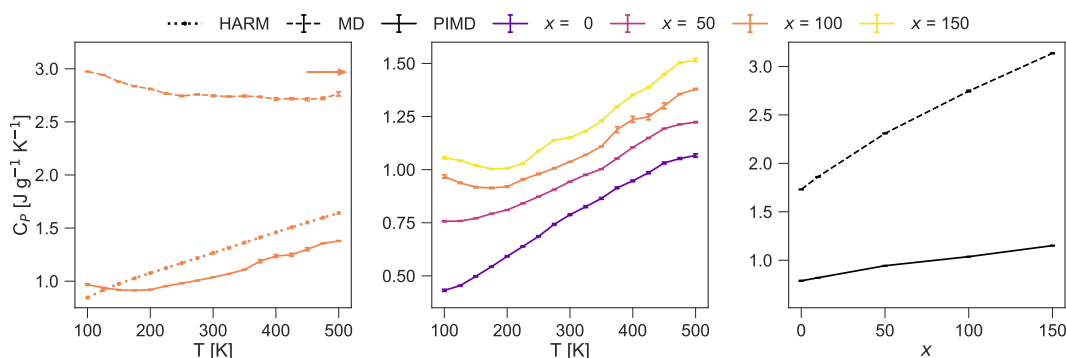


Figure 7.5 – Panel (a) shows the comparison of the classical (dashed), quantum (solid), and harmonic estimates (dotted) of the isobaric heat capacity C_p of MOF-5 with 100 molecules of methane, as a function of the temperature T . Panel (b) shows the temperature dependence of the quantum isobaric heat capacity of MOF-5 with x molecules of methane. Panel (c) shows the quantum isobaric heat capacity of the MOF with x molecules of methane as a function of x at 300 K. Error bars indicate statistical uncertainty.

simple and computationally inexpensive harmonic approximation (dotted line) are in good

agreement with the exact values computed with PIMD. This implies that anharmonic quantum contributions to the heat capacity and the effect of an adequate anharmonic sampling are small for the empty MOF.

Another notable detail of our calculations is that the harmonic approximation was used to estimate the isochoric heat capacity instead of the isobaric one. As the isobaric and isochoric heat capacities are almost the same, the MOF behaves like a regular solid, despite its large negative thermal expansion coefficient. The harmonic approximation could therefore serve as an efficient procedure to accurately estimate the heat capacity of the empty framework in the increasing number of high-throughput MOF screenings.^[259,308]

In order to study the effect of adsorbates, we start by considering the case of a loading x of 100 methane molecules per conventional unit cell ($8(\text{Zn}_4\text{O}(\text{CO}_2)_6) \cdot 100 \text{CH}_4$). Contrary to the case of the empty framework, we could not find published experimental data for the heat capacity of the guest-loaded framework. Although a high-level PIMD strategy might not be required to estimate the heat capacity of the empty MOF host, PIMD proves to be crucial to capture the correct temperature dependence of the loaded system, as can be seen in the left panel of Figure 7.5. Here, anharmonic effects become important as we observe differences between the PIMD results and the harmonic approximation. The discrepancy in the qualitative behavior of the heat capacity between both techniques can be understood through the mobility of the guest molecules in the large pores of the framework, which cannot be adequately captured by a harmonic approximation.^[309] These low-frequency anharmonic motions explain why we find at the classical MD level a similar low-temperature dependence as in PIMD, but only PIMD simulations include both anharmonic and nuclear quantum effects correctly. Interestingly, the combination of both effects yields a heat capacity that does not change monotonically, but exhibits a minimum at about 200 K.

Extending towards other loadings of methane in the middle panel of Figure 7.5, it becomes clear that the heat-capacity minimum as a function of temperature depends strongly on the number of guests and becomes more pronounced at higher loadings. Even when expressing the heat capacity normalized to the total mass of the system, one can see that at a fixed temperature C_p increases almost linearly with the loading (see Figure 7.5 (c) at 300 K).

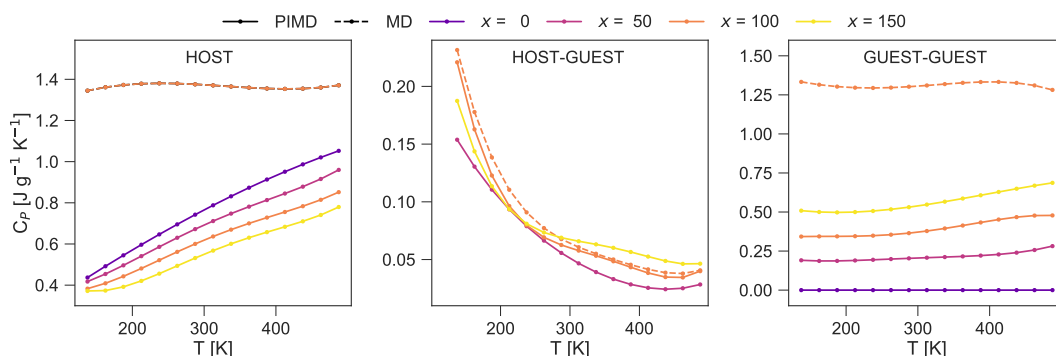


Figure 7.6 – Panels (a), (b) and (c) respectively show the decomposition of the specific heat capacity of the MOF and the adsorbate system into host, host-guest, and guest-guest contributions for different gas loadings (x). The curves were obtained by deriving a polynomial fit to the energy as a function of temperature.

To rationalize the origin of the non-monotonic temperature dependence of the heat capacity, we determine which interactions give the most substantial contribution to C_p . To this end, the force-field energy contributions are decomposed in terms of the host, host-guest, and guest-guest interactions. Figure 7.6 displays the most important results of this analysis. The host and guest-guest contributions to the specific heat capacity are visualized in panels (a) and (c). The shape of the different host curves appears to be independent of the loading. In fact, when rescaled to the mass of the empty MOF, the curves coincide with one another and with the curve obtained within the harmonic approximation. This demonstrates that the degrees of freedom of the MOF-5 framework, which are more strongly quantized, are predominantly harmonic and do not change significantly due to the interaction with methane. Their contribution to the total heat capacity per unit mass, however, decreases with the loading due to a change in the mass balance. The guest-guest interactions, on the other hand, are relatively constant and only show a small increase when going from 100 K to 500 K, due to the activation of high-frequency vibrational modes. The most interesting contribution arises from the host-guest interactions, which explains the non-monotonic behavior of the specific heat capacity of the guest-loaded system. The contribution of these interactions decreases with a sharp temperature dependence when sufficient guest molecules are present inside the pores. The large heat capacity at temperatures lower than 100 K originates from the known first-order structural phase transition of methane in MOF-5 at 60 K,^[310,311] from

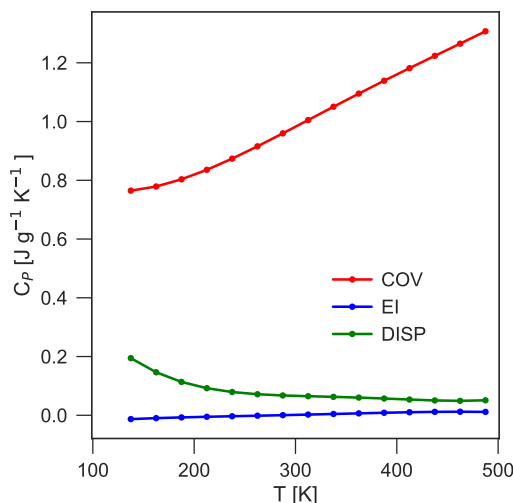


Figure 7.7 – The decomposition of the total heat capacity of MOF-5 with 100 methane molecules per unit cell into covalent (COV), electrostatic (EI), and van der Waals (DISP) contributions. The curves were obtained by deriving a polynomial fit to the energy as a function of temperature.

which we observe the decreasing tail. Since the methane molecules are more localized at low temperatures, the attractive host-guest interactions allow to efficiently store thermal energy. At higher temperatures, from 250 K to 500 K, the host-guest contributions become negligible as the confined guests become more mobile and less bound to the framework, so that the increase in thermal energy can no longer be stored in the physical interactions between the methane guests and the MOF-5 host.

Another decomposition of the force-field energy in terms of the covalent, electrostatic, and van der Waals interactions shows that the short-range covalent interactions and thus the network of chemical bonds (Figure 7.7) dominates the contributions to the heat capacity. For the empty MOF-5 framework, the noncovalent interactions are negligible. This confirms that the heat capacity of empty MOF-5 can be approximated by considering only contributions from the separate molecular fragments of the material^[161] and suggests why the harmonic approximation works well for this material. For the loaded framework, the noncovalent part starts to play a role, which is especially true for the host-guest interactions. Not surprisingly, in the case of nonpolar methane molecules, these interactions are dominated by the van der

Waals terms in the force field. This suggests that the use of different, more polar, guests in which electrostatic interactions play a more prominent role (*e.g.* CO₂) could give rise to other interesting phenomena. However, care must be taken in interpreting these different terms, as a separation is not unambiguously defined and might be force-field dependent.

7.5.2 The interplay of gas loading, anharmonicities, and quantum effects

Our analysis of the structural and thermal properties of methane-loaded MOF-5 shows that the total system does not always need a full treatment of anharmonicities and NQEs. This suggests that a full path integral sampling of the entire system may not be necessary, especially if qualitative trends are to be studied. Hence, inspired by our results, we propose an empirical formula for the volume and the heat capacity in which the most important effects, *i.e.*, anharmonicities and/or NQEs, are captured and which might prove to be beneficial for future studies of guest-loaded MOFs.

As discussed above, the main difference between the volume with or without NQEs comes from zero-point effects in the lattice.^[270] The correct volume can therefore be estimated as follows:

$$\mathcal{V} \approx \mathcal{V}_{\text{qn}}^{\text{anh}}[\text{MOF-5}] - \mathcal{V}_{\text{cl}}^{\text{anh}}[\text{MOF-5}] + \mathcal{V}_{\text{cl}}^{\text{anh}}[\text{MOF-5} + \text{CH}_4],$$

where anh stands for the inclusion of anharmonicities with MD, and cl and qn denote the use of classical or path integral MD respectively. The left most panel of Figure 7.8 indicates that this approximate volume agrees very well with the exact results obtained from PIMD simulations. A more stringent test is the thermal expansion coefficient, which is – as shown in Figure 7.8 (b) – also in excellent agreement with the PIMD results. For systems where a first-principles treatment of the potential energy surface is required and PIMD simulations are too expensive, other approximate techniques such as the quasi-harmonic approximation or classical MD with a quantum thermostat could be used to estimate the zero-point effects^[270].

In contrast, we observed in the previous section that the heat capacity of the framework could be estimated with a harmonic approximation, while the guest-host interactions are dominated

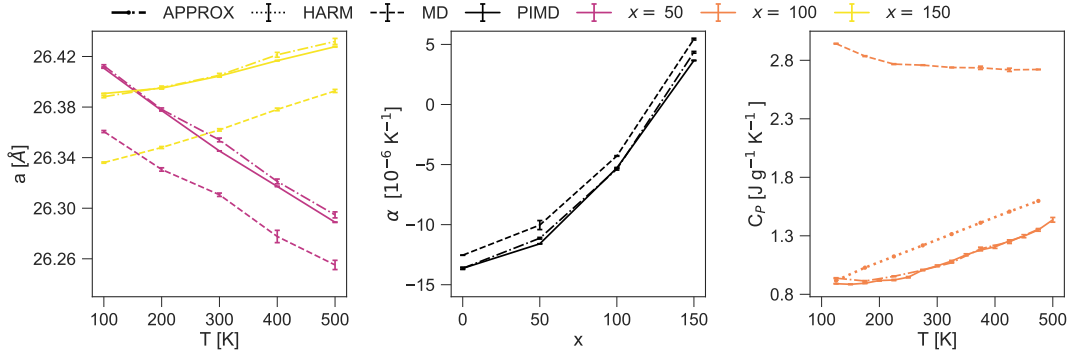


Figure 7.8 – Panels (a), (b) and (c) respectively show the temperature dependence of the cell parameter (a) for MOF-5 with 50 and 150 molecules of methane, the linear thermal expansion coefficient α of MOF-5 with x molecules of methane as function of x , and the isobaric heat capacity C_p of MOF-5 with 100 molecules of methane, obtained with classical MD (dashed), PIMD (solid), and the approximation introduced in the work (dot-dashed). Error bars indicate statistical uncertainty.

by anharmonicities. For that reason, we propose:

$$C \approx \left(C_{\text{qn}}^{\text{har}} - C_{\text{cl}}^{\text{har}} + C_{\text{cl}}^{\text{anh}} \right) [\text{MOF-5} + \text{CH}_4],$$

in which the high frequency modes of adsorbate and the MOF are treated in a harmonic fashion and the host-guest interactions are treated classically. For $C_{\text{cl}}^{\text{har}}$, the Dulong-Petit law can be used. The proposed relation is analogous to similar corrections that are routinely used to compare the heat capacity obtained from classical simulations with experimental data.^[312] As shown in the rightmost panel of Figure 7.8, this approximation works really well for this system, and is even able to qualitatively reproduce the heat capacity minimum for a loading of 100 methane molecules. Beyond 200 K, the agreement between the empirical expression and the exact PIMD becomes quantitatively correct. This method could thus be an inexpensive route to estimate the heat capacity of guest-loaded MOFs.

7.6 Discussion

To summarize, we used the high-order path integral method to calculate the isobaric thermo-physical properties of materials, that is generally applicable and therefore ideally suited to the

study of guest-loaded MOFs. The method is rigorous and can be seamlessly combined with other cost-reduction techniques, which facilitates a huge reduction of the computational cost compared to standard techniques.

We investigated the heat capacity of the prototypical MOF-5 loaded with different numbers of methane molecules. We observed that the level of statistical sampling that is needed to achieve quantitative accuracy depends on the property of interest. For all the cases we considered, we found the framework to behave in a strongly quantized manner, but to be largely amenable to a harmonic treatment. The adsorbates, on the other hand, show only mild quantum effects in their intermolecular interactions, but require a full anharmonic description. The heat capacity shows a particularly subtle interplay of quantum and anharmonic fluctuations, that results in a non-monotonic temperature dependence of the heat capacity, with a minimum around 200 K.

Through a decomposition of the heat capacity into molecular interactions, we find that the host-guest interactions are responsible for this behavior, as their contribution to the total heat capacity decreases with temperature. By comparing the behavior of different classes of framework materials and guest molecules, this may reveal new design rules to optimize the thermal behavior of a storage material over a broad range of temperatures and levels of loading. Our approach provides an affordable route to perform benchmark studies and approximation strategies to carry out the high-throughput studies that are needed to obtain a complete understanding of the interplay between framework, adsorbate, and quantum mechanical and anharmonic fluctuations that determine the thermophysical properties of MOFs.

8 Assessment of approximate methods for anharmonic free energies

The free energy is a key thermodynamic quantity which provides a measure of phase stability. Knowledge of the free energy and its derivatives with respect to temperature and applied fields can, in principle, be used to calculate every other thermodynamic observable. Reliable predictions of free energies from atomistic simulations remain a challenge because they require an accurate description of inter-atomic interactions, as well as proper treatment of the statistical mechanics of the nuclear degrees of freedom. The availability of computational resources combined with developments in electronic structure theory^[28,29,30,31,32,33,313,314,315,316] have made it possible to calculate the Born-Oppenheimer (BO) surfaces that govern nuclear motion routinely and accurately. Consequently, the accuracy of free energy calculations is often limited by the statistical sampling of the nuclear degrees of freedom^[317]. For molecules and for crystalline phases, this is most commonly performed within a harmonic approximation, which is reasonable for weakly anharmonic systems, such as metals at low temperatures^[318], but fails close to the melting temperature^[319,318] and in the presence of defects^[320]. The problem is exacerbated for the case of organic solids, which require a proper description of anharmonicity arising from quantum nuclear motion even at room temperature^[321,322,98,64]. Anharmonic and quantum effects are also often important for systems containing light elements, such as hydrogen^[323,324,325], helium^[60], water^[57], ice^[326,327,328], the record high- T_c conventional

The following chapter is partially based on the journal article V. Kapil, E. Engel, M. Rossi, and M. Ceriotti, "Assessment of Approximate Methods for Anharmonic Free Energies," *Journal of Chemical Theory and Computation*, vol. 15, pp. 5845–5857, Nov. 2019. VK was involved in the design of the research, in the development and implementation of the method, in running the calculations, in performing analysis and in the writing of the manuscript.

superconductor SH_3 ^[61], metal organic frameworks ^[62,63], as well as most molecular materials ^[64,65]. Systems of reduced effective dimensionality such as graphene ^[329,330] often exhibit stronger anharmonicity than their three-dimensional analogues.

Within the BO approximation ^[34] and given the BO potential, exact anharmonic free energies can be calculated using approaches based on imaginary time path integral (PI) simulations ^[331,70]. However, these simulations have historically been considered computationally prohibitive, when combined with an accurate electronic structure method. Consequently, a small zoo of frameworks has been developed which approximately account for quantum anharmonic motion at a much lower computational cost. These invoke different approximations and exhibit different scaling behaviour with system size.

Here, we present an extensive benchmark of the accuracy of some of the most common approximate techniques, namely, the harmonic approximation (HAR) ^[332], self-consistent phonons (SCP) ^[333,334], the independent mode framework (IMF) ^[335], and the vibrational self-consistent field (VSCF) ^[335], against reference results obtained using PI thermodynamic integration (QTI) ^[336,64]. Computationally efficient algorithms for these methods have been developed and implemented in the universal force engine *i*-PI ^[337]. The accuracy and the computational efficiency of the methods is tested on a set of solids ranging from simple allotropes of carbon, anharmonic but relatively rigid polymorphs of ice to polymorphs of paracetamol that contain (nearly) freely rotating internal degrees of freedom.

8.1 Theory

To briefly outline the different free energy methods we discuss, we consider a three-dimensional periodic system, whose minimum potential energy, “equilibrium” atomic positions form a Bravais lattice, noting that finite and aperiodic systems simply represent the limit of infinite period. The full ionic Hamiltonian of such a system is

$$\hat{H} = - \sum_{p,i} \frac{\hbar^2}{2m_i} \nabla_{\mathbf{r}_{pi}}^2 + V(\{\mathbf{r}_{pi}\}) \quad (8.1)$$

where m_i is the mass of nucleus i , V is the BO potential governing nuclear motion, and p and i run over the Bravais points and the nuclei within a unit cell, respectively. In practice we perform supercell simulations using periodic Born-von Karman simulation cell \mathbf{h} consisting of $N_a \times N_b \times N_c$ replicas of the unit cell and with cell vectors $N_a \mathbf{R}_a$, $N_b \mathbf{R}_b$, and $N_c \mathbf{R}_c$.

In the following we only consider Γ -point vibrational motion within the simulation cell. We thereby sample those \mathbf{K} -points within the first vibrational Brillouin Zone (BZ) of the underlying unit cell, for which $\exp(-i\mathbf{K} \cdot \mathbf{R}) = 1 \quad \forall \quad \mathbf{R} = n_a \mathbf{R}_a + n_b \mathbf{R}_b + n_c \mathbf{R}_c$, with $0 \leq n_x < N_x$. The Hamiltonian of the system is then uniquely defined given the positions of the N particles within the simulation cell \mathbf{h} :

$$\hat{H} = -\frac{\hbar^2}{2} \nabla_{\mathbf{r}} \mathbf{M}^{-1} \nabla_{\mathbf{r}}^T + V(\mathbf{r}, \mathbf{h}) \quad (8.2)$$

where $(\nabla_{\mathbf{r}}, \mathbf{r}) \equiv (\{\nabla_{r_1}, \dots, \nabla_{r_{3N}}\}, \{r_1, \dots, r_{3N}\})$ denotes the momenta and positions of the $3N$ degrees of freedom associated with the N particles and $\mathbf{M} = \text{Diag}[m_1, \dots, m_{3N}]$. The canonical partition function of the system at inverse temperature $\beta = (k_B T)^{-1}$ and volume $\mathcal{V} = \text{Det}[\mathbf{h}]$ is defined as

$$Z(N, \mathcal{V}, \beta) = \text{Tr}[\exp(-\beta \hat{H})], \quad (8.3)$$

where the trace can be performed over any complete basis set. In the thermodynamic limit, the Helmholtz free energy of the system is

$$A(N, \mathcal{V}, \beta) = -\beta^{-1} \ln Z(N, \mathcal{V}, \beta). \quad (8.4)$$

Direct computation of A is hindered by the computational complexity of solving the Schrödinger equation associated with the Hamiltonian \hat{H} , motivating approximate but computationally more affordable approaches.

8.1.1 Harmonic Approximation

For small displacements, $\mathbf{r} - \mathbf{r}^0$, of the particles from their equilibrium positions, $\mathbf{r}^0 \equiv \text{argmin}_{\mathbf{r}} V(\mathbf{r}, \mathbf{h})$, the potential can be Taylor expanded. Truncation after the quadratic term amounts to the

harmonic approximation

$$V^{\text{har}}(\mathbf{r}, \mathbf{h}) = V^{(0)} + \frac{1}{2}(\mathbf{r} - \mathbf{r}^0) \mathbf{K} (\mathbf{r} - \mathbf{r}^0)^T \quad (8.5)$$

with $V^{(0)} \equiv V(\mathbf{r}^0)$ and $\mathbf{K} = \nabla^2 V(\mathbf{r}, \mathbf{h})|_{\mathbf{r}=\mathbf{r}^0}$. The spectral decomposition of the Hessian,

$$\mathbf{K} = \mathbf{M}^{\frac{1}{2}} \mathbf{U} \mathbf{\Omega}^2 \mathbf{U}^T \mathbf{M}^{\frac{1}{2}} = \tilde{\mathbf{U}} \mathbf{\Omega}^2 \tilde{\mathbf{U}}^T \quad (8.6)$$

provides the unitary matrix \mathbf{U} , the mass-scaled transformation matrix $\tilde{\mathbf{U}}$, and the diagonal matrix containing the normal mode frequencies $\mathbf{\Omega} = \text{Diag}[\omega_1, \dots, \omega_{3N}]$. After transformation to the normal mode coordinates $\nabla_{\mathbf{q}} \equiv \tilde{\mathbf{U}}^T \nabla_{\mathbf{r}}$ and $\mathbf{q} \equiv \tilde{\mathbf{U}}(\mathbf{r} - \mathbf{r}^0)$, the Hamiltonian

$$\begin{aligned} \hat{H}^{\text{har}} &= -\frac{\hbar^2}{2} \nabla_{\mathbf{q}}^2 + \frac{1}{2} \mathbf{q} \mathbf{\Omega}^2 \mathbf{q}^T + V(\mathbf{r}^0) \\ &= V^{(0)} + \hat{H}^{\text{com}} + \sum_{i=1}^d \left[-\frac{\hbar^2}{2} \nabla_{q_i}^2 + \frac{1}{2} \omega_i^2 q_i^2 \right] \end{aligned} \quad (8.7)$$

separates into $V^{(0)}$, the centre of mass term \hat{H}^{com} , and a term describing a system of $d = 3N - 3$ independent simple harmonic oscillators (SHO) whose energies and wave functions for a given excitation state s_i , $E_{i,(s_i)}^{\text{har}} = (s_i + 1/2)\hbar\omega_i$ and $|\phi_i^{(s_i)}\rangle$, are known analytically. In finite systems, global rotations decouple analogously. The centre of mass contribution to the free energy A^{com} is that of a free particle in a three dimensional box with a volume and shape equal to that of the Wigner-Seitz cell of the system, while the contribution from the free rotations of finite systems can be computed within the rigid rotor approximation^[338].

For the remaining system of harmonic oscillators, the d -body wave function of the global state described by the d -tuple $\mathbf{s} = (s_1, \dots, s_d)$ is a Hartree product of the independent normal mode wave functions:

$$|\Psi_{(\mathbf{s})}^{\text{har}}\rangle = \prod_{i=1}^d \otimes |\phi_{i,(s_i)}\rangle \quad (8.8)$$

and the free energy is:

$$A^{\text{har}}(N, \mathcal{V}, \beta) = V^{(0)} + A^{\text{com}} + \sum_{i=1}^d \left[\frac{\hbar \omega_i}{2} + \beta^{-1} \ln \left(1 - e^{-\beta \hbar \omega_i} \right) \right]. \quad (8.9)$$

8.1.2 Independent Mode Framework

A first approximation to anharmonic quantum nuclear motion is detailed in the work of Monserrat *et al.* ^[335]. The potential is expanded in terms of the normal mode coordinates

$$V(\mathbf{q}) = V^0 + \sum_i^d V^{(1)}(q_i) + \frac{1}{2} \sum_i^d \sum_{j \neq i}^d V^{(2)}(q_i, q_j) + \dots, \quad (8.10)$$

where

$$V^{(1)}(q_i) = V(0, \dots, q_i, \dots, 0) - V^{(0)}, \quad (8.11)$$

is the (anharmonic) independent mode term and

$$\begin{aligned} V^{(2)}(q_i, q_j) = & V(0, \dots, q_i, \dots, q_j, \dots, 0) \\ & - V^{(1)}(q_i) - V^{(1)}(q_j) - V^{(0)}. \end{aligned} \quad (8.12)$$

describes pairwise coupling between normal modes. This expansion can be continued for more general n -body terms $V^{(n)}$. Since one starts with the harmonic approximation, in which the normal modes are non-interacting, the hope is that higher-order terms decrease in size with increasing n . The validity of this assumption is discussed in section 8.3. Truncation after $V^{(1)}$ amounts to the independent mode approximation with the Hamiltonian

$$\hat{H}^{\text{imf}} = V^{(0)} + \sum_i^d \left[-\frac{\hbar^2}{2} \nabla_{q_i}^2 + V^{(1)}(q_i) \right]. \quad (8.13)$$

Despite the presence of anharmonicity the normal modes remain independent. A Hartree product analogous to Eq. (8.8) of anharmonic normal mode wave functions solves the Schrödinger

equation yielding the eigenvalues $E_{i,(s_i)}^{\text{imf}}$. The Helmholtz free energy is

$$A^{\text{imf}}(N, \mathcal{V}, \beta) = V^{(0)} + A^{\text{com}} - \sum_{i=1}^d \left[\beta^{-1} \ln \sum_{s_i} \exp \left(-\beta E_{i,(s_i)}^{\text{imf}} \right) \right]. \quad (8.14)$$

8.1.3 Vibrational Self-consistent Field

Retaining terms involving $V^{(2)}$ (and/or higher order terms) leads to coupling of the previously independent normal modes and complicates the solution of the Schrödinger equation. Monserrat *et al.* ^[335] solve the equation

$$\left[-\frac{\hbar^2}{2} \nabla_{\mathbf{q}}^2 + V(\mathbf{q}) \right] \left| \Psi_{(\mathbf{s})}^{\text{vscf}} \right\rangle = E_{(\mathbf{s})}^{\text{vscf}} \left| \Psi_{(\mathbf{s})}^{\text{vscf}} \right\rangle \quad (8.15)$$

within the iterative Vibrational Self-Consistent Field (VSCF) approach, where $V(\mathbf{q})$ represents the truncated form of Eq. (8.10). Using a Hartree product trial wavefunction amounts to a mean-field (MF) treatment and leads to the VSCF equations

$$\left[-\frac{\hbar^2}{2} \nabla_{q_i}^2 + \bar{V}_i(q_i) \right] \left| \psi_{i,(s_i)}^{\text{vscf}} \right\rangle = E_i^{\text{vscf}} \left| \psi_{i,(s_i)}^{\text{vscf}} \right\rangle \quad (8.16)$$

where $\bar{V}_i(q_i)$ is the mean-field potential experienced by normal mode i ,

$$\bar{V}_i(q_i) = \sum_{j \neq i} \rho(q_j) V(\mathbf{q}) \quad (8.17)$$

with

$$\rho(q_j) = \frac{\sum_{s_j} \exp \left(-\beta E_{j,(s_j)}^{\text{vscf}} \right) \left| \psi_{j,(s_j)}^{\text{vscf}} \right\rangle \left\langle \psi_{j,(s_j)}^{\text{vscf}} \right|}{\sum_{s_j} \exp \left(-\beta E_{j,(s_j)}^{\text{vscf}} \right)}. \quad (8.18)$$

To lowest order the VSCF free energy becomes

$$A^{\text{vscf}}(N, \mathcal{V}, \beta) = V^{(0)} + A^{\text{com}} - \beta^{-1} \ln \sum_{\mathbf{s}} \exp \left(-\beta \sum_i E_{i,(s_i)}^{\text{vscf}} \right). \quad (8.19)$$

A perturbation theory can be constructed in terms of the (assumed to be small) difference between the mapped out and the MF potential, $V(\mathbf{q}) - \sum_i \bar{V}_i(q_i)$, leading to a second-order

MP2 correction to the energy of state \mathbf{s} given by

$$E_{(\mathbf{s})}^{\text{vscf},(2)} = \sum_{\mathbf{s}' \neq \mathbf{s}} \frac{\left\langle \Psi_{(\mathbf{s}')}^{\text{vscf}} \left| V(\mathbf{q}) - \sum_i \tilde{V}_i(q_i) \right| \Psi_{(\mathbf{s})}^{\text{vscf}} \right\rangle^2}{E_{\mathbf{s}}^{\text{vscf},(1)} - E_{\mathbf{s}'}^{\text{vscf},(1)}} \quad (8.20)$$

and to the approximate free energy

$$A^{\text{mp2}}(N, \mathcal{V}, \beta) = A^{\text{vscf}}(N, \mathcal{V}, \beta) - \beta^{-1} \ln \sum_{\mathbf{s}} \exp \left(-\beta E_{\mathbf{s}}^{\text{vscf},(2)} \right). \quad (8.21)$$

In all the examples we discuss below, we never consider terms beyond $V^{(2)}$ in Eq. (8.10), similarly to what was done by Monserrat *et al.*. In all cases we considered, the MP2 correction is very small and therefore we decided not to include it. Thus, every time we refer to IMF in what follows, we imply that we only considered $V^{(1)}$, and whenever we refer to VSCF we imply the Hartree energy computed on the potential of Eq. (8.10) including terms up to $V^{(2)}$.

8.1.4 Self-consistent Phonons

Another way of calculating an anharmonic correction to the harmonic approximation exploits the Gibbs-Bogoliubov inequality^[339], which states that the true free energy of a system is always bounded from above by the free energy A^{scp} computed using a trial density matrix, $\hat{\rho}^{\text{scp}}$:

$$A < A^{\text{scp}} = \langle \hat{H} + \beta^{-1} \ln \hat{\rho}^{\text{scp}} \rangle_{\hat{H}^{\text{scp}}} \quad ; \quad \hat{\rho}^{\text{scp}} = \frac{\exp(-\beta \hat{H}^{\text{scp}})}{\text{Tr}[\exp(-\beta \hat{H}^{\text{scp}})]}, \quad (8.22)$$

where $\langle \square \rangle_{\hat{H}^{\text{scp}}} = \text{Tr}[\rho^{\text{scp}} \square]$ is an ensemble average defined by the trial density matrix $\hat{\rho}^{\text{scp}}$. Within the self-consistent phonons method^[340,341], $\hat{\rho}^{\text{scp}}$ is the density matrix of a harmonic Hamiltonian with Hessian \mathbf{K}^{scp} and equilibrium positions \mathbf{r}^{scp} ,

$$\hat{\rho}^{\text{scp}}(\mathbf{r}) = (2\pi\mathbf{D})^{-\frac{1}{2}} \exp \left(-\frac{1}{2}(\mathbf{r} - \mathbf{r}^{\text{scp}})\mathbf{D}^{-1}(\mathbf{r} - \mathbf{r}^{\text{scp}})^T \right), \quad (8.23)$$

where $\mathbf{D} = \mathbf{M}^{-\frac{1}{2}} \mathbf{U} \mathbf{Q}^2 \mathbf{U}^T \mathbf{M}^{-\frac{1}{2}}$ and \mathbf{Q} is a diagonal matrix containing the root-mean-square (RMS) displacements^[333]

$$\tilde{q}_i^{\text{scp}}(T) = \sqrt{\frac{\hbar}{2\omega_i^{\text{scp}}} \coth \frac{\hbar\omega_i^{\text{scp}}}{2k_B T}} \quad (8.24)$$

of the normal modes. The lowest upper bound to the true free energy is obtained by minimizing the free energy with respect to \mathbf{r}^{scp} and \mathbf{K}^{scp} . This leads to the steady state conditions^[333]

$$\begin{aligned} \langle \mathbf{f}_r(\mathbf{r}) \rangle_{H^{\text{scp}}} &= 0 \\ \langle \mathbf{K}(\mathbf{r}) \rangle_{H^{\text{scp}}} &= \mathbf{K}^{\text{scp}} \end{aligned} \quad (8.25)$$

where $\mathbf{f}_r(\mathbf{r})$ correspond to the forces of the potential $V(\mathbf{r})$. The solution is obtained in a self consistent manner by starting with educated guesses of $(\mathbf{r}^{\text{scp}}, \mathbf{K}^{\text{scp}}) = (\mathbf{r}_0^{\text{scp}}, \mathbf{K}_0^{\text{scp}})$, which are in practice chosen to be those obtained within the harmonic approximation, and updating

$$\begin{aligned} \mathbf{K}_{\hat{H}_{l+1}^{\text{scp}}}^{\text{scp}} &= \langle \mathbf{K}(\mathbf{r}) \rangle_{\hat{H}_l^{\text{scp}}}, \\ \mathbf{r}_{\hat{H}_{l+1}^{\text{scp}}}^{\text{scp}} &= \mathbf{r}_{\hat{H}_l^{\text{scp}}}^{\text{scp}} + \mathbf{K}_{\hat{H}_{l+1}^{\text{scp}}}^{\text{scp}^{-1}} \langle \mathbf{f}_r \rangle_{\hat{H}_l^{\text{scp}}}, \end{aligned} \quad (8.26)$$

until convergence is achieved. Here \hat{H}_l^{scp} denotes the the trial Hamiltonian of the l -th SCP iteration. The resultant free energy at the l -th iteration is calculated as

$$\begin{aligned} A^{\text{scp}} &= A^{\text{com}} + \left[\frac{\hbar\omega_i^{\text{scp},l}}{2} + \beta^{-1} \ln \left(1 - e^{-\beta\hbar\omega_i^{\text{scp},l}} \right) \right] \\ &+ \left\langle V(\mathbf{r}) - \frac{1}{2} (\mathbf{r} - \mathbf{r}_l^{\text{scp}}) \mathbf{K}_l^{\text{scp}} (\mathbf{r} - \mathbf{r}_l^{\text{scp}})^T \right\rangle_{\hat{H}_l^{\text{scp}}}. \end{aligned} \quad (8.27)$$

8.1.5 Thermodynamic Integration

Within the thermodynamic integration scheme the *exact*¹ free energy difference between two states is calculated as the work to reversibly transform one state into the other^[342,343,344]. For solids this method can be used to calculate the classical anharmonic correction to the

¹The free energy obtained from path-integral based thermodynamic integration will be exact for a given interatomic potential and within the statistical accuracy afforded by the length of the simulations.

harmonic Helmholtz free energy as the reversible work done while “switching on” the anharmonic part of the potential^[345,346]. In the Hamiltonian $H^\lambda = (1 - \lambda)H^{\text{har}} + \lambda H$ the Kirkwood coupling parameter λ smoothly switches the potential from harmonic ($\lambda = 0$) to fully anharmonic ($\lambda = 1$). The free energy difference at a given temperature is obtained by computing the integral of the thermodynamic force along the switching path:

$$\Delta A_{\text{cl}} = A_{\text{cl}} - A_{\text{cl}}^{\text{har}} = \int_0^1 d\lambda \left(\frac{\partial A}{\partial \lambda} \right) = \int_0^1 d\lambda \langle V - V^{\text{har}} \rangle_{H^\lambda}, \quad (8.28)$$

where, $\langle \square \rangle_{H^\lambda}$ represents an average over the classical canonical ensemble sampled by the intermediate Hamiltonian and $A_{\text{cl}}^{\text{har}}$ is the classical harmonic free energy. Setting aside statistical errors, ΔA_{cl} can be computed exactly by sampling the thermodynamic forces at multiple values of $\lambda \in [0, 1]$ using molecular dynamics simulations in the canonical ensemble. Depending on computational convenience, ΔA_{cl} can also be calculated by alternative TI paths, that involve computing the anharmonic free energy at a low temperature T_0 , and then the change in free energy between T_0 and the desired temperature^[347,344],

$$\frac{A(N, \mathcal{V}, (k_B T)^{-1})}{k_B T} = \frac{A(N, \mathcal{V}, (k_B T_0)^{-1})}{k_B T_0} - \int_{T_0}^T \frac{\langle V \rangle_{\tilde{T}} + \frac{3N}{2} k_B \tilde{T}}{k_B \tilde{T}^2} d\tilde{T}, \quad (8.29)$$

where $\langle \square \rangle_{\tilde{T}}$ is an average over the classical $N\mathcal{V}\tilde{T}$ ensemble. A similar expression can be used to compute the full value of ΔA_{cl} , by taking the $T_0 \rightarrow 0$ limit of Eq. (8.29)^[348]

$$\Delta A_{\text{cl}} = -T \int_0^T d\tilde{T} \frac{\langle V - V^{(0)} - \frac{3N}{2} k_B \tilde{T} \rangle_{\tilde{T}}}{\tilde{T}^2}. \quad (8.30)$$

To include quantum anharmonic corrections due to zero-point energy, tunnelling, etc., a second thermodynamic integration must be performed to calculate the work required to reversibly transform the particles from classical to quantum^[349,350]. This can be achieved by defining the Hamiltonian $\hat{H}^g = -g \frac{\hbar^2}{2} \nabla_{\mathbf{r}} \mathbf{M}^{-1} \nabla_{\mathbf{r}}^T + V(\mathbf{r}, \mathbf{h})$, where g scales the mass of the particles^[336,322,99]. As g is varied from 1 to 0 (i.e. the limit of infinite mass), the de Broglie wavelength of the particles smoothly drops from its physical value to zero, yielding the desired transformation from quantum to classical particles. The corresponding free energy difference

is:

$$\Delta A_{\text{qn}} = A_{\text{qn}} - A_{\text{cl}} = \int_0^1 dg \, g^{-1} \langle \hat{T} - T_{\text{cl}} \rangle_{\hat{H}^g}. \quad (8.31)$$

where $\langle \hat{T} \rangle_{\hat{H}^g}$ represents the average quantum kinetic energy for the intermediate Hamiltonian and T_{cl} is the classical kinetic energy which is independent of the mass of the system. Eq. (8.31) can be computed exactly (modulo statistical error) by sampling the quantum canonical ensembles for $g \in [0, 1]$ using PI molecular dynamics (PIMD). The difference between the classical and quantum kinetic energy can be computed directly using a centroid-virial kinetic energy estimator^[351,60]. The total anharmonic free energy is computed as:

$$A = A^{\text{com}} + A_{\text{cl}}^{\text{har}} + \Delta A_{\text{cl}} + \Delta A_{\text{qn}}. \quad (8.32)$$

For a given PES, Eq. 8.32 gives the exact anharmonic free energy (aside from statistical error) provided that a sufficient number of intermediate states are used to perform the classical and quantum thermodynamic integrations.

8.2 Implementation

In order to perform a direct comparison of the different approximate methods, IME, VSCE, and SCP were implemented within i-PI^[337]. i-PI is an open-source python package for atomistic simulations which collects energies and forces computed from one of the many of density-functional-theory (DFT), empirical and machine-learning potential codes it is interfaced to. The IME, VSCE, and SCP implementations are schematically shown in Figs. 8.1 to 8.3. The reference free energies can be evaluated by post-processing (PI)MD simulations and required no new dedicated implementation, since i-PI is able to collect energy and forces from different codes, and to combine them to realize simulations with mixed and weighted potentials.

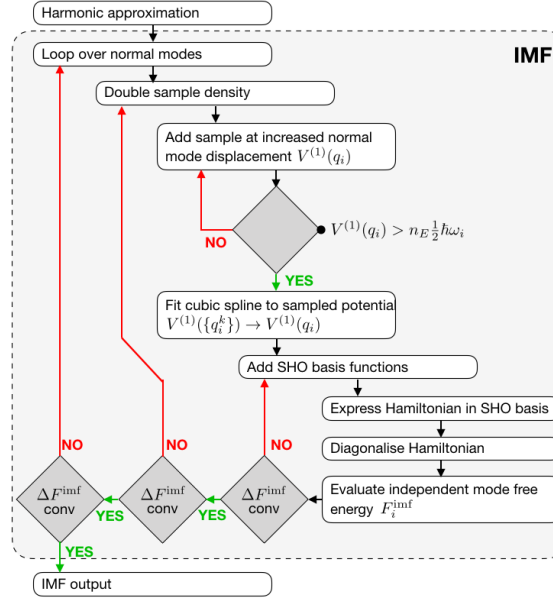


Figure 8.1 – Schematic representation of the independent mode approximation module. See the main text for a detailed discussion of the implementation.

8.2.1 Independent Mode Framework

Given the transformation matrix, $\tilde{\mathbf{U}}$, and normal-mode frequencies from the harmonic approximation, we perform single-point energy and force evaluations for equally spaced configurations $q_i^j = j f \tilde{q}_i(T)$ along each normal mode i where \tilde{q}_i is the RMS displacement of the normal mode at a target temperature T . We increase j by one at a time until the sampled energy $V^{(1)}(q_i^j)$ exceeds a user-defined multiple n_E of the thermal harmonic energy $V^{(1)}(q_i^j) > n_E E_i^{\text{har}}(T)$. This ensures that the potential is always mapped out far enough into the classically forbidden region (but only as far as necessary) to localize the nuclear density, at temperatures lower or equal than the chosen target. The independent mode potential $\sum_i V^{(1)}(q_i)$ is then reconstructed by fitting cubic splines to $\{(q_i^j, V^{(1)}(q_i^j))\}$. The corresponding independent mode Hamiltonian is expanded in a basis of SHO eigenstates and diagonalized to evaluate the independent mode anharmonic Helmholtz free energy. The Helmholtz free energy is converged with respect to the density of the frozen-phonon samples q_i^j by repeatedly halving f and supplementing the already collected $\{(q_i^j, V^{(1)}(q_i^j))\}$ with corresponding samples, until the required convergence threshold is met. For each f the Helmholtz free energy is converged with respect to the size of the SHO basis.

8.2.2 Vibrational Self-Consistent Field

The implementation of the VSCF framework is split into two modules: one for mapping the potential energy surface (PES), and one for solving the VSCF problem. The mapping strategy mirrors that employed in the IMF module. In a first loop over normal modes, we collect $\{(jf \tilde{q}_i(T), V^{(1)}(jf \tilde{q}_i(T)))\}$ until the sampled potential exceeds a user-defined multiple of the harmonic energy $V^{(1)}(jf \tilde{q}_i(T)) > n_E E_i^{\text{har}}(T)$, thereby also determining the sampling range for the coupling corrections. In a second loop over n -tuples of normal modes, we then sample $\{(q_{i_1}^{j_1}, \dots, q_{i_n}^{j_n}), V^{(n)}(q_{i_1}^{j_1}, \dots, q_{i_n}^{j_n})\}$ in a similar fashion and extract the coupling corrections $V^{(n)}(q_{i_1}, \dots, q_{i_n})$ using cubic spline fits. Currently, sampling and fitting of $n = 2, 3$ are implemented. The extracted coupling corrections are stored for use within the VSCF solver module.

The module for solving the VSCF problem consists of two submodules, the first of which performs the VSCF calculation itself. The thermal density determining the mean-field potentials $\{\bar{V}_i(q_i)\}$ is initialised as the IMF thermal density. Within a VSCF step the MF independent mode Hamiltonians for the given MF potentials are constructed, expanded in an SHO basis, and diagonalized to determine the updated MF thermal vibrational density and the free energy A^{vscf} . To stabilize the VSCF convergence, 50% of the thermal density resulting from the previous VSCF iteration are mixed in before the mean-field potentials $\{\bar{V}_i(q_i)\}$ are updated and the next VSCF step is initiated. This is repeated until self-consistency has been reached as indicated by convergence of the associated free energy A^{vscf} to within the required threshold.

The second submodule allows the calculation of an MP2 correction on top of the MF eigenstates and -energies by looping over pairs of eigenstates $(\mathbf{s}, \mathbf{s}')$ to evaluate the MP2 corrections in Eq. (8.20) on a real-space grid of predefined density. We only consider eigenstates of the self-consistent MF description with eigenenergies $E_{\mathbf{s}}^{\text{vscf}}$ within a set multiple of $k_B T$.

8.2.3 Self-Consistent Phonons

Our implementation of the SCP method is schematically shown in Fig. 8.3. In the first loop over SCP steps, we construct the trial density matrix $\hat{\rho}^{\text{scp}}(\mathbf{r})$ using the mean position \mathbf{r}^{scp} and Hessian \mathbf{K}^{scp} obtained from the previous step. In the first step $\hat{\rho}^{\text{scp}}(\mathbf{r})$ is built based on

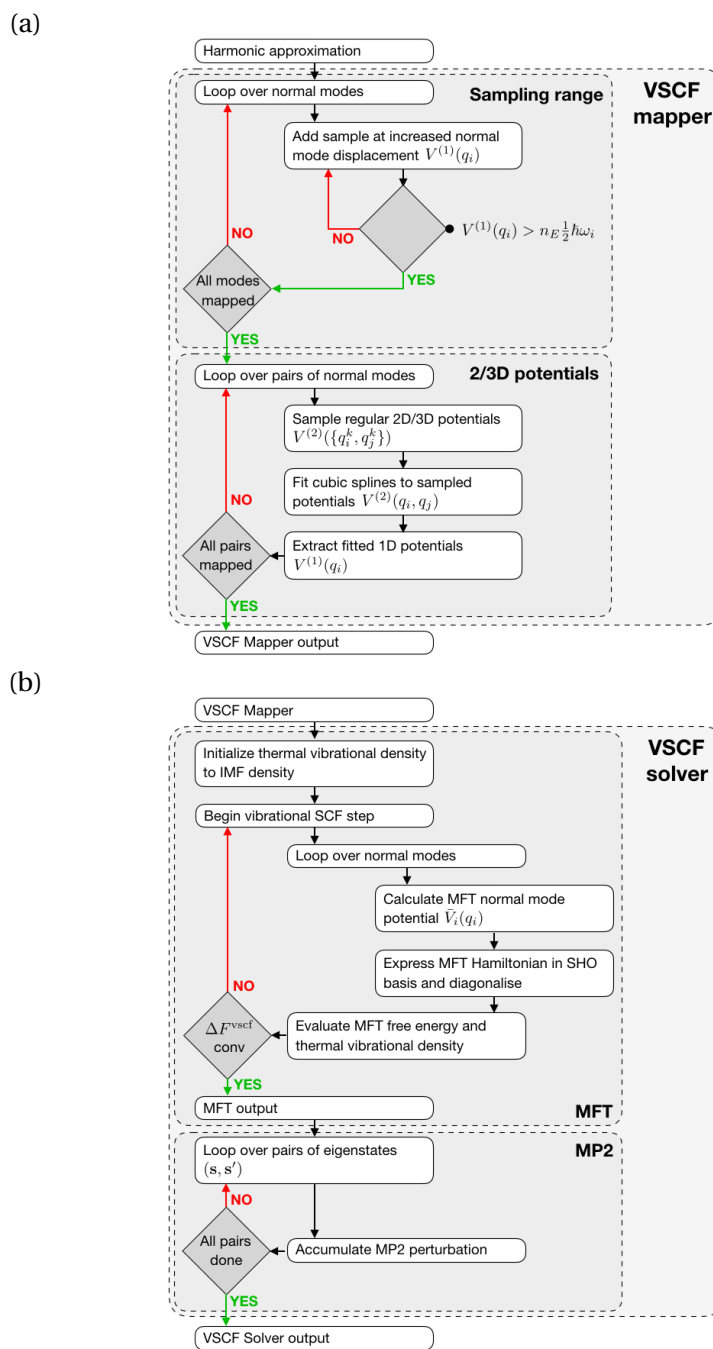


Figure 8.2 – Schematic representation of the (a) VSCF mapper and (b) VSCF solver module. See the main text for a detailed discussion of the implementation

the equilibrium geometry and the harmonic Hessian. In the second loop we calculate the ensemble averages of the forces and the Hessian, necessary to perform the optimization steps described in Eq. 8.26. These are realized as Gaussian integrals and computed using MC importance sampling as

$$\langle \square \rangle_{H^{\text{scp}}} = \frac{1}{N_s} \sum_{i=1}^{N_s} \square + \mathcal{O}\left(1/\sqrt{N_s}\right) \quad (8.33)$$

where N_s is the number of samples. Samples are generated by translating $3(N-1)$ -tuples of (quasi-)random numbers on the interval $[0, 1]$ into atomic displacements from the mean position \mathbf{r}^{scp} using the inverse cumulative distribution function of $\rho^{\text{scp}}(\mathbf{r})$ with a Beasley-Springer-Moro algorithm^[352,353]. To speed up the convergence of the averages with respect to the number of samples, we employ the following tricks:

1. For small system sizes, instead of drawing pseudo random numbers, we use low-discrepancy quasi-random numbers – specifically Sobol sequences^[354] – as was done in the implementation of Brown and coworkers^[333]. This leads to a more uniform sampling, so that error in equation 8.33 decays as $\mathcal{O}(\ln(N_s)^d/N_s)$ which becomes $\sim \mathcal{O}(1/N)$ for low dimensional integrals^[355]. For large system sizes, we resort to pseudo random numbers as the performance of Sobol sequences degrades^[356]. We use the FORTRAN implementation of Burkardt^[357] to generate Sobol sequences.
2. As was done in the implementation of Errea and co-workers^[358], we re-use samples from previous SCP iterations via a reweighting scheme. Given the updated trial density $\hat{\rho}_l^{\text{scp}}$ at the l -th SCF iteration, the reweighted average using the N_s samples $\{\mathbf{r}\}^k$ drawn from the trial density $\hat{\rho}_k^{\text{scp}}$ at the k -th SCP iteration is

$$\begin{aligned} \langle \square \rangle_{\hat{H}_l^{\text{scp}}}^k &= \frac{1}{N_s} \sum_{\mathbf{r} \in \{\mathbf{r}\}^k} \left[w_l^k(\mathbf{r}) \square(\mathbf{r}) \right] \\ w_l^k(\mathbf{r}) &\equiv \rho_l^{\text{scp}}(\mathbf{r}) / \rho_k^{\text{scp}}(\mathbf{r}) \end{aligned} \quad (8.34)$$

We minimize the error in the global estimates $\langle \square \rangle_l$ at the l -th SCP iteration by weighting

samples drawn in the k -th SCP iteration according to a “batch weight”

$$\begin{aligned}\langle \square \rangle_{\hat{H}_l^{\text{scp}}} &\approx \sum_{k=1}^l \left[W^k \langle \square \rangle_l^k \right] \\ W^k &= \text{Var}(\square^k)^{-1} \left[\sum_{k'} \text{Var}(\square^{k'})^{-1} \right]^{-1}\end{aligned}\quad (8.35)$$

where $\square^k \equiv \square(\{\mathbf{r}\}^k)$ and the variance of a generic observable over samples from k -th SCP iteration is^[359]

$$\text{Var}(\square^k) = \left(\text{Var}(\square) + \langle -\square \ln w_l^k \rangle \right) \frac{\exp(\text{Var}(-\ln w_l^k))}{N_s} \quad (8.36)$$

where $w_l^k \equiv w_l^k(\{\mathbf{r}\}^k)$, provided both \square and $-\ln w_l^k$ are normally distributed. Neglecting $\langle -\square \ln w_l^k \rangle$ renders the batch weights independent of the observable being considered

$$W_l^k = \frac{\exp(-\text{Var}(\ln w_l^k(\{\mathbf{r}\}^k)))}{\sum_m \exp(-\text{Var}(\ln w_l^m(\{\mathbf{r}\}^m)))} \quad (8.37)$$

and thereby also suitable for both Hessians and forces.

3. Taking inspiration from stochastic over-relaxation algorithms^[360,361], we always draw pairs of configurations $(\mathbf{q}^i, \mathbf{q}^{i+1})$, where $\mathbf{q}^{i+1} = -\mathbf{q}^i$, ensuring that forces from the symmetric part of V cancel out exactly.
4. To compute the average Hessian we use integration by parts, as suggested in the Ref. 333, and to further reduce the variance, we express it in terms of the difference between the harmonic and the anharmonic forces:

$$\begin{aligned}\langle \mathbf{K}(\mathbf{r}) \rangle_{\hat{H}_l^{\text{scp}}}^k &= \mathbf{K}_l^{\text{scp}} - \mathbf{D}^{-1} \left\langle [\mathbf{r} - \mathbf{r}_l^{\text{scp}}]^T [\mathbf{f}_{\mathbf{r}}(\mathbf{r}) - \mathbf{f}_{\mathbf{r}}^{\text{scp}}(\mathbf{r})] \right\rangle_{\hat{H}_l^{\text{scp}}}^k \\ \mathbf{f}_{\mathbf{r}}^{\text{scp}}(\mathbf{r}) &\equiv -\mathbf{K}_l^{\text{scp}}(\mathbf{r} - \mathbf{r}_l^{\text{scp}})\end{aligned}\quad (8.38)$$

Samples are drawn in sets of N_s until at least one component of the average forces (in terms of normal mode coordinates) is statistically significant, as assessed by whether the average over samples is larger than the standard deviation.

Direct application of Eq. 8.26 in Cartesian space may lead to instability due to the residual

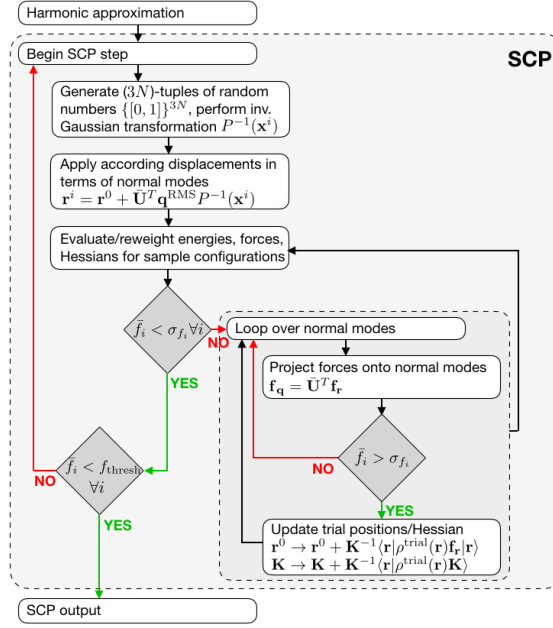


Figure 8.3 – Schematic representation of the SCP module. See the main text for a detailed discussion of the implementation

statistical errors. Therefore, we only update \mathbf{r}^{SCP} along those normal modes which exhibit a statistically significant net force, $\langle \mathbf{f}_{\mathbf{q}_i} \rangle$, where $\langle \mathbf{f}_{\mathbf{q}} \rangle = \tilde{\mathbf{U}}^T \langle \mathbf{f}_{\mathbf{r}}(\mathbf{r}) \rangle$. The optimization continues until no statistically significant force component remain or the “batch weights” become smaller than a preset threshold, at which point a new SCP iteration begins.

While all modes must be real upon convergence, insufficient statistics may lead to spurious imaginary modes with $\omega_i^2 < 0$ before convergence is achieved. Such imaginary modes are treated by setting $\omega_i^2 = -\omega_i^2$ in the effective harmonic description. It is also possible to define a lower threshold for the permissible value of ω_i , although it was not necessary to apply this threshold for any of the examples discussed here.

8.3 Results

We first describe the systems that have been studied and the potentials that have been used to compute inter-particle interactions. We then investigate the scaling of the computational cost of the methods with respect to the system size, before assessing their accuracy by systematically comparing the approximate free energies to reference quantum thermodynamic

integrations. We neglect the centre of mass contribution to the free energy throughout as it cancels out when comparing methods. We converge all results with respect to the sampling of the vibrational BZ by increasing the simulation cell size, allowing us to compare the different methods in equivalent and physically meaningful conditions.

8.3.1 Systems and Computational Details

Three sets of materials are studied in this work in order of increasing complexity. As a first example we consider the diamond^[362] and lonsdaleite^[363] allotropes of carbon. These differ only in the stacking of hexagonal bilayers of tetrahedrally-coordinated carbon atoms. Their room temperature densities are identical to within experimental error^[362,363], and equal to 3.51g/cm^3 . We consider simulation cells containing up to 64 atoms, starting from the two- and four-atom primitive cells for diamond and lonsdaleite respectively. All the cells were designed to be as close to cubic as possible to render the effective sampling of the vibrational BZ as uniform as possible. Inter-atomic interactions are modelled using the Gaussian Approximation Potential (GAP) of Deringer and Csányi^[364], which is based on LDA DFT calculations on configurations from MD simulations of liquid and amorphous carbon. For crystalline carbon (including diamond and graphite) it has been shown to reproduce DFT energies and forces to within RMS errors of 2 meV/atom and 0.1 eV/Angstrom, respectively.

As a second benchmark, we discuss two proton-ordered polymorphs of ice, hexagonal (XIh)^[365,366] and cubic (XIc)^[367] ice. These become thermodynamically (meta-)stable below the experimental transition temperature for proton-disordering of 72 K^[365,366]. For XIc we assume $I4_1/amd$ symmetry, noting that the true experimental structure of XIc is still under debate^[367]. In direct analogy to the above carbon allotropes the oxygen sublattices of XIh and XIc only differ in the stacking of bilayers of tetrahedrally-coordinated oxygen atoms. In view of the absence of experimental data for sufficiently pure XIc we take its density to be identical to that of XIh. We use the experimental density at ambient pressure and 10 K of 0.93g/cm^3 , noting that the thermal expansion of ice XIh between 10 K and 70 K is less than 0.5 %. We use simulation cells containing up to 16 molecules to allow for the possibility of coupling between pseudo-translations, which are not present at the Γ -point of the unit cell, and librational, bending and O–H bond stretching modes. The interatomic interactions are described using a Behler-

Parinello type neural network (NN)^[368], based on B3LYP+D3 DFT reference calculations for around 20,000 liquid water and hexagonal ice configurations from MD and PIMD trajectories. This potential successfully reproduces the density of states, pair correlation functions and energy fluctuations of B3LYP+D3 liquid water^[369] and has been used to study the quantum kinetic energy, proton momentum distribution, and vibrational density of states of solid and liquid water^[370,371].

Finally we analyze two polymorphs of paracetamol (N-acetyl-p-aminophenol), the monoclinic form I^[372] and the orthorhombic form II^[373]. The two forms differ in the packing of hydrogen bonded sheets of molecules – zig-zag for form I and almost planar for form II. We consider the conventional unit cells containing four and eight formula units for forms I and II respectively, at room temperature experimental densities. Inter-atomic interactions are described on the basis of the Merck Molecular Force Field also used in Ref.^[64]. The accuracy parameter of the PPPM method^[374] used for calculating electrostatic interactions was set to 10^{-6} fractional error in the individual force components, which is smaller than the value usually required, to ensure a smooth PES. While this simple potential contains harmonic terms for bonds and angles, it remains highly anharmonic as the dihedral interaction term describes a (almost) free rotation of the methyl groups at room temperature.

8.3.2 Computational cost

We define the computational cost of an approximate method (\square) as the minimum number of energy or force evaluations required to converge the per atom anharmonic free energy ($A^\square - A^{\text{har}}$) to within 10% of the reference value for the largest system size. For the case of diamond, this tolerance is equal to a stringent 0.2 meV/atom. Fig. 8.4 shows how the cost of these methods and the reference QTI scales with the number of atoms in the simulation cell.

The cost of IMF calculations depends linearly on the number of normal modes – which scales linearly with system size – and the number of points sampled along each mode. In our implementation, the latter remains weakly dependent on the potential due to the variable, dynamically optimized sampling point density.

Analogously the VSCF approach exhibits a rough scaling behaviour of N^n where n is the

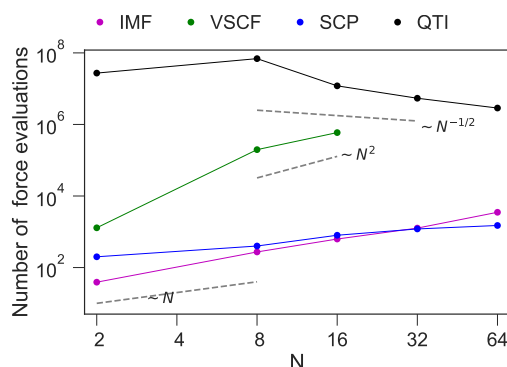


Figure 8.4 – Scaling of computational costs in the case of diamond in terms of the number of energy and force evaluations for IMF (pink), VSCF (green), SCP (blue) and QTI (black) with the number of atoms in the simulation cell. Here, we do not make use of any crystal symmetries to reduce the cost of calculations. For the approximate methods the free energy was converged to within 10% of the fully converged QTI results for 64 atoms with respect to all relevant convergence parameters. Analogously, the cost of QTI is defined as the total number of force evaluations required to drive the statistical error to below 10 % of the fully converged QTI results for 64 atoms.

dimensionality of the potential surfaces that are being sampled. For the case of diamond we use $n = 2$ and therefore observe a N^2 dependence for large N .

The cost for the SCP scheme, using pseudo random numbers, arises from the use of Monte Carlo importance sampling of the optimal effective harmonic description, which scales independently of system size. The statistical reweighting scheme reduces the cost for small sizes but becomes increasingly less beneficial as N is increased^[359]. The net result is a near linear scaling behaviour for the system sizes that we have considered. We note that the use of Sobol sequences improves the convergence of the MC integrals for small system sizes, and thereby reduces the cost of the SCP, but leads to an unfavourable exponential scaling for large N .

The reference calculations (QTI) were performed using a combination of a TI from the harmonic reference to the anharmonic potential using classical MD and a quantum TI over mass using PIMD. We define the cost of a QTI simulation as the total number of force evaluations required to drive the statistical error – the dominant source of error, given we converged the discretization of the integral^[375,376] and the number of replicas in a high-order PIMD scheme^[369] beyond that level – to below 10 % of the reference value for the largest system size obtained from a fully converged QTI simulation. One should note that the variance of the

integrands and the cost of performing one molecular dynamics step is different for classical and quantum MD. Thus the minimum number of force evaluations required to reduce the error to within the tolerance is an optimization problem. Since both the integrals are effective energies and the fluctuations of the potential energy in the canonical ensemble display a $\sim \frac{1}{\sqrt{N}}$ behaviour with respect to its mean, QTI also displays a $\sim \frac{1}{\sqrt{N}}$ scaling behaviour.

In summary, in the limit of small system size, SCP and IMF display the most favourable scaling. The reference technique QTI displays a $\mathcal{O}(1/\sqrt{N})$ behaviour, making it the least expensive method in the limit of large system size. For carbon, however, as well as for all the system discussed in this work, we do not reach this limit, and QTI requires a substantially larger number of force evaluations than either SCP or IMF. It is worth noting that at fixed cell size (unlike the other anharmonic methods) IMF and VSCF also provide the temperature dependence of the free energy without any additional force evaluations.

In the current implementation, none of the above free energy methods exploit crystal symmetries. Exploiting crystal symmetries in HAR, IMF and VSCF is straightforward and the reduction in computational cost is simply related to the reduction in the number of independent normal modes. Crystal symmetries can similarly, albeit not quite as trivially, be exploited in SCP^[358] and the other methods. However, crystal symmetries do not affect the overall scaling behaviour with respect to the number of degrees of freedom considered in a given calculation and have therefore not been regarded in the benchmarks for the computational cost.

8.3.3 Accuracy

We gauge the accuracy of the approximate methods by studying the error incurred in the absolute anharmonic free energy, and in the free energy differences between two phases of the same material.

Allotropes of Carbon

Diamond and lonsdaleite are mildly anharmonic systems which serve as excellent starting points for our study. We find that 32-atom simulation cells suffice to converge the free energy

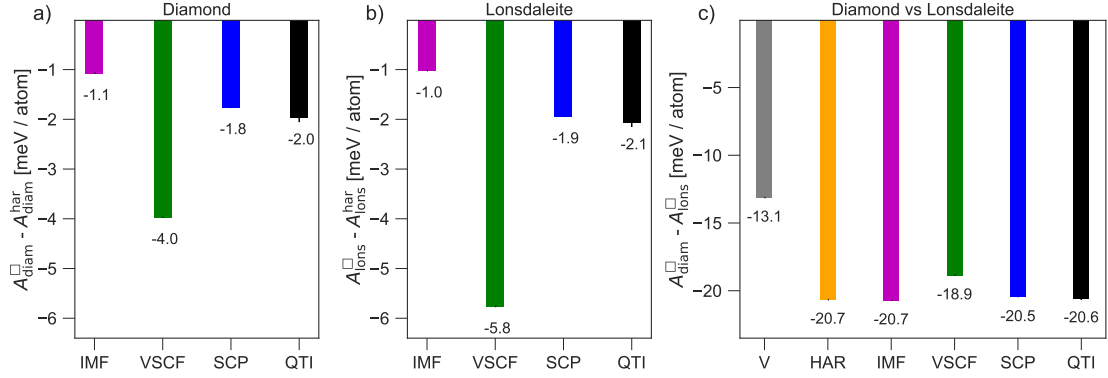


Figure 8.5 – Panels (a) and (b) respectively show the quantum anharmonic Helmholtz free energies $A_{\text{diam}}^{\square} - A_{\text{diam}}^{\text{har}}$ of diamond and lonsdaleite allotropes of carbon at 300 K with IMF (pink), VSCF (green), SCP (blue), and QTI (black). Panel (c) shows the difference in the potential energies of the minimum-energy configurations V (grey), and the free energy difference $A_{\text{diam}}^{\square} - A_{\text{lons}}^{\square}$ obtained using the harmonic approximation (yellow) and the aforementioned anharmonic methods.

difference between diamond and lonsdaleite, with respect to BZ sampling.

As shown in panels (a) and (b) of Fig. 8.5, the quantum anharmonic contribution to the free energy of both diamond and lonsdaleite is approximately 2 meV/atom. IMF, which considers anharmonicity only along normal modes, underestimates the anharmonic free energy by around 1 meV/atom. Including pairwise mean-field coupling using VSCF leads to a large over correction that increases the error to over 2 meV/atom, while SCP (which also includes a mean-field anharmonic corrections within Gaussian statistics) gives excellent results in comparison to the reference. This indicates that the error in VSCF arises from the truncation of the potential.

We also study the accuracy of the methods in reproducing the free energy difference between diamond and lonsdaleite, as shown in panel (c) of Fig. 8.5. Notably, the free energy contribution from quantum anharmonic motion for the two allotropes is almost identical so that there is almost no anharmonic correction to the free energy between the two phases.

Fortuitously IMF and VSCF benefit from large amounts of error cancellation and reproduce the exact result within the errors in the anharmonic free energies. Overall, all approximate methods perform reasonably well at reproducing both the (very small) anharmonic corrections and the free energy difference.

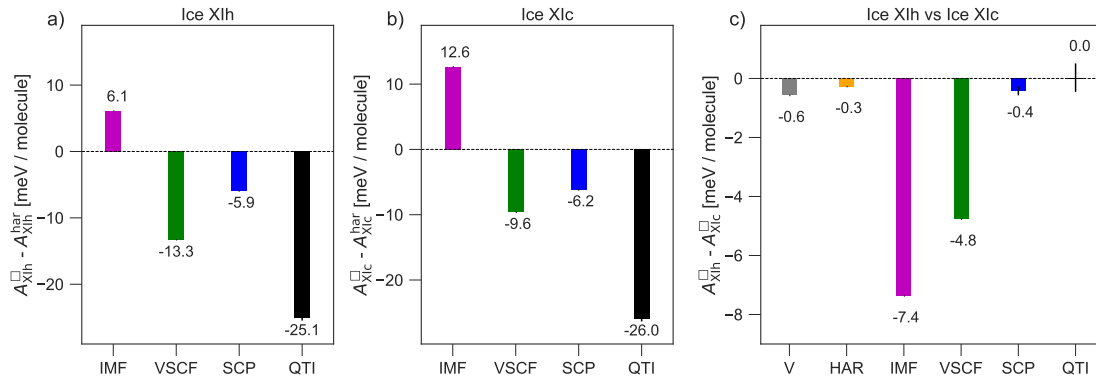


Figure 8.6 – Panels (a), (b) and (c) respectively show the quantum anharmonic Helmholtz free energies $A^{\square} - A^{\text{har}}$ of hexagonal and cubic polymorphs of ice XI at 70 K with IMF (pink), VSCF (green), SCP (blue) and QTI (black). Panel (c) shows the difference in the potential energies of the minimum-energy configurations V (grey), and the free energy difference $A^{\square}_{\text{XIh}} - A^{\square}_{\text{XIc}}$ obtained using the harmonic approximation (yellow) and the aforementioned anharmonic methods.

Polymorphs of Ice

Ice XIh and ice XIc are a more challenging test-case because of the large anharmonic contribution due to the pronounced anharmonicity of the O–H bond, the coupling between the high and low frequency modes^[377], and the small free energy difference between the hexagonal and cubic polymorphs^[327,328]. Supercells containing 16 molecules of water suffice to converge the free energy difference for all methods.

As shown in panels (a) and (b) of Fig. 8.6, the overall contribution from quantum anharmonicity to the free energy is around 25 meV/molecule for both systems. Contrary to the case of carbon, we find the approximate methods do not accurately reproduce the reference anharmonic free energy. For instance, the IMF technique produces qualitatively incorrect anharmonic corrections. The VSCF approach with pairwise couplings of normal modes provides the best approximation, but remains off by over 10 meV/molecule. The SCP scheme incurs errors of around 20 meV/molecule.

In line with previous path integral calculations on hexagonal and cubic ice^[328], we find the free energy difference between the polymorphs of ice XI to be almost zero, as shown in panel (c) of Fig. 8.6. IMF predicts the hexagonal form to be more stable by around 7 meV/molecule. After adding mean field coupling corrections within VSCF the margin of stability reduces to

around 5 meV/molecule. The SCP scheme benefits from cancellation of errors and fortuitously gives the correct result within 1 meV/molecule.

Polymorphs of Paracetamol

As a final test, we consider forms 1 and 2 of crystalline paracetamol. These are more complex molecular crystals, for which free energy calculations are complicated by the presence of quasi-free rotations of the methyl groups. Reference free energies are obtained by first integrating from the harmonic reference to the full potential at 10 K, then by performing classical TI with respect to temperature^[348] from 10 K to 300 K, and finally by quantum TI over masses. We found that, for this system, classical anharmonicity is almost completely suppressed at 10 K, and the classical anharmonic correction for the fluctuations around the potential energy minimum is essentially zero. We expect that, at an appropriately low temperature, a similar behavior will be shared by many systems, making this route more efficient than integrating between harmonic and full potential at the target temperature^[64]. The subtle issues connected with the degeneracy of the rotational conformers of the methyl group are discussed below. For reference, the free energies were recalculated using the TI route employed in Ref.^[64], and we were able to reproduce the same result within statistical error. As shown in Figs. 8.7 (a) and (b), the overall quantum anharmonic corrections for forms I and II are around -58 and -46 meV/molecule. All approximate anharmonic methods produce qualitatively incorrect anharmonic free energy corrections.

As shown in Fig. 8.7 (c), the free energy difference between the two forms is around 12 meV/molecule. The difference with respect to Ref.^[64] arises due to the use of slightly different lattice constants, a more accurate path integral sampling technique^[369] and a finer PPPM mesh for the Ewald summation of electrostatics. As in the cases of carbon and ice, the IMF and SCP benefit from significant error cancellation. The former correctly predicts form II to be more stable but performs worse than a harmonic approximation in getting the correct magnitude. The latter also predicts the correct sign but fortuitously estimates the magnitude to within 5 meV/molecule of the exact result. VSCF doesn't benefit from error cancellation to the same extent and overestimates the stability of form II by over 70 meV/molecule.

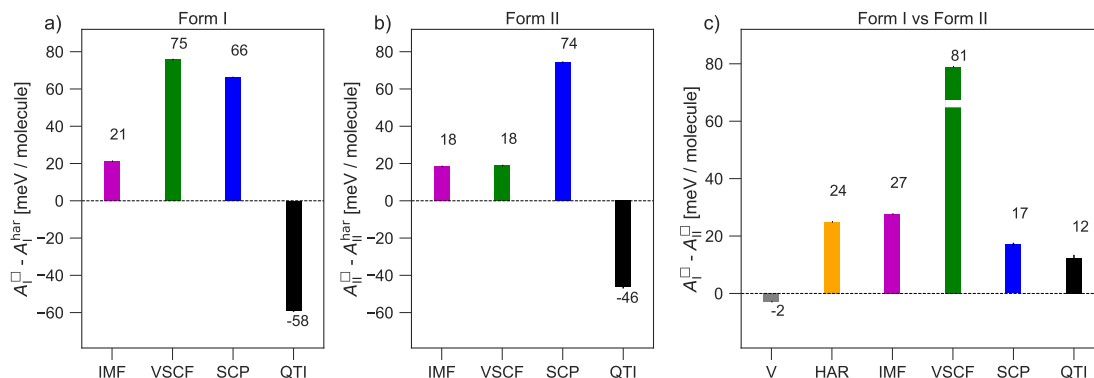


Figure 8.7 – Panels (a), (b) and (c) respectively show the quantum anharmonic Helmholtz free energies $A^\square - A^{\text{har}}$ of form I and form II polymorphs of crystalline paracetamol at 300 K with IMF (pink), VSCF (green), SCP (blue) and QTI (black). Panel (c) shows the difference in the potential energies of the minimum-energy configurations V (grey), and the free energy difference $A_I^\square - A_{II}^\square$ obtained using the harmonic approximation (yellow) and the aforementioned anharmonic methods.

An important contribution to the anharmonic free energy correction for each phase is due to the degeneracy of the rotational conformers of the methyl groups. In the presence of D degenerate, non-overlapping states, the conformational contribution to the free energy amounts to $-k_B T \ln D$ (for reference, at room temperature a three-fold degeneracy contributes about 30 meV, which is comparable with the overall anharmonic corrections in this system). We take this term into account in QTI calculations by computing harmonic to anharmonic corrections for a single minimum at low temperature, explicitly including the configurational entropy. This term is very small at sufficiently low temperature (< 1 meV per molecule at 10K, in this case) and can be safely ignored in general cases, in which knowledge of possible degeneracies is not available. We then perform a TI over temperature using replica exchange molecular dynamics^[378], so that the degenerate states are sampled even at the lowest temperatures. We don't include this term in the approximate methods, because at 300 K the distributions associated with the three states show some overlap, and so a simple configurational entropy term ceases to be rigorous.

The failure of normal mode based approaches for paracetamol is unsurprising, as the description of quasi-free rotations requires curvilinear coordinates. In paracetamol the potential energy barrier for rotational motion corresponds to approximately $200 k_B K$, implying (even classically) quasi-free rotation of the methyl groups at room temperature. For the force field

used to describe paracetamol, the potential governing rotation and breathing of methyl groups can be extracted explicitly (neglecting coupling to the remainder of the molecule) and takes the simple form

$$V(r, \theta) = \frac{1}{2}k(r - r_0)^2 + V_\theta(1 - \cos(3\theta)) \quad (8.39)$$

where $k = 53.114 \text{ eV/\AA}^2$, $r_0 = 1.09 \text{ \AA}$, and $V_\theta = 8.54 \text{ meV}$. This allows us to study the failure of the above approximate free energy methods with the activation of angular motion on the basis of a simple toy model,

$$\hat{H} = -\frac{\hbar^2}{2\mu} \left[\frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} \right] + V(r, \theta) \quad (8.40)$$

which can easily be studied over a range of temperatures. The exact solution for this simple model is obtained by exact diagonalization (ED) of the Hamiltonian matrix on a regular, square two dimensional real space grid of 256×256 points spanning from $(x, y) = (-1.5r_0, -1.5r_0)$ to $(x, y) = (1.5r_0, 1.5r_0)$. We find that the reference free energy is converged to within 0.2 meV .

The temperature dependence of the free energy of the model system is shown in Fig. 8.8. At high temperature, a large fraction of the discrepancy is due to the degeneracy, that is not captured by the approximate descriptions. Even at the lowest temperature, where the configurational entropy term is small, however, only VSCF – which in this case treats the full dimensionality of the problem, albeit at the mean-field level – accurately captures the full anharmonic correction. With rising temperature the VSCF approximation becomes increasingly inaccurate, as the amplitude of angular motion of methyl groups increases and the vibrational density delocalizes over the three equivalent potential energy minima (see Fig. 8.8). The harmonic, IMF, and SCP approximations severely overestimate the free energy throughout. For the harmonic approximation this can be explained by the fact that linear coordinates mix angular and much higher frequency radial motion, so that the effective “angular mode” is stiffened substantially, while the radial mode retains the true harmonic frequency. IMF and SCP exacerbate this effect, and yield essentially the same free-energy estimate.

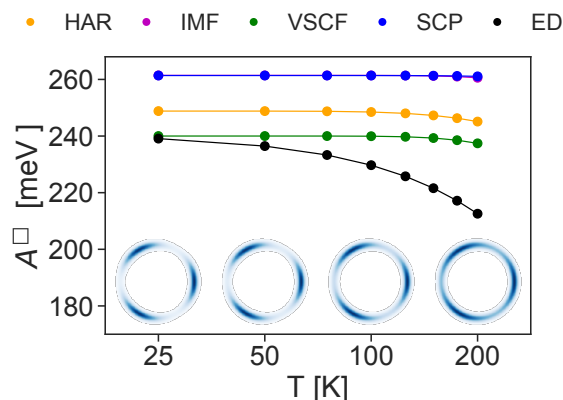


Figure 8.8 – Temperature dependence of the Helmholtz free energy of a quasi-free rotor obtained with IMF (pink), VSCF (green), SCP (blue). Reference data obtained from exact diagonalization are shown in black. The insets show the position distribution function at temperatures of 25 K, 50 K, 100 K and 200 K with blue and white indication high and low probability.

8.4 Discussion

Diamond and lonsdaleite, as examples of simple weakly anharmonic solids, highlight the utility of approximate free energy methods. While the accuracy of the approximate Helmholtz free energies varies, all approaches achieve sub-2 meV/atom accuracy and, more importantly, consistently yield a systematic improvement over the harmonic approximation at a substantially lower computational cost than the reference QTI. On the other hand, ice and paracetamol, as examples of more complex, molecular crystals, highlight the limitations of approximate techniques. The free energies of the molecular crystals are substantially overestimated due to the inherent limitations of normal modes based descriptions in the presence of large-amplitude curvilinear librational or quasi-free rotational motion. The simple model description of the rotation of the methyl group in paracetamol demonstrates that SCP, IMF and VSCF artificially stiffen the rotational modes, leading to an overestimation of the total free energy. The failure of these methods for ice can also be understood along the same lines: at larger displacements along the normal modes initially corresponding to librational motion, O–H bonds are stretched and bent, leading to an overestimation of the effective frequency of librational motion and consequently the free energy. This is confirmed by the blue shifts of the librational modes observed in the case of IMF and SCP with respect to the harmonic approximation. Consequently, these methods do not consistently yield systematic improvements over the

harmonic approximation for free energies of solids that possess high amplitude librational or quasi-rotational modes. However, we expect these techniques to perform well for atomic and ionic solids, where the point-like nature of the atomic/ionic building blocks suppresses large-amplitude curvilinear motion, including also cases, such as ferroelectrics, in which anharmonicities are very strong but do not have a curvilinear character. It is worth mentioning that (in suitable applications) normal modes based approximate methods lend themselves to identifying the atomistic/structural origins of anharmonicity and facilitate analyses of, for example, spectral properties of strongly anharmonic phonons, as probed by inelastic scattering processes, the formation of charge-density-waves, and ferroelectric instabilities^[379].

We also find that the approximate methods benefit from error cancellation, leading to errors in free energy differences that are consistently smaller than the errors in the absolute Helmholtz free energies. One should consider, however, that we have compared systems with very similar local environments. Within that perspective, the quasi-harmonic approximation^[380] and its extension to approximate anharmonic methods which minimizes the free energy w.r.t. to the volume of the system, may lead to a more significant error cancellation. However, such beneficial cancellation of errors is not guaranteed in general.

As the approximate results can vary from almost quantitatively accurate to qualitatively incorrect results, QTI is the only free energy methods among those considered in this work that provides reliable anharmonic free energies for large and complex organic solids. Given that it displays a $\mathcal{O}(1/\sqrt{N})$ computational cost, it may furthermore require comparable or fewer force evaluations than a SCP or VSCF calculation for systems of interest, in particular when considering biological or pharmaceutical compounds that involve large unit cells with flexible molecular units. It is further worth noting that QTI (and other statistical sampling methods such as SCP) are substantially less susceptible to noise in the underlying PES than the harmonic approximation, IMF and VSCE. While random noise largely cancels out in the ensemble averages calculated in statistical sampling methods, especially in combination with stochastic thermostats^[381,382], the harmonic approximation relies on the ability to determine a meaningful dynamical matrix and thus a differentiable PES, and the IMF and VSCF require an interpolatable PES. This is demonstrated by performing the Ewald summation in the description of paracetamol using a coarser PPPM mesh, which leads to discontinuities in the

PES of 0.50 meV. While the resultant QTI and SCP free energies remain largely unaffected, the free energy estimates obtained from the analytic methods – at least in the implementation we discuss here – depend heavily on the size of the finite displacements underlying the mapping of the PES, and the VSCF in particular eventually fails to converge altogether. In cases in which achieving sufficient absolute accuracy is impossible, one should consider using regularized approximations of the PES rather than a strict interpolation.

Efforts towards obtaining a computationally feasible anharmonic free energy should therefore be channelled towards reducing the cost of performing a QTI or, at least, a classical-nuclei TI with nuclear quantum contributions evaluated at a more approximate level^[64,63]. This includes streamlining hierarchical frameworks^[383] that perform the full free energy calculations using inexpensive bespoke potentials^[384] or cheaper basis sets^[318], reducing the cost to that of reversibly switching an *ab initio* potential. Machine learning potentials offer exceptional promise to provide ab-initio-quality potential energy surfaces to evaluate the anharmonic free energy, and approximate methods could also constitute an effective sampling approach to generate data to train and validate such ML potentials.

9 Conclusions

Accurate and reliable prediction of material properties requires high accuracy simulations that model the quantum mechanics (QM) of electrons and nuclei in a computationally efficient manner. While a lot of attention has been given to the QM of electrons, a classical approximation has been routinely adopted to model the motion of nuclei which gives good results for studying systems at high temperatures, and those containing only heavy nuclei. To capture the quantum mechanical motion of nuclei in systems at low temperatures or those containing light atoms at room temperature and beyond, imaginary time path integral simulations are required which simulate multiple replicas of the system and thus are typically one to two orders of magnitude more expensive than standard simulations that treat nuclei within the classical approximation.

The high computational cost and non-trivial implementation of path integral methods has prevented their widespread adoption in combination with first principles electronic structure methods. Within this perspective, the use of generalized Langevin equation (GLE) thermostats that reduce the associated computational cost of path integral simulations, and the open source software *i-PI* which allows simulations with several electronic structure codes have ensured a surge in the use of path integral methods for modelling the quantum nature of nuclei. However, GLE based methods only provide a semi quantitative description of nuclear quantum effects, and can be applied only to simple structural observables. The study of dynamical properties – such as vibrational spectrum – heat capacities, and momentum dependent observables, within an efficient and transparent framework remains an open challenge.

Chapter 9. Conclusions

In this thesis, we have presented an array of methodologies that accelerate the convergence of the imaginary time path integral method in a systematic manner. First of all, we have combined ring polymer contraction (RPC) that allow different number of replicas for slow and fast component of the potential energy, with multiple timestep (MTS) methods that allow slow components of the forces to be evaluated less frequently, in the context of ab initio electronic structure path integral simulations. We have shown that this combination leads to a dramatic reduction in cost of simulations performed at a high level of electronic structure theory by utilizing low cost methodologies which can accurately model short ranged forces on the cheap. The method also displays excellent performance for the computation of dynamical properties such as vibrational spectra and diffusion coefficients.

To be able to perform simulations in the high accuracy regime, we have presented a revamp of the high-order path integral method which is based on a more accurate splitting of the Boltzmann operator than the standard method. This technique has not been popular due to the need to compute the Hessian of the potential at each time step. We have developed an integration scheme which prevents the evaluation of the Hessian, and makes it possible to harness the efficiency of the technique. The cost of this method can be further reduced by using range separation methods like RPC and MTS. We expect this method to be useful when highly accurate comparisons need to be made with experiments, or when quantities requiring large number of replicas such as the heat capacity or free energies, are to be studied.

We have also developed a simple methodology that simplifies the implementation of the "open" path integral method, useful for studying momentum dependent observables. We also present a highly efficient "virial like" estimator of the particle momentum distribution which reduces the associated computational cost by over an order of magnitude than standard methods, making it possible to make comparisons with highly accurate Deep Inelastic Neutron Scattering experiments.

In order to model dynamical properties, such as vibrational spectra, we present an inexpensive methodology based on the use of GLE thermostats to accelerate the convergence of quantum nuclear effects. Typically, GLE thermostats use artificial noise to model quantum fluctuations that induces major artefacts in the vibrational spectra – such as severe broadening and shifting of peaks. We show that it is possible to analytically estimate the dynamical perturbation induced by the artificial noise, and to a good extent correct for these artefacts using a deconvolution algorithm, making it possible to recover meaningful dynamical properties at a low computational cost.

All the methods have been implemented in the second release of the *i-PI* code which has allowed their combination with softwares that evaluate potential energies and forces. Access to first principle as well as machine learning potential energy surfaces has allowed application of these methods to a wide class of materials. For instance, we have studied the proton momentum distributions (PMDs) and their kinetic energies, in several phases of water such as room temperature liquid water, supercooled water, and ice. This quantity is influenced by the local environment, of protons, giving subtle information on chemical environments in different phases of water. The computation of the PMD is challenging for both theory and experiments, and its deviation from theoretical results will have implications on our understanding of the hydrogen bond network in water. Our results have shown an excellent agreement between theoretical and experimental results for supercooled water and ice, but a disagreement for liquid water. Our results also show an absence of the experimentally observed anomalous non-monotonic temperature dependence of the proton kinetic energies, which we confirm using several accurate water models, indicating either an experimental problem for the measurement at room temperature, or an incomplete understanding of liquid water.

We also complement the PMDs with calculations of the IR and Raman spectrum of water, made inexpensive using the simple deconvolution method developed by us together with machine learning predictions of polarization and polarizability computed at first principles electronic structure level. Our results show an overall good agreement with experimental IR and Raman spectra for liquid water, and IR spectra for ice, and that low cost modelling of more

Chapter 9. Conclusions

advanced spectroscopic quantities such as sum frequency generation and second harmonic scattering should be possible by combining our deconvolution method with machine learning predictions of dielectric responses.

We have also applied the methodologies we developed to study the impact of nuclear quantum effects on the thermal and structural properties of nano-porous materials such as metal organic framework. In particular, we have studied the temperature dependent heat capacity of MOF-5, a promising candidate for the absorption of green house gases, as a function of the loading with methane gas. The use of the high-order path integral method, in combination with RPC and MTS, paves the way for an efficient computation of this quantity, which would have otherwise been prohibitive. Our results show that nuclear quantum effects lead to an increase in the heat capacity of the system as the temperature is increased, while anharmonic effects reduce this quantity. The extent of increase and decrease can be tuned by carefully selecting a MOF and a guest molecule, allowing for a novel mechanism for controlling the heat capacity of the total system to increase, decrease or stay constant over a range of temperature.

Finally, we have studied the role of the quantum nature of nuclei in influencing the free energy of solids. Even though this quantity can be calculated using path integral free energy calculations, a number of methods have been developed which approximate it, beyond the harmonic approximation, and provide a computationally inexpensive estimate of the anharmonic free energy. Understanding which of the approximations involved are justified for a given system is complicated by the lack of comparative benchmarks. To facilitate this choice we have assessed the accuracy of some of the commonly used approximate methods. We have compared anharmonic corrections to Helmholtz free energies against reference path integral calculations for a diverse set of systems, ranging from simple weakly anharmonic solids to flexible molecular crystals with freely-rotating units. The results suggest that, for simple solids such as allotropes of carbon, these methods yield results that are in excellent agreement with the reference calculations, at a considerably lower computational cost. For more complex molecular systems such as polymorphs of ice and paracetamol the methods do not consistently provide a reliable approximation of the anharmonic correction. Despite

substantial cancellation of errors when comparing the stability of different phases, we do not observe a systematic improvement over the harmonic approximation even for relative free energies. We conclude that efforts toward obtaining computationally feasible anharmonic free energies of organic molecular crystals should therefore be directed toward streamlining path integral free energy calculations.

To conclude, this thesis provides the necessary infrastructure for modelling the properties of materials while explicitly including the quantum nature of nuclei. The methods that we have developed provide a means to efficient calculations of a wide array of properties, including vibrational spectra, heat capacities, particle momentum distributions and free energies, while systematically improving upon the limitations of already existing cost reduction schemes. The implementation of these methods in an open-source code that can connect to first principles softwares, ensures the combination of these methods with the latest electronic structure methods. Our methods eliminate the systematic errors due approximations in sampling finite temperature vibrations and quantum mechanical fluctuations. Therefore, any disagreement of static properties with experimental results can now be traced to underlying approximations in potential energy landscapes. Therefore, to reach the ultimate level of accuracy, focus should now be using highly accurate energetics from advanced wave function methods such as quantum Monte Carlo. Since, the cost of such simulations can be large despite the saving brought by the methods developed in this thesis, we recommend the use of machine learning potentials as surrogate models for sampling. Combination of path integral methods with methods such as quantum Monte Carlo would lead to near exact results at given thermodynamic conditions, and will pave the way for solving open problems such as organic crystal structure prediction, and in silico materials design and discovery.

Bibliography

- [1] V. Kapil, J. VandeVondele, and M. Ceriotti, “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,” *The Journal of Chemical Physics*, vol. 144, p. 054111, Feb. 2016.
- [2] V. Kapil, J. Behler, and M. Ceriotti, “High order path integrals made easy,” *The Journal of Chemical Physics*, vol. 145, p. 234103, Dec. 2016.
- [3] M. Rossi, V. Kapil, and M. Ceriotti, “Fine tuning classical and quantum molecular dynamics using a generalized Langevin equation,” *The Journal of Chemical Physics*, vol. 148, p. 102301, July 2017.
- [4] V. Kapil, A. Cuzzocrea, and M. Ceriotti, “Anisotropy of the Proton Momentum Distribution in Water,” *The Journal of Physical Chemistry B*, vol. 122, pp. 6048–6054, June 2018.
- [5] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. V. Speybroeck, and M. Ceriotti, “i-PI 2.0: A universal force engine for advanced molecular simulations,” *Comput. Phys. Commun.*, vol. 236, pp. 214–223, oct 2019.
- [6] V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. Van Speybroeck, and M. Ceriotti, “Modeling the Structural and Thermal Properties of Loaded Metal–Organic Frameworks.

Bibliography

- An Interplay of Quantum and Anharmonic Fluctuations,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 3237–3249, May 2019.
- [7] V. Kapil, E. Engel, M. Rossi, and M. Ceriotti, “Assessment of Approximate Methods for Anharmonic Free Energies,” *Journal of Chemical Theory and Computation*, vol. 15, pp. 5845–5857, Nov. 2019.
- [8] V. Kapil, D. M. Wilkins, J. Lan, and M. Ceriotti, “Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats,” *The Journal of Chemical Physics*, vol. 152, p. 124104, Mar. 2020.
- [9] K. Druzbicki, M. Krzystyniak, D. Hollas, V. Kapil, P. Slavíček, G. Romanelli, and F. Fernandez-Alonso, “Hydrogen dynamics in solid formic acid: Insights from simulations with quantum colored-noise thermostats,” *Journal of Physics: Conference Series*, vol. 1055, p. 012003, July 2018.
- [10] J. Wieme, S. Vandenbrande, A. Lemaire, V. Kapil, L. Vanduyfhuys, and V. V. Speybroeck, “Thermal engineering of metal-organic frameworks for adsorption applications: A molecular simulations perspective,” *ACS Applied Materials & Interfaces*, 2019.
- [11] M. Liu, L. Zhang, M. A. Little, V. Kapil[†], M. Ceriotti, S. Yang, L. Ding, D. L. Holden, R. Balderas-Xicohtencatl, D. He, R. Clowes, S. Y. Chong, G. Schütz, L. Chen, M. Hirscher, and A. I. Cooper, “Barely porous organic cages for hydrogen isotope separation,” *Science*, vol. 366, pp. 613–620, Oct. 2019.
- [12] S. Awasthi, V. Kapil, and N. N. Nair, “Sampling free energy surfaces as slices by combining umbrella sampling and metadynamics,” *Journal of Computational Chemistry*, vol. 37, pp. 1413–1424, Apr. 2016.
- [13] A. Forsyth, “The information age: Economy, society and culture,” *Journal of Planning Education and Research*, vol. 19, pp. 211–213, dec 1999.
- [14] M. Mendelson, “Practical solutions to the antibiotic resistance crisis,” *South African Medical Journal*, vol. 105, p. 413, May 2015.
- [15] P. A. Owusu and S. Asumadu-Sarkodie, “A review of renewable energy sources, sustainability issues and climate change mitigation,” *Cogent Engineering*, vol. 3, Apr. 2016.

-
- [16] I. Tuomi, “The lives and death of moore's law,” *First Monday*, vol. 7, Nov. 2002.
- [17] “Quantum mechanics of many-electron systems,” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 123, pp. 714–733, Apr. 1929.
- [18] J. M. Wing, “Computational thinking,” *Communications of the ACM*, vol. 49, p. 33, Mar. 2006.
- [19] A. Rahman, “Correlations in the motion of atoms in liquid argon,” *Physical Review*, vol. 136, pp. A405–A411, Oct. 1964.
- [20] J. F. Wendt, ed., *Computational Fluid Dynamics*. Springer Berlin Heidelberg, 2009.
- [21] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. London: Academic Press, second ed., 2002.
- [22] M. Tuckerman, *Statistical Mechanics and Molecular Simulations*. Oxford University Press, 2008.
- [23] J. A. Elliott, “Novel approaches to multiscale modelling in materials science,” *International Materials Reviews*, vol. 56, pp. 207–225, July 2011.
- [24] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill, 1964.
- [25] G. Reiter, J. C. Li, J. Mayers, T. Abdul-Redah, and P. Platzman, “The proton momentum distribution in water and ice,” *Brazilian J. Phys.*, vol. 34, pp. 142–147, 2004.
- [26] F. Pietrucci and W. Andreoni, “Graph Theory Meets Ab Initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale,” *Phys. Rev. Lett.*, vol. 107, p. 085504, 2011.
- [27] E. Schrödinger, “An undulatory theory of the mechanics of atoms and molecules,” *Physical Review*, vol. 28, pp. 1049–1070, Dec. 1926.
- [28] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. D. Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. D. Marco,

Bibliography

- C. Draxl, M. Duřak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Gianozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuřan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. V. Speybroeck, J. M. Wills, J. R. Yates, G. Zhang, and S. Cottenier, “Reproducibility in density functional theory calculations of solids,” *Science*, vol. 351, p. 6280, 2016.
- [29] B. M. Austin, D. Y. Zubarev, and W. A. Lester, “Quantum Monte Carlo and related methods,” *Chem. Rev.*, vol. 112, p. 263, 2012.
- [30] G. Onida, L. Reining, and A. Rubio, “Electronic excitations: density-functional versus many-body Green’s-function approaches,” *Rev. Mod. Phys.*, vol. 74, p. 601, 2002.
- [31] G. Kotliar, S. Y. Savrasov, K. Haule, V. S. Oudovenko, O. Parcollet, and C. A. Marianetti, “Electronic structure calculations with dynamical mean-field theory,” *Rev. Mod. Phys.*, vol. 78, p. 865, 2006.
- [32] G. Rohringer, H. Hafermann, A. Toschi, A. A. Katanin, A. E. Antipov, M. I. Katsnelson, A. I. Lichtenstein, A. N. Rubtsov, and K. Held, “Diagrammatic routes to nonlocal correlations beyond dynamical mean field theory,” *Rev. Mod. Phys.*, vol. 90, p. 025003, 2018.
- [33] F. Neese, M. Atanasov, G. Bistoni, D. Maganas, and S. Ye, “Chemistry and quantum mechanics in 2019: Give us insight and numbers,” *Journal of the American Chemical Society*, vol. 141, pp. 2814–2824, Jan. 2019.
- [34] M. Born and R. Oppenheimer, “Zur Quantentheorie der Molekeln,” *Annalen der Physik*, vol. 389, p. 457, 1927.
- [35] M. P. Bircher, E. Liberatore, N. J. Browning, S. Brickel, C. Hofmann, A. Patoz, O. T. Unke, T. Zimmermann, M. Chergui, P. Hamm, U. Keller, M. Meuwly, H.-J. Woerner, J. Vaníček, and U. Rothlisberger, “Nonadiabatic effects in electronic and nuclear dynamics,” *Structural Dynamics*, vol. 4, p. 061510, Nov. 2017.

-
- [36] A. Tehlar and H. J. Wörner, "Time-resolved high-harmonic spectroscopy of the photodissociation of CH₃I and CF₃I," *Molecular Physics*, vol. 111, pp. 2057–2067, Apr. 2013.
- [37] W. Fang, R. A. Zarotiadis, and J. O. Richardson, "Revisiting nuclear tunnelling in the aqueous ferrous–ferric electron transfer," *Physical Chemistry Chemical Physics*, 2020.
- [38] H. Kahn and T. E. Harris, "Estimation of particle transmission by random sampling.monte carlo method," *National Bureau, of Standards Applied Mathematics Series*, vol. 12, pp. 27–30, 1951.
- [39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1092, 1953.
- [40] M. Tuckerman, B. J. Berne, and G. J. Martyna, "Reversible multiple time scale molecular dynamics," *J. Chem. Phys.*, vol. 97, p. 1990, 1992.
- [41] L. Verlet, "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1967.
- [42] G. Bussi and M. Parrinello, "Accurate sampling using Langevin dynamics," *Phys. Rev. E*, vol. 75, p. 56707, 2007.
- [43] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *J. Chem. Phys.*, vol. 72, pp. 2384–2393, 1980.
- [44] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, pp. 511–519, July 1984.
- [45] G. J. Martyna, M. E. Tuckerman, and M. L. Klein, "Nosé-Hoover chains: The canonical ensemble via continuous dynamics," *J. Chem. Phys.*, vol. 97, p. 2635, 1992.
- [46] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, vol. 126, p. 14101, 2007.
- [47] P. Minari, G. J. Martyna, and M. E. Tuckerman, "Algorithms and novel applications based on the isokinetic ensemble. i. biophysical and path integral molecular dynamics," *The Journal of Chemical Physics*, vol. 118, no. 6, p. 2510, 2003.

Bibliography

- [48] W. G. Hoover, “Canonical dynamics: Equilibrium phase-space distributions,” *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.
- [49] W. Hoover, “Constant-pressure equations of motion,” *Phys. Rev. A*, vol. 34, pp. 2499–2500, 1986.
- [50] S. Melchionna, G. Ciccotti, and B. L. Holian, “Hoover NPT dynamics for systems varying in shape and size,” *Molecular Physics*, vol. 78, pp. 533–544, Feb. 1993.
- [51] G. J. Martyna, D. J. Tobias, and M. L. Klein, “Constant pressure molecular dynamics algorithms,” *J. Chem. Phys.*, vol. 101, p. 4177, 1994.
- [52] G. Bussi, T. Zykova-Timan, and M. Parrinello, “Isothermal-isobaric molecular dynamics using stochastic velocity rescaling,” vol. 130, p. 74101, 2009.
- [53] R. Kubo, “Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems,” *Journal of the Physical Society of Japan*, vol. 12, pp. 570–586, June 1957.
- [54] A. Pérez, M. E. Tuckerman, and M. H. Müser, “A comparative study of the centroid and ring-polymer molecular dynamics methods for approximating quantum time correlation functions from path integrals,” *J. Chem. Phys.*, vol. 130, p. 184105, 2009.
- [55] D. Marx, M. E. Tuckerman, J. Hutter, and M. Parrinello, “The nature of the hydrated excess proton in water,” *Nature*, vol. 397, pp. 601–604, Feb. 1999.
- [56] D. Marx, A. Chandra, and M. E. Tuckerman, “Aqueous basic solutions: Hydroxide solvation, structural diffusion, and comparison to the hydrated proton,” *Chemical Reviews*, vol. 110, pp. 2174–2216, Apr. 2010.
- [57] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelidis, M. A. Morales, and T. E. Markland, “Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges,” *Chemical Reviews*, vol. 116, p. 7529, 2016.
- [58] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, “Nuclear quantum effects and hydrogen bond fluctuations in water,” *Proc. Natl. Acad. Sci. USA*, vol. 110, pp. 15591–6, 2013.

-
- [59] L. Wang, M. Ceriotti, and T. E. Markland, “Quantum fluctuations and isotope effects in ab initio descriptions of water,” *The Journal of Chemical Physics*, vol. 141, p. 104502, Sept. 2014.
- [60] D. M. Ceperley, “Path integrals in the theory of condensed helium,” *Reviews of Modern Physics*, vol. 67, pp. 279–355, apr 1995.
- [61] I. Errea, M. Calandra, C. J. Pickard, J. Nelson, R. J. Needs, Y. Li, H. Liu, Y. Zhang, Y. Ma, and F. Mauri, “High-Pressure Hydrogen Sulfide from First Principles: A Strongly Anharmonic Phonon-Mediated Superconductor,” *Physical Review Letters*, vol. 114, p. 157004, 2015.
- [62] A. Lemaire, J. Wieme, S. M. J. Rogge, M. Waroquier, and V. V. Speybroeck, “On the importance of anharmonicities and nuclear quantum effects in modelling the structural properties and thermal expansion of MOF-5,” *The Journal of Chemical Physics*, vol. 150, p. 094503, mar 2019.
- [63] V. Kapil, J. Wieme, S. Vandenbrande, A. Lemaire, V. V. Speybroeck, and M. Ceriotti, “Modeling the structural and thermal properties of loaded metal-organic frameworks. an interplay of quantum and anharmonic fluctuations,” *arXiv*, 2019.
- [64] M. Rossi, P. Gasparotto, and M. Ceriotti, “Anharmonic and Quantum Fluctuations in Molecular Crystals: A First-Principles Study of the Stability of Paracetamol,” *Physical Review Letters*, vol. 117, p. 115702, 2016.
- [65] H.-Y. Ko, R. A. DiStasio, B. Santra, and R. Car, “Thermal expansion in dispersion-bound molecular crystals,” *Phys. Rev. Materials*, vol. 2, p. 055603, May 2018.
- [66] D. Chandler and P. G. Wolynes, “Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids,” *The Journal of Chemical Physics*, vol. 74, pp. 4078–4095, Apr. 1981.
- [67] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein, “Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals,” *J. Chem. Phys.*, vol. 99, p. 2796, 1993.
- [68] M. Parrinello and A. Rahman, “Study of an F center in molten KCl,” *J. Chem. Phys.*, vol. 80, p. 860, 1984.

Bibliography

- [69] J. A. Morrone, V. Srinivasan, D. Sebastiani, and R. Car, “Proton momentum distribution in water: an open path integral molecular dynamics study,” *J. Chem. Phys.*, vol. 126, p. 234504, 2007.
- [70] M. Parrinello and A. Rahman, “Study of an f center in molten KCl,” *The Journal of Chemical Physics*, vol. 80, pp. 860–867, jan 1984.
- [71] M. Ceriotti and D. E. Manolopoulos, “Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei,” *Physical Review Letters*, vol. 109, p. 100604, Sept. 2012.
- [72] G. J. Martyna, A. Hughes, and M. E. Tuckerman, “Molecular dynamics algorithms for path integrals at constant pressure,” *J. Chem. Phys.*, vol. 110, p. 3275, 1999.
- [73] M. Ceriotti, J. More, and D. E. Manolopoulos, “i-pi: A python interface for ab initio path integral molecular dynamics simulations,” *Comp. Phys. Commun.*, vol. 185, pp. 1019–1026, 2014.
- [74] M. Ceriotti, G. Bussi, and M. Parrinello, “Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations,” *Physical Review Letters*, vol. 102, p. 020601, Jan. 2009.
- [75] B. Leimkuhler and C. Matthews, “Robust and efficient configurational molecular sampling via langevin dynamics,” *The Journal of Chemical Physics*, vol. 138, p. 174102, may 2013.
- [76] Z. Zhang, X. Liu, Z. Chen, H. Zheng, K. Yan, and J. Liu, “A unified thermostat scheme for efficient configurational sampling for classical/quantum canonical ensembles via molecular dynamics,” *The Journal of Chemical Physics*, vol. 147, p. 034109, July 2017.
- [77] R. Korol, N. Bou-Rabee, and T. F. Miller, “Cayley modification for strongly stable path-integral and ring-polymer molecular dynamics,” *The Journal of Chemical Physics*, vol. 151, p. 124103, Sept. 2019.
- [78] R. Korol, J. L. Rosa-Raíces, N. Bou-Rabee, and T. F. Miller, “Dimension-free path-integral molecular dynamics without preconditioning,” *The Journal of Chemical Physics*, vol. 152, p. 104102, Mar. 2020.

-
- [79] T. J. H. Hele, M. J. Willatt, A. Muolo, and S. C. Althorpe, "Boltzmann-conserving classical dynamics in quantum time-correlation functions: "Matsubara dynamics"," *The Journal of Chemical Physics*, vol. 142, p. 134103, Apr. 2015.
- [80] S. Jang and G. a. Voth, "A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables," *J. Chem. Phys.*, vol. 111, p. 2371, 1999.
- [81] I. R. Craig and D. E. Manolopoulos, "Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics," *J. Chem. Phys.*, vol. 121, p. 3368, 2004.
- [82] S. Habershon, G. S. Fanourgakis, and D. E. Manolopoulos, "Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water," *J. Chem. Phys.*, vol. 129, p. 74501, 2008.
- [83] T. J. H. Hele, M. J. Willatt, A. Muolo, and S. C. Althorpe, "Communication: Relation of centroid molecular dynamics and ring-polymer molecular dynamics to exact quantum dynamics," *The Journal of Chemical Physics*, vol. 142, p. 191101, May 2015.
- [84] T. J. H. Hele, "On the relation between thermostatted ring-polymer molecular dynamics and exact quantum dynamics," *Molecular Physics*, vol. 114, pp. 1461–1471, Jan. 2016.
- [85] B. J. Braams and D. E. Manolopoulos, "On the short-time limit of ring polymer molecular dynamics," *The Journal of Chemical Physics*, vol. 125, p. 124105, Sept. 2006.
- [86] M. J. Willatt, M. Ceriotti, and S. C. Althorpe, "Approximating Matsubara dynamics using the planetary model: Tests on liquid water and ice," *The Journal of Chemical Physics*, vol. 148, p. 102336, Jan. 2018.
- [87] T. D. Hone, P. J. Rossky, and G. A. Voth, "A comparative study of imaginary time path integral based methods for quantum dynamics," *The Journal of Chemical Physics*, vol. 124, p. 154103, Apr. 2006.
- [88] I. R. Craig and D. E. Manolopoulos, "Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics," *The Journal of Chemical Physics*, vol. 121, pp. 3368–3373, Aug. 2004.

Bibliography

- [89] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, "How to remove the spurious resonances from ring polymer molecular dynamics," *The Journal of Chemical Physics*, vol. 140, p. 234116, June 2014.
- [90] M. Rossi, M. Ceriotti, and D. E. Manolopoulos, "How to remove the spurious resonances from ring polymer molecular dynamics.," *J. Chem. Phys.*, vol. 140, p. 234116, 2014.
- [91] T. E. Markland and D. E. Manolopoulos, "An efficient ring polymer contraction scheme for imaginary time path integral simulations.," *J. Chem. Phys.*, vol. 129, p. 024105, 2008.
- [92] W. Shinoda and M. Shiga, "Quantum simulation of the heat capacity of water," *Phys. Rev. E*, vol. 71, p. 041204, 2005.
- [93] M. Suzuki, "Hybrid exponential product formulas for unbounded operators with possible applications to Monte Carlo simulations," *Phys. Lett. A*, vol. 201, pp. 425–428, 1995.
- [94] S. A. Chin, "Symplectic integrators from composite operator factorizations," *Phys. Lett. A*, vol. 226, pp. 344–348, 1997.
- [95] A. Pérez and M. E. Tuckerman, "Improving the convergence of closed and open path integral molecular dynamics via higher order Trotter factorization schemes.," *J. Chem. Phys.*, vol. 135, p. 064104, 2011.
- [96] M. Ceriotti, D. E. Manolopoulos, and M. Parrinello, "Accelerating the convergence of path integral dynamics with a generalized Langevin equation," *The Journal of Chemical Physics*, vol. 134, p. 084104, Feb. 2011.
- [97] R. Dupuis, M. Benoit, M. E. Tuckerman, and M. Méheut, "Importance of a fully anharmonic treatment of equilibrium isotope fractionation properties of dissolved ionic species as evidenced by Li^+ (aq)," *Accounts of Chemical Research*, vol. 50, pp. 1597–1605, June 2017.
- [98] M. Rossi, W. Fang, and A. Michaelides, "Stability of complex biomolecular structures: van der waals, hydrogen bond cooperativity, and nuclear quantum effects," *The Journal of Physical Chemistry Letters*, vol. 6, pp. 4233–4238, oct 2015.

-
- [99] M. Ceriotti and T. E. Markland, "Efficient methods and practical guidelines for simulating isotope effects," *The Journal of Chemical Physics*, vol. 138, p. 014112, Jan 2013.
- [100] J. A. Morrone and R. Car, "Nuclear Quantum Effects in Water," *Phys. Rev. Lett.*, vol. 101, p. 17801, 2008.
- [101] L. Lin, J. A. Morrone, R. Car, and M. Parrinello, "Displaced Path Integral Formulation for the Momentum Distribution of Quantum Particles," *Phys. Rev. Lett.*, vol. 105, p. 110602, 2010.
- [102] I. R. Craig and D. E. Manolopoulos, "Chemical reaction rates from ring polymer molecular dynamics," *The Journal of Chemical Physics*, vol. 122, p. 084106, Feb. 2005.
- [103] E. Geva, Q. Shi, and G. A. Voth, "Quantum-mechanical reaction rate constants from centroid molecular dynamics simulations," *The Journal of Chemical Physics*, vol. 115, pp. 9209–9222, Nov. 2001.
- [104] W. H. Miller, "The Semiclassical Initial Value Representation: A Potentially Practical Way for Adding Quantum Effects to Classical Molecular Dynamics Simulations," *The Journal of Physical Chemistry A*, vol. 105, pp. 2942–2955, Apr. 2001.
- [105] J. Liu, "Path integral liouville dynamics for thermal equilibrium systems," *The Journal of Chemical Physics*, vol. 140, p. 224107, June 2014.
- [106] J. R. Cendagorta, Z. Bačić, and M. E. Tuckerman, "An open-chain imaginary-time path-integral sampling approach to the calculation of approximate symmetrized quantum time correlation functions," *The Journal of Chemical Physics*, vol. 148, p. 102340, Mar. 2018.
- [107] G. Trenins, M. J. Willatt, and S. C. Althorpe, "Path-integral dynamics of water using curvilinear centroids," *The Journal of Chemical Physics*, vol. 151, p. 054109, Aug. 2019.
- [108] D. M. Ceperley, "Path integrals in the theory of condensed helium," *Reviews of Modern Physics*, vol. 67, pp. 279–355, Apr. 1995.
- [109] B. Hirshberg, V. Rizzi, and M. Parrinello, "Path integral molecular dynamics for bosons," *Proceedings of the National Academy of Sciences*, vol. 116, pp. 21445–21449, Oct. 2019.

Bibliography

- [110] B. Hirshberg, M. Invernizzi, and M. Parrinello, "Path integral molecular dynamics for fermions: Alleviating the sign problem with the bogoliubov inequality," *The Journal of Chemical Physics*, vol. 152, p. 171102, May 2020.
- [111] R. Iftimie and J. Schofield, "Efficient ab initio sampling methods in rate constant calculations for proton-transfer reactions," *J. Chem. Phys.*, vol. 114, pp. 6763–6773, 2001.
- [112] H. Y. Geng, "Accelerating ab initio path integral molecular dynamics with multilevel sampling of potential surface," *J. Comput. Phys.*, vol. 283, pp. 299–311, 2015.
- [113] T. E. Markland and D. E. Manolopoulos, "A refined ring polymer contraction scheme for systems with electrostatic interactions," *Chem. Phys. Lett.*, vol. 464, p. 256, 2008.
- [114] N. Luehr, T. E. Markland, and T. J. Martínez, "Multiple time step integrators in ab initio molecular dynamics," *J. Chem. Phys.*, vol. 140, p. 084116, 2014.
- [115] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics," *J. Chem. Phys.*, vol. 101, pp. 6168–6183, 1994.
- [116] J. a. Morrone, T. E. Markland, M. Ceriotti, and B. J. Berne, "Efficient multiple time scale molecular dynamics: Using colored noise thermostats to stabilize resonances," *J. Chem. Phys.*, vol. 134, p. 14103, 2011.
- [117] M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, "Efficient stochastic thermostating of path integral molecular dynamics," *J. Chem. Phys.*, vol. 133, p. 124104, 2010.
- [118] M. Rossi, H. Liu, F. Paesani, J. Bowman, and M. Ceriotti, "Communication: On the consistency of approximate quantum dynamics simulation methods for vibrational spectra in the condensed phase," *J. Chem. Phys.*, vol. 141, p. 181101, 2014.
- [119] B. Leimkuhler, D. T. Margul, and M. E. Tuckerman, "Stochastic, resonance-free multiple time-step algorithm for molecular dynamics with very large time steps," *Molecular Physics*, vol. 111, pp. 3579–3594, Dec. 2013.

-
- [120] P. Minary, M. E. Tuckerman, and G. J. Martyna, “Long time molecular dynamics for enhanced conformational sampling in biomolecular systems,” *Physical Review Letters*, vol. 93, Oct. 2004.
- [121] M. Del Ben, J. Hutter, and J. VandeVondele, “Probing the structural and dynamical properties of liquid water with models including non-local electron correlation,” *J. Chem. Phys.*, vol. 143, p. 054506, 2015.
- [122] M. Del Ben, J. Hutter, and J. VandeVondele, “Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme,” *J. Chem. Theory Comput.*, vol. 9, pp. 2654–2671, 2013.
- [123] M. Del Ben, O. Schütt, T. Wentz, P. Messmer, J. Hutter, and J. VandeVondele, “Enabling simulation at the fifth rung of DFT: Large scale RPA calculations with excellent time to solution,” *Comput. Phys. Commun.*, vol. 187, pp. 120–129, 2015.
- [124] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, “cp2k: atomistic simulations of condensed matter systems,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, pp. 15–25, June 2013.
- [125] J. P. Perdew, K. Burke, and M. Ernzerhof, “Generalized Gradient Approximation made simple,” *Phys. Rev. Lett.*, vol. 77, p. 3865, 1996.
- [126] M. Del Ben, J. Hutter, and J. VandeVondele, “Second-order møller-plesset perturbation theory in the condensed phase: An efficient and massively parallel gaussian and plane waves approach,” *J. Chem. Theory Comput.*, vol. 8, pp. 4177–4188, 2012.
- [127] M. Del Ben, J. Hutter, and J. VandeVondele, “Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution-of-identity Gaussian and plane waves approach,” *J. Chem. Phys.*, vol. 143, p. 102803, 2015.
- [128] S. Goedecker, M. Teter, and J. Hutter, “Separable dual-space gaussian pseudopotentials,” *Phys. Rev. B*, vol. 54, pp. 1703–1710, Jul 1996.
- [129] E. E. Dahlke, , and D. G. Truhlar*, “Improved density functionals for water,” *The Journal of Physical Chemistry B*, vol. 109, no. 33, pp. 15677–15683, 2005.

Bibliography

- [130] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu," *The Journal of Chemical Physics*, vol. 132, no. 15, p. 154104, 2010.
- [131] R. P. Steele, "Multiple-timestep ab initio molecular dynamics with electron correlation.," *J. Chem. Phys.*, vol. 139, p. 011102, 2013.
- [132] M. Ceriotti and D. E. Manolopoulos, "Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei," *Phys. Rev. Lett.*, vol. 109, p. 100604, 2012.
- [133] M. Takahashi and M. Imada, "Monte Carlo calculation of quantum systems. II. Higher order correction," *J. Phys. Soc. Japan*, vol. 53, pp. 3765–3769, 1984.
- [134] S. S. Jang and G. A. Voth, "Applications of higher order composite factorization schemes in imaginary time path integral simulations," *J. Chem. Phys.*, vol. 115, pp. 7832–7842, 2001.
- [135] I. Poltavsky and A. Tkatchenko, "Modeling quantum nuclei with perturbed path integral molecular dynamics," *Chem. Sci.*, vol. 7, pp. 1368–1372, 2016.
- [136] M. Ceriotti, G. a. R. Brain, O. Riordan, and D. E. Manolopoulos, "The inefficiency of re-weighted sampling and the curse of system size in high order path integration," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 468, pp. 2–17, 2011.
- [137] I. Poltavsky and A. Tkatchenko, "Modeling quantum nuclei with perturbed path integral molecular dynamics," *Chem. Sci.*, vol. 7, pp. 1368–1372, 2016.
- [138] C. Predescu, "Reconstruction of thermally symmetrized quantum autocorrelation functions from imaginary-time data," *Phys. Rev. E*, vol. 70, p. 066705, 2004.
- [139] M. Buchowiecki, J. Vaníček, and J. Vaníček, "Monte Carlo evaluation of the equilibrium isotope effects using the Takahashi-Imada factorization of the Feynman path integral," *Chem. Phys. Lett.*, vol. 588, pp. 11–16, 2013.
- [140] A. Putrino, D. Sebastiani, and M. Parrinello, "Generalized variational density functional perturbation theory," *J. Chem. Phys.*, vol. 113, p. 7102, 2000.

-
- [141] M. Ceriotti, J. More, and D. E. Manolopoulos, “i-PI: A Python interface for ab initio path integral molecular dynamics simulations,” *Comput. Phys. Commun.*, vol. 185, pp. 1019–1026, 2014.
- [142] V. Kapil, J. VandeVondele, and M. Ceriotti, “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.*, vol. 144, p. 054111, 2016.
- [143] M. Ceriotti, D. E. Manolopoulos, and M. Parrinello, “Accelerating the convergence of path integral dynamics with a generalized Langevin equation,” *J. Chem. Phys.*, vol. 134, p. 84104, 2011.
- [144] G. Brain, *Higher Order Propagators in Path Integral Molecular Dynamics*. PhD thesis, Part II Chemistry Thesis, Oxford University, 2011.
- [145] M. Ceriotti, G. Bussi, and M. Parrinello, “Colored-Noise Thermostats à la Carte,” *J. Chem. Theory Comput.*, vol. 6, pp. 1170–1180, 2010.
- [146] M. Ceriotti, “GLE4MD.” <http://epfl-cosmo.github.io/gle4md>, 2010.
- [147] A. D. Becke, “Density-functional thermochemistry. III. The role of exact exchange,” *J. Chem. Phys.*, vol. 98, p. 5648, 1993.
- [148] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *J. Chem. Phys.*, vol. 132, p. 154104, 2010.
- [149] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, “Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach,” *Comput. Phys. Commun.*, vol. 167, pp. 103–128, 2005.
- [150] B. Cheng, J. Behler, and M. Ceriotti, “Nuclear quantum effects in water at the triple point: Using theory as a link between experiments,” *The Journal of Physical Chemistry Letters*, vol. 7, pp. 2210–2215, May 2016.
- [151] A. Singraber, T. Morawietz, J. Behler, and C. Dellago *to be published*.

Bibliography

- [152] S. Plimpton, “Fast Parallel Algorithms for Short-Range Molecular Dynamics,” *J. Comput. Phys.*, vol. 117, pp. 1–19, 1995.
- [153] F. Uhl, D. Marx, and M. Ceriotti, “Accelerated path integral methods for atomistic simulations at ultra-low temperatures,” *J. Chem. Phys.*, vol. 145, p. 054101, 2016.
- [154] S. Ganeshan, R. Ramírez, and M. V. Fernández-Serra, “Simulation of quantum zero-point effects in water using a frequency-dependent thermostat,” *Phys. Rev. B*, vol. 87, p. 134207, 2013.
- [155] L. Wang, M. Ceriotti, and T. E. Markland, “Quantum fluctuations and isotope effects in ab initio descriptions of water,” *J. Chem. Phys.*, vol. 141, p. 104502, 2014.
- [156] J. Cao and G. A. Voth, “A new perspective on quantum time correlation functions,” *J. Chem. Phys.*, vol. 99, pp. 10070–10073, 1993.
- [157] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller, “Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space,” *Annu. Rev. Phys. Chem.*, vol. 64, pp. 387–413, 2013.
- [158] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, “On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy,” *J. Chem. Phys.*, vol. 130, p. 194510, 2009.
- [159] D. J. Wales, “Decoding heat capacity features from the energy landscape,” *Physical Review E*, vol. 95, Mar. 2017.
- [160] J. Wieme, S. Vandenbrande, A. Lemaire, V. Kapil, L. Vanduyfhuys, and V. Van Speybroeck, “Thermal Engineering of Metal–Organic Frameworks for Adsorption Applications: A Molecular Simulation Perspective,” *ACS Applied Materials & Interfaces*, vol. 11, pp. 38697–38707, Oct. 2019.
- [161] B. Mu and K. S. Walton, “Thermal analysis and heat capacity study of metal-organic frameworks,” *J. Phys. Chem. C*, vol. 115, pp. 22748–22754, 2011.
- [162] K. R. Glaesemann and L. E. Fried, “Improved heat capacity estimator for path integral simulations,” *J. Chem. Phys.*, vol. 117, p. 3020, 2002.

-
- [163] T. M. Yamamoto, "Path-integral virial estimator based on the scaling of fluctuation coordinates: Application to quantum clusters with fourth-order propagators," *J. Chem. Phys.*, vol. 123, p. 104101, 2005.
- [164] M. Kolár and S. F. O'Shea, "A high-temperature approximation for the path-integral quantum monte carlo method," *Journal of Physics A: Mathematical and General*, vol. 29, no. 13, p. 3471, 1996.
- [165] T. M. Yamamoto, "Path-integral virial estimator based on the scaling of fluctuation coordinates: Application to quantum clusters with fourth-order propagators," *J. Chem. Phys.*, vol. 123, p. 104101, 2005.
- [166] S. Habershon, T. E. Markland, and D. E. Manolopoulos, "Competing quantum effects in the dynamics of a flexible water model," *J. Chem. Phys.*, vol. 131, p. 024501, jul 2009.
- [167] G. F. Reiter, J. Mayers, and J. Noreland, "Momentum-distribution spectroscopy using deep inelastic neutron scattering," *Phys. Rev. B*, vol. 65, p. 104305, 2002.
- [168] C. J. Burnham, G. F. Reiter, J. Mayers, T. Abdul-Redah, H. Reichert, and H. Dosch, "On the origin of the redshift of the OH stretch in ice ih: evidence from the momentum distribution of the protons and the infrared spectral density," *Physical Chemistry Chemical Physics*, vol. 8, no. 34, p. 3966, 2006.
- [169] C. Andreani, G. Romanelli, and R. Senesi, "Direct Measurements of Quantum Kinetic Energy Tensor in Stable and Metastable Water near the Triple Point: An Experimental Benchmark," *J. Phys. Chem. Letters*, vol. 7, pp. 2216–2220, 2016.
- [170] R. Quhe, M. Nava, P. Tiwary, and M. Parrinello, "Path integral metadynamics," *Journal of Chemical Theory and Computation*, vol. 11, pp. 1383–1388, Mar. 2015.
- [171] Y. Wu and R. Car, "Quantum momentum distribution and quantum entanglement in the deep tunneling regime," *The Journal of Chemical Physics*, vol. 152, p. 024106, Jan. 2020.
- [172] C. Andreani, D. Colognesi, J. Mayers, G. F. Reiter, and R. Senesi, "Measurement of momentum distribution of light atoms and molecules in condensed matter systems using inelastic neutron scattering," *Adv. Phys.*, vol. 54, pp. 377–469, 2005.

Bibliography

- [173] D. Flammini, A. Pietropaolo, R. Senesi, C. Andreani, F. McBride, A. Hodgson, M. a. Adams, L. Lin, and R. Car, “Spherical momentum distribution of the protons in hexagonal ice from modeling of inelastic neutron scattering data.,” *J. Chem. Phys.*, vol. 136, p. 024504, 2012.
- [174] F. Perakis, L. De Marco, A. Shalit, F. Tang, Z. R. Kann, T. D. Kühne, R. Torre, M. Bonn, and Y. Nagata, “Vibrational spectroscopy and dynamics of water,” *Chemical reviews*, vol. 116, no. 13, pp. 7590–7607, 2016.
- [175] S. Woutersen, U. Emmerichs, and H. J. Bakker, “Femtosecond Mid-IR Pump-Probe Spectroscopy of Liquid Water: Evidence for a Two-Component Structure,” *Science*, vol. 278, pp. 658–660, Oct. 1997.
- [176] S. Roke and G. Gonella, “Nonlinear Light Scattering and Spectroscopy of Particles and Droplets in Liquids,” *Annual Review of Physical Chemistry*, vol. 63, no. 1, pp. 353–378, 2012.
- [177] Y. R. Shen, “Optical Second Harmonic Generation at Interfaces,” *Annual Review of Physical Chemistry*, vol. 40, pp. 327–350, Oct. 1989.
- [178] H. Bakker and J. Skinner, “Vibrational spectroscopy as a probe of structure and dynamics in liquid water,” *Chemical reviews*, vol. 110, no. 3, pp. 1498–1517, 2009.
- [179] “Map: Principles of Instrumental Analysis (Skoog et al.),” Jan. 2017.
- [180] X. Xue, Z.-Z. He, and J. Liu, “Detection of water–ice phase transition based on Raman spectrum,” *Journal of Raman Spectroscopy*, vol. 44, no. 7, pp. 1045–1048, 2013.
- [181] Y. Han, J. Liu, L. Huang, X. He, and J. Li, “Predicting the phase diagram of solid carbon dioxide at high pressure from first principles,” *npj Quantum Materials*, vol. 4, pp. 1–7, Mar. 2019.
- [182] G. Li and H. Guo, “The Vibrational Level Spectrum of $\text{H}_2\text{O}(\tilde{X}^1\text{A})$ on a Partridge–Schwenke Potential up to the Dissociation Limit,” *Journal of Molecular Spectroscopy*, vol. 210, pp. 90–97, Nov. 2001.

-
- [183] O. Vendrell, F. Gatti, and H.-D. Meyer, "Full dimensional (15-dimensional) quantum-dynamical simulation of the protonated water dimer. II. Infrared spectrum and vibrational dynamics," *The Journal of Chemical Physics*, vol. 127, p. 184303, Nov. 2007.
- [184] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller, "Ring-Polymer Molecular Dynamics: Quantum Effects in Chemical Dynamics from Classical Trajectories in an Extended Phase Space," *Annual Review of Physical Chemistry*, vol. 64, no. 1, pp. 387–413, 2013.
- [185] Y. Wang and J. M. Bowman, "Ab initio potential and dipole moment surfaces for water. II. Local-monomer calculations of the infrared spectra of water clusters," *The Journal of Chemical Physics*, vol. 134, p. 154510, Apr. 2011.
- [186] J. Cao and G. A. Voth, "The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties," *The Journal of Chemical Physics*, vol. 100, pp. 5106–5117, Apr. 1994.
- [187] G. R. Medders, A. W. Götz, M. A. Morales, P. Bajaj, and F. Paesani, "On the representation of many-body interactions in water," *The Journal of Chemical Physics*, vol. 143, p. 104102, Sept. 2015.
- [188] Y. Litman, J. O. Richardson, T. Kumagai, and M. Rossi, "Elucidating the Nuclear Quantum Dynamics of Intramolecular Double Hydrogen Transfer in Porphycene," *Journal of the American Chemical Society*, vol. 141, pp. 2526–2534, Feb. 2019.
- [189] O. Marsalek and T. E. Markland, "Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects," *The Journal of Physical Chemistry Letters*, vol. 8, pp. 1545–1551, Apr. 2017.
- [190] T. E. Markland and M. Ceriotti, "Nuclear quantum effects enter the mainstream," *Nature Reviews Chemistry*, vol. 2, pp. 1–14, Feb. 2018.
- [191] F. Brieuc, H. Dammak, and M. Hayoun, "Quantum Thermal Bath for Path Integral Molecular Dynamics Simulation," *J. Chem. Theory Comput.*, vol. 12, pp. 1351–1359, 2016.

Bibliography

- [192] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, “Nuclear quantum effects and hydrogen bond fluctuations in water,” *Proceedings of the National Academy of Sciences*, vol. 110, pp. 15591–15596, Sept. 2013.
- [193] M. Liu, L. Zhang, M. A. Little, V. Kapil, M. Ceriotti, S. Yang, L. Ding, D. L. Holden, R. Balderas-Xicohténcatl, D. He, R. Clowes, S. Y. Chong, G. Schütz, L. Chen, M. Hirscher, and A. I. Cooper, “Barely porous organic cages for hydrogen isotope separation,” *Science*, vol. 366, pp. 613–620, Nov. 2019.
- [194] F. Uhl, D. Marx, and M. Ceriotti, “Accelerated path integral methods for atomistic simulations at ultra-low temperatures,” *The Journal of Chemical Physics*, vol. 145, p. 054101, Aug. 2016.
- [195] C. W. Gardiner, *Handbook of Stochastic Methods*. Berlin: Springer, third ed., 2003.
- [196] V. Kapil, J. Behler, and M. Ceriotti, “High order path integrals made easy,” *J. Chem. Phys.*, vol. 145, p. 234103, 2016.
- [197] M. E. Daube-Witherspoon and G. Muehllehner, “An Iterative Image Space Reconstruction Algorithm Suitable for Volume ECT,” *IEEE Transactions on Medical Imaging*, vol. 5, pp. 61–66, 1986.
- [198] G. Archer and D. Titterton, “The iterative image space reconstruction algorithm (ISRA) as an alternative to the EM algorithm for solving positive linear inverse problems,” *Statistica Sinica*, vol. 5, pp. 77–96, 1995.
- [199] F. Krajewski and M. Parrinello, “Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces,” *Phys. Rev. B*, vol. 73, p. 041105, 2006.
- [200] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, “Efficient and Accurate Car-Parrinello-like Approach to Born-Oppenheimer Molecular Dynamics,” *Phys. Rev. Lett.*, vol. 98, p. 66401, 2007.
- [201] R. Baer, D. Neuhauser, and E. Rabani, “Self-Averaging Stochastic Kohn-Sham Density-Functional Theory,” *Phys. Rev. Lett.*, vol. 111, p. 106402, 2013.

-
- [202] G. Mazzola, S. Yunoki, and S. Sorella, “Unexpectedly high pressure for molecular dissociation in liquid hydrogen by electronic simulation,” *Nature Comm.*, vol. 5, p. 3487, 2014.
- [203] M. Ceriotti, G. Bussi, and M. Parrinello, “Nuclear Quantum Effects in Solids Using a Colored-Noise Thermostat,” *Phys. Rev. Lett.*, vol. 103, p. 30603, 2009.
- [204] H. Dammak, Y. Chalopin, M. Laroche, M. Hayoun, and J.-J. Greffet, “Quantum Thermal Bath for Molecular Dynamics Simulation,” *Phys. Rev. Lett.*, vol. 103, p. 190601, 2009.
- [205] S. Habershon and D. E. Manolopoulos, “Zero point energy leakage in condensed phase dynamics: An assessment of quantum simulation methods for liquid water,” *J. Chem. Phys.*, vol. 131, p. 244518, 2009.
- [206] A. Witt, S. D. Ivanov, M. Shiga, H. Forbert, and D. Marx, “On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy,” *The Journal of Chemical Physics*, vol. 130, p. 194510, May 2009.
- [207] S. Habershon, T. E. Markland, and D. E. Manolopoulos, “Competing quantum effects in the dynamics of a flexible water model,” *The Journal of Chemical Physics*, vol. 131, p. 024501, July 2009.
- [208] M. Rossi, H. Liu, F. Paesani, J. Bowman, and M. Ceriotti, “Communication: On the consistency of approximate quantum dynamics simulation methods for vibrational spectra in the condensed phase,” *The Journal of Chemical Physics*, vol. 141, p. 181101, Nov. 2014.
- [209] S. D. Ivanov, A. Witt, M. Shiga, and D. Marx, “Communications: On artificial frequency shifts in infrared spectra obtained from centroid molecular dynamics: Quantum liquid water,” *The Journal of Chemical Physics*, vol. 132, p. 031101, Jan. 2010.
- [210] P. Gallo and H. E. Stanley, “Supercooled water reveals its secrets,” *Science*, vol. 358, no. 6370, pp. 1543–1544, 2017.
- [211] M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. Morales, and T. E. Markland, “Nuclear Quantum Effects in Water and Aqueous Systems: Experiment, Theory, and Current Challenges,” *Chem. Rev.*, vol. 116, pp. 7529–7550, 2016.

Bibliography

- [212] C. Andreani, M. Krzystyniak, G. Romanelli, R. Senesi, and F. Fernandez-Alonso, “Electron-volt neutron spectroscopy: beyond fundamental systems,” *Advances in Physics*, vol. 66, no. 1, pp. 1–73, 2017.
- [213] C. Andreani, R. Senesi, M. Krzystyniak, G. Romanelli, and F. Fernandez-Alonso, “Atomic quantum dynamics in materials research,” in *Neutron Scattering - Applications in Biology, Chemistry, and Materials Science*, pp. 403–457, Elsevier, 2017.
- [214] A. Pietropaolo, R. Senesi, C. Andreani, A. Botti, M. a. Ricci, and F. Bruni, “Excess of Proton Mean Kinetic Energy in Supercooled Water,” *Phys. Rev. Lett.*, vol. 100, pp. 2–5, 2008.
- [215] D. Flammini, M. a. Ricci, and F. Bruni, “A new water anomaly: the temperature dependence of the proton mean kinetic energy,” *J. Chem. Phys.*, vol. 130, p. 236101, 2009.
- [216] R. Ramírez and C. P. Herrero, “Kinetic energy of protons in ice Ih and water: A path integral study,” *Phys. Rev. B*, vol. 84, p. 064130, 2011.
- [217] C. J. Burnham, T. Hayashi, R. L. Napoleon, T. Keyes, S. Mukamel, and G. F. Reiter, “The proton momentum distribution in strongly h-bonded phases of water: A critical test of electrostatic models,” *The Journal of Chemical Physics*, vol. 135, p. 144502, oct 2011.
- [218] B. Cheng, J. Behler, and M. Ceriotti, “Nuclear Quantum Effects in Water at the Triple Point: Using Theory as a Link Between Experiments,” *J. Phys. Chem. Letters*, vol. 7, pp. 2210–2215, 2016.
- [219] C. Pantalei, A. Pietropaolo, R. Senesi, S. Imberti, C. Andreani, J. Mayers, C. Burnham, and G. Reiter, “Proton momentum distribution of liquid water from room temperature to the supercritical phase,” *Phys. Rev. Lett.*, vol. 100, p. 177801, 2008.
- [220] M. Ceriotti, G. Miceli, A. Pietropaolo, D. Colognesi, A. Nale, M. Catti, M. Bernasconi, and M. Parrinello, “Nuclear quantum effects in ab initio dynamics: Theory and experiments for lithium imide,” *Phys. Rev. B*, vol. 82, p. 174306, 2010.
- [221] R. Senesi, D. Flammini, A. I. Kolesnikov, É. D. Murray, G. Galli, and C. Andreani, “The quantum nature of the OH stretching mode in ice and water probed by neutron scattering experiments,” *J. Chem. Phys.*, vol. 139, p. 074504, 2013.

-
- [222] S. Habershon, T. E. Markland, and D. E. Manolopoulos, "Competing quantum effects in the dynamics of a flexible water model.," *J. Chem. Phys.*, vol. 131, p. 24501, 2009.
- [223] V. Babin, C. Leforestier, and F. Paesani, "Development of a "first principles" water potential with flexible monomers: Dimer potential energy surface, VRT spectrum, and second virial coefficient," *Journal of Chemical Theory and Computation*, vol. 9, pp. 5395–5403, Nov. 2013.
- [224] J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146401, 2007.
- [225] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der Waals interactions determine the unique properties of water," *Proc. Natl. Acad. Sci. USA*, vol. 113, pp. 8368–8373, 2016.
- [226] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *The Journal of Physical Chemistry*, vol. 98, pp. 11623–11627, nov 1994.
- [227] G. Imbalzano, A. Anelli, D. Giofr , S. Klees, J. Behler, and M. Ceriotti, "Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials," *J. Chem. Phys.*, vol. 148, p. 241730, 2018.
- [228] R. Senesi, G. Romanelli, M. A. Adams, and C. Andreani, "Temperature dependence of the zero point kinetic energy in ice and water above room temperature," *Chemical Physics*, vol. 427, pp. 111–116, 2013.
- [229] C. Andreani and R. Senesi
- [230] M. Ceriotti, "Ab initio simulation of particle momentum distributions in high-pressure water," *J. Phys. Conf. Ser.*, vol. 571, p. 012011, 2014.
- [231] G. R. Medders, A. W. G tz, M. A. Morales, P. Bajaj, and F. Paesani, "On the representation of many-body interactions in water," *J. Chem. Phys.*, vol. 143, p. 104102, 2015.
- [232] S. K. Reddy, D. R. Moberg, S. C. Straight, and F. Paesani, "Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations

Bibliography

- with the MB-pol potential energy function,” *The Journal of Chemical Physics*, vol. 147, p. 244504, dec 2017.
- [233] F. Paesani, “Getting the right answers for the right reasons: Toward predictive molecular simulations of water with many-body potential energy functions,” *Accounts of Chemical Research*, vol. 49, pp. 1844–1851, aug 2016.
- [234] E. Grifoni, G. Piccini, and M. Parrinello, “Microscopic description of acid–base equilibrium,” *Proceedings of the National Academy of Sciences*, vol. 116, pp. 4054–4057, Feb. 2019.
- [235] L. Li, T. Totton, and D. Frenkel, “Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials,” *The Journal of Chemical Physics*, vol. 149, p. 054102, Aug. 2018.
- [236] A. Grisafi, D. M. Wilkins, G. Csányi, and M. Ceriotti, “Symmetry-adapted machine learning for tensorial properties of atomistic systems,” *Phys. Rev. Lett.*, vol. 120, p. 036002, 2018.
- [237] D. M. Wilkins, A. Grisafi, Y. Yang, K. U. Lao, R. A. DiStasio, and M. Ceriotti, “Accurate molecular polarizabilities with coupled cluster theory and machine learning,” *Proc. Natl. Acad. Sci.*, vol. 116, p. 3401, 2019.
- [238] N. Raimbault, A. Grisafi, M. Ceriotti, and M. Rossi, “Using gaussian process regression to simulate the vibrational raman spectra of molecular crystals,” *New J. Phys.*, vol. 21, p. 105001, 2019.
- [239] R. Resta and D. Vanderbilt, “Theory of Polarization: A Modern Approach,” in *Physics of Ferroelectrics: A Modern Perspective*, Topics in Applied Physics, pp. 31–68, Berlin, Heidelberg: Springer, 2007.
- [240] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, “Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials,” *The Journal of Chemical Physics*, vol. 148, p. 241730, Apr. 2018.

- [241] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, “Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons,” *Physical Review Letters*, vol. 104, p. 136403, Apr. 2010.
- [242] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.
- [243] A. P. Bartók, R. Kondor, and G. Csányi, “On representing chemical environments,” *Physical Review B*, vol. 87, p. 184115, May 2013.
- [244] M. J. Willatt, F. Musil, and M. Ceriotti, “Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements,” *Physical Chemistry Chemical Physics*, vol. 20, pp. 29661–29668, Dec. 2018.
- [245] A. V. Gubskaya and P. G. Kusalik, “The total molecular dipole moment for liquid water,” *The Journal of Chemical Physics*, vol. 117, pp. 5290–5302, Aug. 2002.
- [246] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, “Ab initio thermodynamics of liquid and solid water,” *Proceedings of the National Academy of Sciences*, vol. 116, pp. 1110–1115, Jan. 2019.
- [247] M. H. Brooker, G. Hancock, B. C. Rice, and J. Shapter, “Raman frequency and intensity studies of liquid H₂O, H₂¹⁸O and D₂O,” *Journal of Raman Spectroscopy*, vol. 20, no. 10, pp. 683–694, 1989.
- [248] J. E. Bertie and Z. Lan, “Infrared Intensities of Liquids XX: The Intensity of the OH Stretching Band of Liquid Water Revisited, and the Best Current Values of the Optical Constants of H₂O(l) at 25°C between 15,000 and 1 cm⁻¹,” *Applied Spectroscopy*, vol. 50, pp. 1047–1057, Aug. 1996.
- [249] D. R. Moberg, S. C. Straight, C. Knight, and F. Paesani, “Molecular Origin of the Vibrational Structure of Ice Ih,” *The Journal of Physical Chemistry Letters*, vol. 8, pp. 2579–2583, June 2017.
- [250] X. Liu and J. Liu, “Critical role of quantum dynamical effects in the raman spectroscopy of liquid water,” *Molecular Physics*, vol. 116, pp. 755–779, Feb. 2018.

Bibliography

- [251] A. G. Slater and A. I. Cooper, "Function-led design of new porous materials," *Science*, vol. 348, p. aaa8075, 2015.
- [252] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi, "The chemistry and applications of metal-organic frameworks," *Science*, vol. 341, p. 1230444, 2013.
- [253] J. A. Mason, M. Veenstra, and J. R. Long, "Evaluating metal-organic frameworks for natural gas storage," *Chem. Sci.*, vol. 5, pp. 32–51, 2014.
- [254] Y. He, F. Chen, B. Li, G. Qian, W. Zhou, and B. Chen, "Porous metal-organic frameworks for fuel storage," *Coord. Chem. Rev.*, vol. 373, pp. 167–198, 2017.
- [255] M. F. de Lange, K. J. F. M. Verouden, T. J. H. Vlugt, J. Gascon, and F. Kapteijn, "Adsorption-driven heat pumps: The potential of metal-organic frameworks," *Chem. Rev.*, vol. 115, pp. 12205–12250, 2015.
- [256] S. Wang, J. S. Lee, M. Wahiduzzaman, J. Park, M. Muschi, C. Martineau-Corcus, A. Tissot, K. H. Cho, J. Marrot, W. Shepard, G. Maurin, J.-S. Chang, and C. Serre, "A robust large-pore zirconium carboxylate metal-organic framework for energy-efficient water-sorption-driven refrigeration," *Nat. Energy*, vol. 3, pp. 985–993, 2018.
- [257] K. Sumida, D. L. Rogow, J. A. Mason, T. M. McDonald, E. D. Bloch, Z. R. Herm, T.-H. Bae, and J. R. Long, "Carbon dioxide capture in metal-organic frameworks," *Chem. Rev.*, vol. 112, pp. 724–781, 2012.
- [258] C. A. Trickett, A. Helal, B. A. Al-Maythaly, Z. H. Zamani, K. E. Cordova, and O. M. Yaghi, "The chemistry of metal-organic frameworks for CO₂ capture, regeneration and conversion," *Nat. Rev. Mater.*, vol. 2, p. 17045, 2017.
- [259] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, and B. Smit, "The materials genome in action: identifying the performance limits for methane storage," *Energy Environ. Sci.*, vol. 8, pp. 1190–1199, 2015.
- [260] J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, P. L. Llewellyn, N. Masciocchi, and

- J. R. Long, "Methane storage in flexible metal-organic frameworks with intrinsic thermal management," *Nature*, vol. 527, pp. 357–361, 2015.
- [261] J. M. Huck, L.-C. Lin, A. H. Berger, M. N. Shahrak, R. L. Martin, A. S. Bhowm, M. Haranczyk, K. Reuter, and B. Smit, "Evaluating different classes of porous materials for carbon capture," *Energy Environ. Sci.*, vol. 7, pp. 4132–4146, 2014.
- [262] F. A. Kloutse, R. Zacharia, D. Cossement, and R. Chahine, "Specific heat capacities of MOF-5, Cu-BTC, Fe-BTC, MOF-177 and MIL-53(Al) over wide temperature ranges: Measurements and application of empirical group contribution method," *Microporous Mesoporous Mater.*, vol. 217, pp. 1–5, 2015.
- [263] H. Babaei and C. E. Wilmer, "Mechanisms of heat transfer in porous crystals containing adsorbed gases: Applications to metal-organic frameworks," *Phys. Rev. Lett.*, vol. 116, p. 025902, 2016.
- [264] S. R. G. Balestra, R. Bueno-Perez, S. Hamad, D. Dubbeldam, A. R. Ruiz-Salvador, and S. Calero, "Controlling thermal expansion: A metal-organic frameworks route," *Chem. Mater.*, vol. 28, pp. 8296–8304, 2016.
- [265] S. M. J. Rogge, L. Vanduyfhuys, A. Ghysels, M. Waroquier, T. Verstraelen, G. Maurin, and V. Van Speybroeck, "A comparison of barostats for the mechanical characterization of metal-organic frameworks," *J. Chem. Theory Comput.*, vol. 11, pp. 5583–5597, 2015.
- [266] L. Vanduyfhuys, S. M. J. Rogge, J. Wieme, S. Vandenbrande, G. Maurin, M. Waroquier, and V. Van Speybroeck, "Thermodynamic insight into stimuli-responsive behaviour of soft porous crystals," *Nat. Commun.*, vol. 9, p. 204, 2018.
- [267] J. Wieme, K. Lejaeghere, G. Kresse, and V. Van Speybroeck, "Tuning the balance between dispersion and entropy to design temperature-responsive flexible metal-organic frameworks," *Nat. Commun.*, vol. 9, p. 4899, 2018.
- [268] F. Paesani, "Water in metal-organic frameworks: structure and diffusion of H_2O in mil-53(cr) from quantum simulations," *Mol. Simul.*, vol. 38, pp. 631–641, 2012.

Bibliography

- [269] D. D. Borges, R. Semino, S. Devautour-Vinot, H. Jobic, F. Paesani, and G. Maurin, "Computational exploration of the water concentration dependence of the proton transport in the porous uio-66(zr)-(co₂h)₂ metal-organic framework," *Chem. Mater.*, vol. 29, pp. 1569–1576, 2017.
- [270] A. Lamaire, J. Wieme, S. M. J. Rogge, M. Waroquier, and V. Van Speybroeck, "On the importance of anharmonicities and nuclear quantum effects in modelling the structural properties and thermal expansion in MOF-5," *J. Chem. Phys.*, vol. 150, p. 094503, 2019.
- [271] D. Chandler and P. G. Wolynes, "Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids," *J. Chem. Phys.*, vol. 74, pp. 4078–4095, 1981.
- [272] M. Ceriotti and D. E. Manolopoulos, "Efficient first-principles calculation of the quantum kinetic energy and momentum distribution of nuclei," *Phys. Rev. Lett.*, vol. 109, p. 100604, sep 2012.
- [273] T. E. Markland and M. Ceriotti, "Nuclear quantum effects enter the mainstream," *Nat. Rev. Chem.*, vol. 2, p. 0109, 2018.
- [274] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, and V. Van Speybroeck, "QuickFF: A program for a quick and easy derivation of force fields for metal-organic frameworks from ab initio input," *J. Comput. Chem.*, vol. 36, pp. 1015–1027, 2015.
- [275] L. Vanduyfhuys, S. Vandenbrande, J. Wieme, M. Waroquier, T. Verstraelen, and V. Van Speybroeck, "Extension of the quickFF force field protocol for an improved accuracy of structural, vibrational, mechanical and thermal properties of metal-organic frameworks," *J. Comput. Chem.*, vol. 39, pp. 999–1011, 2018.
- [276] H. Li, M. Eddaoudi, M. O’Keeffe, and O. M. Yaghi, "Design and synthesis of an exceptionally stable and highly porous metal-organic framework," *Nature*, vol. 402, pp. 276–279, 1999.
- [277] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O’Keeffe, and O. M. Yaghi, "Sys-

- tematic design of pore size and functionality in isorecticular mofs and their application in methane storage,” *Science*, vol. 295, pp. 469–472, 2002.
- [278] W. Zhou, H. Wu, M. R. Hartman, and T. Yildirim, “Hydrogen and methane adsorption in metal-organic frameworks: A high-pressure volumetric study,” *J. Phys. Chem. C*, vol. 111, pp. 16131–16137, 2007.
- [279] S. Vandenbrande, T. Verstraelen, J. J. Gutiérrez-Sevillano, M. Waroquier, and V. Van Speybroeck, “Methane adsorption in zr-based mofs: Comparison and critical evaluation of force fields,” *J. Phys. Chem. C*, vol. 121, pp. 25309–25322.
- [280] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, “Gaussian16 Revision B.01,” 2016. Gaussian Inc. Wallingford CT.
- [281] A. D. Becke, “Density-functional thermochemistry. iii. the role of exact exchange,” *J. Chem. Phys.*, vol. 98, no. 7, pp. 5648–5652, 1993.
- [282] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, “Self-consistent molecular orbital methods. 20. basis set for correlated wave-functions,” *J. Chem. Phys.*, vol. 72, no. 1, pp. 650–654, 1980.
- [283] P. J. Hay and W. R. Wadt, “Ab initio effective core potentials for molecular calculations. potentials for the transition metal atoms sc to hg,” *J. Chem. Phys.*, vol. 82, p. 270, 1985.
- [284] T. Verstraelen, S. Vandenbrande, F. Heidar-Zadeh, L. Vanduyfhuys, V. Van Speybroeck,

Bibliography

- M. Waroquier, and P. W. Ayers, "Minimal basis iterative stockholder: Atoms in molecules for force-field development," *J. Chem. Theory Comput.*, vol. 12, pp. 3894–3912, 2016.
- [285] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Phys. Rev. Lett.*, vol. 77, p. 3865, 1996.
- [286] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, "Real-space grid implementation of the projector augmented wave method," *Phys. Rev. B*, vol. 71, p. 035109, 2005.
- [287] J. H. Lii and N. L. Allinger, "Molecular mechanics. the mm3 force field for hydrocarbons. 3. the van der waals potentials and crystal data for aliphatic and aromatic hydrocarbons," *J. Am. Chem. Soc.*, vol. 111, no. 23, pp. 8576–8582, 1989.
- [288] N. L. Allinger, X. Zhou, and J. Bergsma, "Molecular mechanics parameters," *J. Mol. Struc.-THEOCHEM*, vol. 312, no. 1, pp. 69–83, 1994.
- [289] H. Sun, "Compass: An ab initio force-field optimized for condensed-phase applications - overview with details on alkane and benzene compounds," *J. Phys. Chem. B*, vol. 102, pp. 7338–7364, 1998.
- [290] D. Dubbeldam, S. Calero, D. E. Ellis, and R. Q. Snurr, "Raspa: molecular simulation software for adsorption and diffusion in flexible nanoporous materials," *Mol. Simul.*, vol. 42, pp. 81–101, 2016.
- [291] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.*, vol. 117, pp. 1–19, 1995.
- [292] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Mol. Phys.*, vol. 52, no. 2, pp. 255–268, 1984.
- [293] W. G. Hoover, "Canonical Dynamics: Equilibrium Phase-Space Distributions," *Phys. Rev. A*, vol. 31, no. 3, pp. 1695–1697, 1985.
- [294] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nose-Hoover Chains - the Canonical Ensemble Via Continuous Dynamics," *J. Chem. Phys.*, vol. 97, no. 4, pp. 2635–2643, 1992.
- [295] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Mol. Phys.*, vol. 87, pp. 1117–1157, 1996.

- [296] G. Bussi, T. Zykova-Timan, and M. Parrinello, "Isothermal-isobaric molecular dynamics using stochastic velocity rescaling," *J. Chem. Phys.*, vol. 130, p. 074101, 2009.
- [297] B. Leimkuhler and C. Matthews, "Robust and efficient configurational molecular sampling via Langevin dynamics," *J. Chem. Phys.*, vol. 138, 2013.
- [298] E. Balog, A. L. Hughes, and G. J. Martyna, "Constant pressure path integral molecular dynamics studies of quantum effects in the liquid state properties of n-alkanes," *J. Chem. Phys.*, vol. 112, pp. 870–880, jan 2000.
- [299] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, "On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, pp. 8878–8882, aug 2018.
- [300] M. Veit, S. K. Jain, S. Bonakala, I. Rudra, D. Hohl, and G. Csányi, "Equation of state of fluid methane from first principles with machine learning potentials," *J. Chem. Theory Comput.*, vol. 15, pp. 2574–2586, 2019.
- [301] P. I. Ravikovitch and A. V. Neimark, "Density functional theory model of adsorption deformation," *Langmuir*, vol. 22, pp. 10864–10868, 2006.
- [302] J. Joo, H. Kim, and S. S. Han, "Volume shrinkage of a metal-organic framework host induced by the dispersive attraction of guest gas molecules," *Phys. Chem. Chem. Phys.*, vol. 15, pp. 18822–18826, 2013.
- [303] W. Zhou, H. Wu, T. Yildirim, J. R. Simpson, and A. R. Hight Walker, "Origin of the exceptional negative thermal expansion in metal-organic framework-5 $\text{Zn}_4\text{O}(\text{1,4-benzenedicarboxylate})_3$," *Phys. Rev. B*, vol. 78, p. 054114, 2008.
- [304] N. Lock, Y. Wu, M. Christensen, L. J. Cameron, V. K. Peterson, A. J. Bridgeman, C. J. Kepert, and B. B. Iversen, "Elucidating negative thermal expansion in mof-5," *J. Phys. Chem. C*, vol. 114, pp. 16181–16186, 2010.
- [305] N. Lock, M. Christensen, Y. Wu, V. K. Peterson, M. K. Thomsen, R. O. Plitz, A. J. Ramirez-Cuesta, G. J. McIntyre, K. Norén, R. Kutteh, C. J. Kepert, G. J. Kearley, and B. B. Iversen,

Bibliography

- “Scrutinizing negative thermal expansion in mof-5 by scattering techniques and ab initio calculations,” *Dalton Trans.*, vol. 42, pp. 1996–2007, 2013.
- [306] Y. Ming, J. Purewal, D. Liu, A. Sudik, C. Xu, J. Yang, M. Veenstra, K. Rhodes, R. Soltis, J. Warner, M. Gaab, U. Müller, and D. J. Siegel, “Thermophysical properties of mof-5 powders,” *Microporous Mesoporous Mater.*, vol. 185, pp. 235–244, 2014.
- [307] D. Liu, J. J. Purewal, J. Yang, A. Sudik, S. Maurer, U. Müller, J. Ni, and D. J. Siegel, “Mof-5 composites exhibiting improved thermal conductivity,” *Int. J. Hydrogen Ener.*, vol. 37, pp. 6109–6117, 2012.
- [308] D. P. Tabor, L. M. Roch, S. K. Saikin, C. Kreisbeck, D. Sheberla, J. H. Montoya, S. Dwarak-nath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C. J. Brabec, B. Maruyama, K. A. Persson, and A.-G. A., “Accelerating the discovery of materials for clean energy in the era of smart automation,” *Nat. Rev. Mater.*, vol. 3, pp. 5–20, 2018.
- [309] E. Tsivion and M. Head-Gordon, “Methane storage: molecular mechanisms underlying room-temperature adsorption in $\text{Zn}_4\text{O}(\text{BDC})_3$ (MOF-5),” *J. Phys. Chem. C*, vol. 121, pp. 12091–12100, 2017.
- [310] H. Wu, W. Zhou, and T. Yildirim, “Methane sorption in nanoporous metal-organic frameworks and first-order phase transition of confined methane,” *J. Phys. Chem. C*, vol. 113, pp. 3029–3035, 2009.
- [311] B. Kuchta, E. Dundar, F. Formalik, P. L. Llewellyn, and L. Firley, “Adsorption-induced structural phase transformation in nanopores,” *Angew. Chem. Int. Ed.*, vol. 56, pp. 16243–16246, 2017.
- [312] M. Levitt, M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett, “Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution,” *J. Phys. Chem. B*, vol. 101, pp. 5051–5061, jun 1997.
- [313] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, “Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. II. linear scaling do-main based pair natural orbital coupled cluster theory,” *The Journal of Chemical Physics*, vol. 144, p. 024109, Jan. 2016.

-
- [314] T. Gruber, K. Liao, T. Tsatsoulis, F. Hummel, and A. Grüneis, “Applying the coupled-cluster ansatz to solids and surfaces in the thermodynamic limit,” *Physical Review X*, vol. 8, May 2018.
- [315] E. Caldeweyher and J. G. Brandenburg, “Simplified DFT methods for consistent structures and energies of large systems,” *Journal of Physics: Condensed Matter*, vol. 30, p. 213001, May 2018.
- [316] A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides, “Fast and accurate quantum monte carlo for molecular crystals,” *Proceedings of the National Academy of Sciences*, vol. 115, pp. 1724–1729, Feb. 2018.
- [317] J. Hoja, H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko, “Reliable and practical computational description of molecular crystal polymorphs,” *Science Advances*, vol. 5, p. eaau3338, jan 2019.
- [318] B. Grabowski, L. Ismer, T. Hickel, and J. Neugebauer, “Ab initio up to the melting point: Anharmonicity and vacancies in aluminum,” *Physical Review B*, vol. 79, apr 2009.
- [319] G. Grimvall, B. Magyari-Köpe, V. Ozolinš, and K. A. Persson, “Lattice instabilities in metallic elements,” *Reviews of Modern Physics*, vol. 84, p. 945, 2012.
- [320] A. Glensk, B. Grabowski, T. Hickel, and J. Neugebauer, “Breakdown of the arrhenius law in describing vacancy formation energies: The importance of local anharmonicity revealed by Ab initio Thermodynamics,” *Physical Review X*, vol. 4, feb 2014.
- [321] X.-Z. Li, B. Walker, and A. Michaelides, “Quantum nature of the hydrogen bond,” *Proceedings of the National Academy of Sciences*, vol. 108, pp. 6369–6373, apr 2011.
- [322] W. Fang, J. Chen, M. Rossi, Y. Feng, X.-Z. Li, and A. Michaelides, “Inverse temperature dependence of nuclear quantum effects in DNA base pairs,” *The Journal of Physical Chemistry Letters*, vol. 7, pp. 2125–2131, may 2016.
- [323] S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, “Dissociation of High-Pressure Solid Molecular Hydrogen: A Quantum Monte Carlo and Anharmonic Vibrational Study,” *Physical Review Letters*, vol. 112, p. 165501, 2014.

Bibliography

- [324] R. Bianco, I. Errea, M. Calandra, and F. Mauri, “High-pressure phase diagram of hydrogen and deuterium sulfides from first principles: Structural and vibrational properties including quantum and anharmonic effects,” *Physical Review B*, vol. 97, p. 214101, 2018.
- [325] W. Fang, J. Chen, Y. Feng, X.-Z. Li, and A. Michaelides, “The quantum nature of hydrogen,” *International Reviews in Physical Chemistry*, vol. 38, pp. 35–61, jan 2019.
- [326] R. Ramírez and C. P. Herrero, “Quantum path integral simulation of isotope effects in the melting temperature of ice Ih,” *Journal of Chemical Physics*, vol. 133, p. 144511, 2010.
- [327] E. A. Engel, B. Monserrat, and R. J. Needs, “Anharmonic Nuclear Motion and the Relative Stability of Hexagonal and Cubic ice,” *Physical Review X*, vol. 5, p. 021033, 2015.
- [328] B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, “Ab initio thermodynamics of liquid and solid water,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 116, p. 1110, 2019.
- [329] J. Lin, L. Guo, Q. Huang, Y. Jia, K. Li, X. Lai, and X. Chen, “Anharmonic phonon effects in Raman spectra of unsupported vertical graphene sheets,” *Physical Review B*, vol. 83, p. 125430, 2011.
- [330] R. Ramírez, E. Chacon, and C. P. Herrero, “Anharmonic effects in the optical and acoustic bending modes of graphene,” *Physical Review B*, vol. 93, p. 235419, 2016.
- [331] D. Chandler and P. G. Wolynes, “Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids,” *The Journal of Chemical Physics*, vol. 74, pp. 4078–4095, apr 1981.
- [332] N. Ashcroft and N. Mermin, “Solid State Physics,” 1976.
- [333] S. E. Brown, I. Georgescu, and V. A. Mandelshtam, “Self-consistent phonons revisited. II. A general and efficient method for computing free energies and vibrational spectra of molecules and clusters,” *Journal of Chemical Physics*, vol. 138, p. 044317, 2013.
- [334] I. Errea, M. Calandra, and F. Mauri, “First-principles theory of anharmonicity and the inverse isotope effect in superconducting palladium-hydride compounds,” *Physical Review Letters*, vol. 111, Oct. 2013.

- [335] B. Monserrat, N. D. Drummond, and R. J. Needs, “Anharmonic vibrational properties in periodic systems: energy, electron-phonon coupling, and stress,” *Physical Review B*, vol. 87, p. 144302, 2013.
- [336] C. McBride, E. G. Noya, J. L. Aragones, M. M. Conde, and C. Vega, “The phase diagram of water from quantum simulations,” *Physical Chemistry Chemical Physics*, vol. 14, no. 29, p. 10140, 2012.
- [337] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, and M. Ceriotti, “i-PI 2.0: A universal force engine for advanced molecular simulations,” *Computer Physics Communications*, vol. 236, p. 214, 2018.
- [338] D. A. McQuarrie and J. D. Simon, *Physical Chemistry: A Molecular Approach*. University Science Books, 1 ed., 1997.
- [339] B. Callen, *Thermodynamics and an Introduction to Thermostatistics*. Wiley, 2 ed., 1985.
- [340] D. Hooton, “LI. A new treatment of anharmonicity in lattice thermodynamics: I,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 7, vol. 46, p. 422, 1955.
- [341] D. Hooton, “LII. A new treatment of anharmonicity in lattice thermodynamics: II,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 7, vol. 46, p. 433, 1955.
- [342] J. G. Kirkwood, “Statistical mechanics of fluid mixtures,” *The Journal of Chemical Physics*, vol. 3, pp. 300–313, May 1935.
- [343] L. M. Ghiringhelli, J. H. Los, E. J. Meijer, A. Fasolino, and D. Frenkel, “Modeling the Phase Diagram of Carbon,” *Physical Review Letters*, vol. 94, p. 145701, 2005.
- [344] B. Cheng and M. Ceriotti, “Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids,” *Physical Review B*, vol. 97, p. 054102, 2018.

Bibliography

- [345] D. Frenkel and A. J. C. Ladd, “New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres,” *The Journal of Chemical Physics*, vol. 81, pp. 3188–3193, Oct. 1984.
- [346] S. Habershon and D. E. Manolopoulos, “Free energy calculations for a flexible water model,” *Physical Chemistry Chemical Physics*, vol. 13, no. 44, p. 19714, 2011.
- [347] W. G. Hoover, S. G. Gray, and K. W. Johnson, “Thermodynamic properties of the fluid and solid phases for inverse power potentials,” *The Journal of Chemical Physics*, vol. 55, pp. 1128–1136, Aug. 1971.
- [348] S. G. Moustafa, A. J. Schultz, and D. A. Kofke, “Very fast averaging of thermal properties of crystals by molecular simulation,” *Physical Review E*, vol. 92, oct 2015.
- [349] J. Morales and K. Singer, “Path integral simulation of the free energy of (lennard-jones) neon,” *Molecular Physics*, vol. 73, pp. 873–880, jul 1991.
- [350] S. Habershon and D. E. Manolopoulos, “Thermodynamic integration from classical to quantum mechanics,” *The Journal of Chemical Physics*, vol. 135, p. 224111, dec 2011.
- [351] M. Kolár and S. F. O’Shea, “A high-temperature approximation for the path-integral quantum monte carlo method,” *Journal of Physics A: Mathematical and General*, vol. 29, pp. 3471–3494, jul 1996.
- [352] J. D. Beasley and S. G. Springer, “Algorithm AS 111: The percentage points of the normal distribution,” *Applied Statistics*, vol. 26, no. 1, p. 118, 1977.
- [353] B. Moro, “The full monte,” *Risk*, vol. 8, no. 2, pp. 57–59, 1995.
- [354] I. M. Sobol, “Distribution of points in a cube and approximate evaluation of integrals,” *USSR Computational Mathematics and Mathematical Physics*, vol. 16, p. 236, 1977.
- [355] R. Schürer, “A comparison between (quasi-)monte carlo and cubature rule based methods for solving high-dimensional integration problems,” *Mathematics and Computers in Simulation*, vol. 62, pp. 509–517, Mar. 2003.
- [356] *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer New York, 2009.

-
- [357] “The Sobol Quasirandom Sequence, code for C++/Fortran 90/Matlab/Python.” http://people.sc.fsu.edu/~jburkardt/f_src/sobol/sobol.html. Accessed: 2018-02-05.
- [358] I. Errea, M. Calandra, and F. Mauri, “Anharmonic free energies and phonon dispersions from the stochastic self-consistent harmonic approximation: Application to platinum and palladium hydrides,” *Physical Review B*, vol. 89, p. 064302, 2014.
- [359] M. Ceriotti, G. A. R. Brain, O. Riordan, and D. E. Manolopoulos, “The inefficiency of re-weighted sampling and the curse of system size in high-order path integration,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 468, pp. 2–17, sep 2011.
- [360] S. L. Adler, “Over-relaxation method for the monte carlo evaluation of the partition function for multiquadratic actions,” *Physical Review D*, vol. 23, pp. 2901–2904, jun 1981.
- [361] M. Ceriotti, G. Bussi, and M. Parrinello, “Conjugate gradient heat bath for ill-conditioned actions,” *Physical Review E*, vol. 76, aug 2007.
- [362] T. Yamanaka and S. Morimoto, “Isotope effect on anharmonic thermal atomic vibration and k-refinement of ^{12}C and ^{13}C diamond,” *Acta Crystallographica, Section B: Structural Science*, vol. 52, p. 232, 1996.
- [363] F. P. Bundy and J. S. Kasper, “Hexagonal Diamond – A New Form of Carbon,” *Journal of Chemical Physics*, vol. 46, p. 3437, 1967.
- [364] V. L. Deringer and G. Csányi, “Machine learning based interatomic potential for amorphous carbon,” *Physical Review B*, vol. 95, p. 094203, 2017.
- [365] A. J. Leadbetter, R. C. Ward, J. W. Clark, P. A. Tucker, T. Matsuo, and S. Suga, “The equilibrium low-temperature structure of ice,” *Journal of Chemical Physics*, vol. 82, p. 424, 1985.
- [366] R. Howe and R. W. Whitworth, “A determination of the crystal structure of ice XI,” *Journal of Chemical Physics*, vol. 90, p. 4450, 1989.

Bibliography

- [367] P. Geiger, C. Dellago, M. Macher, C. Franchini, G. Kresse, J. Bernard, J. N. Stern, and T. Loerting, "Proton ordering of cubic ice ic: Spectroscopy and computer simulations," *The Journal of Physical Chemistry C*, vol. 118, pp. 10989–10997, May 2014.
- [368] J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," *Physical Review Letters*, vol. 98, p. 146401, 2007.
- [369] V. Kapil, J. Behler, and M. Ceriotti, "High order path integrals made easy," *Journal of Chemical Physics*, vol. 145, p. 234103, 2016.
- [370] M. Rossi, V. Kapil, and M. Ceriotti, "Fine tuning classical and quantum molecular dynamics using a generalized langevin equation," *The Journal of Chemical Physics*, vol. 148, p. 102301, mar 2018.
- [371] V. Kapil, A. Cuzzocrea, and M. Ceriotti, "Anisotropy of the proton momentum distribution in water," *The Journal of Physical Chemistry B*, vol. 122, pp. 6048–6054, may 2018.
- [372] M. Haisa, S. Kashino, and H. Maeda, "The orthorhombic form of p-hydroxyacetanilide," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 30, pp. 2510–2512, oct 1974.
- [373] M. Haisa, S. Kashino, R. Kawai, and H. Maeda, "The monoclinic form of p-hydroxyacetanilide," *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, vol. 32, pp. 1283–1285, apr 1976.
- [374] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. Adam Hilger NY, 1989.
- [375] M. Ceriotti and T. E. Markland, "Efficient methods and practical guidelines for simulating isotope effects," *J. Chem. Phys.*, vol. 138, p. 014112, 2013.
- [376] M. Rossi, W. Fang, and A. Michaelides, "Stability of Complex Biomolecular Structures: van der Waals, Hydrogen Bond Cooperativity, and Nuclear Quantum Effects," *J. Phys. Chem. Letters*, vol. 6, pp. 4233–4238, 2015.

-
- [377] L. Monacelli, I. Errea, M. Calandra, and F. Mauri, "Pressure and stress tensor of complex anharmonic crystals within the stochastic self-consistent harmonic approximation," *Physical Review B*, vol. 98, July 2018.
- [378] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physics Letters*, vol. 314, pp. 141–151, Nov. 1999.
- [379] R. Bianco, I. Errea, L. Paulatto, M. Calandra, and F. Mauri, "Second-order structural phase transitions, free energy curvature, and temperature-dependent anharmonic phonons in the self-consistent harmonic approximation: Theory and stochastic implementation," *Physical Review B*, vol. 96, p. 014111, 2017.
- [380] J. Hoja, A. M. Reilly, and A. Tkatchenko, "First-principles modeling of molecular crystals: structures and stabilities, temperature and pressure," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 7, p. e1294, Dec. 2016.
- [381] T. D. Kühne, M. Krack, F. R. Mohamed, and M. Parrinello, "Efficient and accurate car-parrinello-like approach to born-oppenheimer molecular dynamics," *Physical Review Letters*, vol. 98, Feb. 2007.
- [382] S. Mandal, J. Debnath, B. Meyer, and N. N. Nair, "Enhanced sampling and free energy calculations with hybrid functionals and plane waves for chemical reactions," *The Journal of Chemical Physics*, vol. 149, p. 144113, Oct. 2018.
- [383] A. I. Duff, T. Davey, D. Korbmacher, A. Glensk, B. Grabowski, J. Neugebauer, and M. W. Finnis, "Improved method of calculating ab initio high-temperature thermodynamic properties with application to ZrC," *Physical Review B*, vol. 91, Jun 2015.
- [384] O. Hellman, P. Steneteg, I. A. Abrikosov, and S. I. Simak, "Temperature dependent effective potential method for accurate free energy calculations of solids," *Physical Review B*, vol. 87, p. 104111, 2013.

Curriculum vitae

PERSONAL INFORMATION

NAME: Venkat Kapil
WORK ADDRESS: EPFL STI IMX COSMO, MXG 341, Station 12 CH-1015, Lausanne, Switzerland
HOME ADDRESS: Promenade du Jura, Nyon, 1260, Switzerland
EMAIL: venkat.kapil@epfl.ch, venkat.kapil@gmail.com[†]
GOOGLE SCHOLAR: <https://scholar.google.ch/citations?user=F2NMPNQAAAAJ>
ORCID: 0000-0003-0324-2198

EDUCATION AND EMPLOYMENT HISTORY

Oct. 2015 - Present PhD in Department of Materials Science and Engineering
 École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
 Thesis: “Nuclear Quantum Effects: Fast and Efficient”
 Supervisor: Prof. Michele Ceriotti

July 2010 - July 2015 Integrated M.Sc. in Chemistry
 Indian Institute of Technology Kanpur (IITK), Kanpur, India
 Thesis: “Sampling free energy surfaces as slices by combining
 umbrella sampling and metadynamics”
 Supervisor: Prof. Nisanth N. Nair

GRANTS AND AWARDS

1. Principal investigator for a 0.9 million GPU node hour project “Efficient and reliable crystal structure prediction using quantum Monte Carlo and Machine Learning” (CSCS, Switzerland, Mar. 2020)
[<https://www.cscs.ch/>]
2. Awarded the Early Postdoc Mobility Fellowship by the Swiss National Science Foundation (Nov. 2019)
[<http://www.snf.ch/en/funding/careers/early-postdoc-mobility/Pages/default.aspx>]
3. Principal investigator for a VESUVIO direct proposal and three Xpress proposals on “Deep inelastic neutron scattering experiments on solid and liquid phases of H₂O and H₂O¹⁸” (ISIS, UK, Sep. 2019 and Oct. 2018)
[<https://www.isis.stfc.ac.uk/Pages/vesuvio.aspx>]
4. Co-investigator for a 1.5 million CPU hours project “Towards an accurate, high-throughput framework for the prediction of anharmonic free energies in molecular crystals” (CSCS, Switzerland, Oct. 2017)
[<https://www.cscs.ch/>]
5. Best Poster Prize for “Sampling Free Energy Surfaces as Slices by Combining Umbrella Sampling and Metadynamics”, Shalini Awasthi, **Venkat Kapil**, Nisanth N. Nair, CPMD Meeting, University of Chicago, USA (May 2016)
[<https://pme.uchicago.edu/events/4179/>]
6. Academic Excellence Award, IITK, India (June 2014)
[<https://iitk.ac.in/sspc/sspc-academic-awards>]
7. Charpak Scholar of Excellence for a research internship at Chimie ParisTech, France (June 2013)
[<https://www.inde.campusfrance.org/charpak-lab-previously-charpak-research-internship-program>]

[†] preferred

TEACHING

1. Teaching assistant for *Statistical methods in atomistic computer simulations*
Conducted and evaluated tutorial sessions (June 2017)
[<https://edu.epfl.ch/coursebook/en/statistical-methods-in-atomistic-computer-simulations-MSE-639>]
2. Teaching assistant for *Statistical mechanics*
Designed assignments on general statistical mechanics and projects on phase transitions in Ising and Flory models, and conducted tutorial sessions (Feb. 2016 and Feb. 2017)
[<https://edu.epfl.ch/coursebook/en/statistical-mechanics-MSE-421>]

SUPERVISION OF STUDENTS

1. Pengkang Guo, semester project on “Development of a universal engine for the calculation of structural properties of materials” (Sep. 2019)
2. Notash Shafiei, semester project on “Parallelization of a python interface for atomistic modelling” (Feb. 2019)
3. Luca Montanelli, semester project on “Implementation of a variable cell optimization scheme in i-PI*” (Sep. 2018)
4. Yuhang Yuan, masters thesis on “Different Integration schemes in molecular dynamics: Accuracy benchmarks and applications to the calculations of viscosity and diffusion coefficient” (Sep. 2018)
5. Sebastien Bienvenue, semester project on “Implementation of a quasi-random Monte Carlo scheme to improve the self-consistent phonons scheme in i-PI” (Sep. 2017)
6. Alice Cuzzocrea, masters thesis on “Calculation of the proton momentum distribution in different phases of water” (Feb. 2017)
7. Pierre Leroux, semester project on “Implementation of a self-consistent phonons scheme for calculating the free energy of a crystal in i-PI” (Feb. 2016)
8. Sebastien Bienvenue, semester project on “Implementation of a finite difference scheme to calculate the phonons of a crystal in i-PI” (Oct. 2015)

SERVICE AND COMMUNITY RECOGNITION

1. Referee for the Journal of Chemical Physics and Chemical Physics Letters
2. Invited speaker at Atomistic Simulations Webinar Series, International Centre for Theoretical Physics, Trieste (May 2020), Chemistry Department and Theoretical Condensed Matter Physics Department, Cambridge University (Feb. 2020), Materials Department, Chalmers University, Sweden (Sep. 2019), the Workshop on Neutron Scattering, Poznan, Poland (May 2018), the CP2K User Tutorial, Zurich, Switzerland (July 2017) and the Physical Chemistry Seminar, IITK, India (Dec. 2016)
3. Conductor of hands-on sessions (4 hours each) at Chalmers University (Sep. 2019), Sweden, the CECAM Path Integral School, Lausanne, Switzerland (June 2018 and June 2016) and the CP2K User Tutorial, Zurich, Switzerland (July 2017) on simulating nuclear quantum effects using an open-source software, i-PI

TECHNICAL SKILLS

1. Proficient programmer in Python, C, C++ and Fortran – using (Py)MPI – and an avid user of mathematica and jupyter notebooks
2. Principal developer of i-PI v2.0, a python based software for advanced molecular calculations
[<http://ipi-code.org/>]

* A python code for performing atomistic simulations

3. Contributor to prominent (*ab initio*) molecular dynamics softwares CPMD and LAMMPS

[<https://lammmps.sandia.gov/>]

[<http://www.cpmd.org/>]

Research output list

POSTERS

1. “The anisotropy of proton momentum distribution in water”
Water Europe, Lausanne, Switzerland (Oct. 2019)
[\[https://www.cecaml.org/workshop1798/\]](https://www.cecaml.org/workshop1798/)
2. “An assessment of approximate methods for anharmonic free energy”
CECAM 50, Lausanne, Switzerland (Sep. 2019)
[\[https://cecaml50.cecaml.org/\]](https://cecaml50.cecaml.org/)
3. “An assessment of approximate methods for anharmonic free energy”
Marvel Review and Retreat, Lausanne, Switzerland (Sep. 2019)
[\[https://sites.google.com/view/Marvel-rr2019/home\]](https://sites.google.com/view/Marvel-rr2019/home)
4. “An assessment of approximate methods for anharmonic free energy”
CPMD Meeting, Lausanne, Switzerland (July 2019)
[\[https://www.cecaml.org/workshop1738/\]](https://www.cecaml.org/workshop1738/)
5. “Multiple time steps in real and imaginary time”
CAMD Summer School on Electronic Structure Theory and Materials Design, Helsingør, Denmark (Aug. 2018)
[\[https://www.conferencemanager.dk/camd2018/the-camd-summer-school.html\]](https://www.conferencemanager.dk/camd2018/the-camd-summer-school.html)
6. “Multiple time steps in real and imaginary time”
PASC18, Basel, Switzerland (July 2018)
[\[https://pasc18.pasc-conference.org/\]](https://pasc18.pasc-conference.org/)
7. “High order path integrals made easy”
Recent Advances in Modelling Rare Events, Agra, India (Dec. 2017)
[\[https://www.iitk.ac.in/chm/rare2017\]](https://www.iitk.ac.in/chm/rare2017)
8. “Probing the momentum distribution of protons in *ab initio* water”
Neutrons Matter, Rome, Italy (Nov. 2017)
[\[http://sons.uniroma2.it/news-events/workshop/\]](http://sons.uniroma2.it/news-events/workshop/)
9. “Multiple time steps in real and imaginary time”
Path Integral Quantum Mechanics: Theory, Simulation and Application, Lausanne, Switzerland (June 2016)
[\[https://cosmo.epfl.ch/research/page-123319-en-html/page-133709-en-html/\]](https://cosmo.epfl.ch/research/page-123319-en-html/page-133709-en-html/)
10. “Multiple time steps in real and imaginary time”
DPG Meeting, Regensburg, Germany (Mar. 2016)
[\[https://regensburg16.dpg-tagungen.de/\]](https://regensburg16.dpg-tagungen.de/)

SELECTED TALKS

1. “An assessment of approximate statistical sampling methods for anharmonic free energy”
Advanced Electronic Structure Methods in Condensed Matter Physics, Lausanne, Switzerland (July 2019)
[\[https://sites.google.com/view/eth-electronic-structure-2019\]](https://sites.google.com/view/eth-electronic-structure-2019)
2. “Advanced atomistic simulations using i-PI”
Application of Machine-Learning Interatomic Potentials in Materials Design, Moscow, Russia (June 2019)
[\[https://amlipmd.skoltech.ru/\]](https://amlipmd.skoltech.ru/)
3. “Multiple time steps in real and imaginary time: Using DFT to accelerate wave function methods”
Computational Material Science, Warwick, UK (Mar. 2019)
[\[https://warwick.ac.uk/fac/sci/chemistry/news/events/cms2019/\]](https://warwick.ac.uk/fac/sci/chemistry/news/events/cms2019/)

4. “Thermal analysis of guest loaded metal organic frameworks”
Mini Symposium, Workshop of Theoretical Chemistry, Mariapfarr, Austria (Feb. 2019)
[<https://www.tugraz.at/institute/ptc/conferences/wtc-Mariapfarr/previous-workshops/wtc-2019/>]
5. “Advanced path integrals: beyond the benchmarks”
COMDI, Lausanne, Switzerland (Sep. 2018)
[<https://sites.google.com/view/comdi2018>]
6. “Advanced path integrals: beyond the benchmarks”
CECAM Path Integral School, Lausanne, Switzerland (June 2018)
[<https://cosmo.epfl.ch/research/page-123319-en-html/page-133709-en-html/>]
7. “How to simulate nuclear quantum effects accurately at the cost of classical molecular dynamics”
DPG Meeting, Berlin, Germany (Mar. 2018)
[<https://berlin18.dpg-tagungen.de/>]
8. “How to simulate nuclear quantum effects accurately at the cost of classical molecular dynamics”
Marvel Junior Seminars, Lausanne, Switzerland (Jan. 2018)
[<http://nccr-Marvel.ch/events/Marvel-junior-seminar-january2018>]
9. “High order path integrals made easy”
Workshop on Path Integrals, Trieste, Italy (July 2017)
[<http://indico.ictp.it/event/7975/>]
10. “High order path integrals made easy”
DPG Meeting, Dresden, Germany (Mar. 2017)
[<https://dresden17.dpg-tagungen.de/>]

INVITED TALKS AND LECTURES

1. “Nuclear Quantum Effects: Fast and Accurate”
Atomistic Simulation Webinar Series, International Centre for Theoretical Physics, (May. 2020)
[<http://indico.ictp.it/event/9366/overview>]
2. “Nuclear Quantum Effects: Fast and Accurate”
Chemistry Department, Cambridge University, (Feb. 2020)
[<http://talks.cam.ac.uk/talk/index/138652>]
3. “Assessment of approximate methods for anharmonic free energies”
Theoretical Condensed Matter, Cambridge University, (Feb. 2020)
[<https://talks.cam.ac.uk/talk/index/138943>]
4. “Advanced methods for modelling the quantum nature of nuclei: Going beyond benchmarks”
Chalmers University, Sweden (Sep. 2019)
5. “Efficient calculation of particle momentum distribution and computational assessment of deep inelastic neutron scattering”
Neutron Scattering Investigation in Condensed Matter, Poznan, Poland (Mar. 2017)
[<https://tinyurl.com/y3qkqkjj>]
6. “Efficient molecular dynamics: Thermostats, barostats and multiple Time Steps”
CP2K User Tutorial, Zurich, Switzerland (July 2017)
[https://www.cp2k.org/events:2017_user_tutorial:about]
7. “Multiple time steps and path integral molecular dynamics”
CP2K User Tutorial, Zurich, Switzerland (July 2017)
[https://www.cp2k.org/events:2017_user_tutorial:about]
8. “Accelerating the convergence of nuclear quantum effects”
Indian Institute of Technology Kanpur, Kanpur, India (Dec. 2016)

1. **Venkat Kapil**, David M. Wilkins, Jinggang Lan, and Michele Ceriotti. Inexpensive modeling of quantum dynamics using path integral generalized langevin equation thermostats. *The Journal of Chemical Physics*, 152(12):124104, March 2020
[<https://aip.scitation.org/doi/full/10.1063/1.5141950>]
2. Igor Poltavsky, **Venkat Kapil**, Michele Ceriotti, Kwang S. Kim, and Alexandre Tkatchenko. Accurate description of nuclear quantum effects with high-order perturbed path integrals (HOPPI). *Journal of Chemical Theory and Computation*, 16(2):1128–1135, January 2020
[<https://pubs.acs.org/doi/full/10.1021/acs.jctc.9b00881>]
3. Ming Liu, Linda Zhang, Marc A. Little, **Venkat Kapil**[†], Michele Ceriotti, Siyuan Yang, Lifeng Ding, Daniel L. Holden, Rafael Balderas-Xicohtencatl, Donglin He, Rob Clowes, Samantha Y. Chong, Gisela Schütz, Linjiang Chen, Michael Hirscher, and Andrew I. Cooper. Barely porous organic cages for hydrogen isotope separation. *Science*, 366(6465):613–620, October 2019
[<https://science.sciencemag.org/content/366/6465/613.full>]
4. Jelle Wieme, Steven Vandenbrande, Aran Lemaire, **Venkat Kapil**, Louis Vanduyfhuys, and Veronique Van Speybroeck. Thermal engineering of metal-organic frameworks for adsorption applications: A molecular simulations perspective. *ACS Applied Materials & Interfaces*, 2019
[<https://pubs.acs.org/doi/10.1021/acsami.9b12533>]
5. **Venkat Kapil**, Edgar Engel, Mariana Rossi, and Michele Ceriotti. Assessment of approximate methods for anharmonic free energies. *Journal of Chemical Theory and Computation*, 15(11):5845–5857, September 2019
[<https://pubs.acs.org/doi/10.1021/acs.jctc.9b00596>, <https://arxiv.org/abs/1906.06524>]
6. **Venkat Kapil**, Jelle Wieme, Steven Vandenbrande, Aran Lemaire, Veronique Van Speybroeck, and Michele Ceriotti. Modeling the structural and thermal properties of loaded metal-organic frameworks. An interplay of quantum and anharmonic fluctuations. *Journal of Chemical Theory and Computation*, 15(5):3237–3249, April 2019
[<https://pubs.acs.org/doi/10.1021/acs.jctc.8b01297>, <https://arxiv.org/abs/1901.03770>]
7. **Venkat Kapil**, Mariana Rossi, Ondrej Marsalek, Riccardo Petraglia, Yair Litman, Thomas Spura, Bingqing Cheng, Alice Cuzzocrea, Robert H. Meißner, David M. Wilkins, Benjamin A. Helfrecht, Przemysław Juda, Sébastien P. Bienvenue, Wei Fang, Jan Kessler, Igor Poltavsky, Steven Vandenbrande, Jelle Wieme, Clemence Corminboeuf, Thomas D. Kühne, David E. Manolopoulos, Thomas E. Markland, Jeremy O. Richardson, Alexandre Tkatchenko, Gareth A. Tribello, Veronique Van Speybroeck, and Michele Ceriotti. i-PI 2.0: A universal force engine for advanced molecular simulations. *Computer Physics Communications*, 236:214–223, March 2019
[<https://www.sciencedirect.com/science/article/pii/S0010465518303436?via%3Dihub>, <https://arxiv.org/abs/1808.03824>]
8. **Venkat Kapil**, Alice Cuzzocrea, and Michele Ceriotti. Anisotropy of the proton momentum distribution in water. *The Journal of Physical Chemistry B*, 122(22):6048–6054, May 2018
[<https://pubs.acs.org/doi/abs/10.1021/acs.jpcc.8b03896>, <https://arxiv.org/abs/1805.01193>]
9. Mariana Rossi, **Venkat Kapil**, and Michele Ceriotti. Fine tuning classical and quantum molecular dynamics using a generalized langevin equation. *The Journal of Chemical Physics*, 148(10):102301, March 2018
[<https://aip.scitation.org/doi/abs/10.1063/1.4990536>, <https://arxiv.org/abs/1704.05099>]
10. **Venkat Kapil**, Jörg Behler, and Michele Ceriotti. High order path integrals made easy. *The Journal of Chemical Physics*, 145(23):234103, December 2016
[<https://aip.scitation.org/doi/abs/10.1063/1.4971438>, <https://arxiv.org/abs/1606.00920>]
11. **Venkat Kapil**, Joost VandeVondele, and Michele Ceriotti. 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. *The Journal of Chemical Physics*, 144(5):054111, February 2016
[<https://aip.scitation.org/doi/abs/10.1063/1.4941091>, <https://arxiv.org/abs/1512.00176>]

[†] first computational author

12. Shalini Awasthi, **Venkat Kapil**, and Nisanth N. Nair. Sampling free energy surfaces as slices by combining umbrella sampling and metadynamics. *Journal of Computational Chemistry*, 37(16):1413–1424, April 2016
[<https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.24349>, <https://arxiv.org/abs/1507.06764>]

For more details check out <https://scholar.google.ch/citations?user=F2NMPNQAAAAJ&hl=en>

PUBLICATIONS IN PEER-REVIEWED CONFERENCE PROCEEDINGS

1. Kacper Drużbicki, Matthew Krzystyniak, Daniel Hollas, **Venkat Kapil**, Petr Slavíček, G Romanelli, and Felix Fernandez-Alonso. Hydrogen dynamics in solid formic acid: Insights from simulations with quantum colored-noise thermostats. *Journal of Physics: Conference Series*, 1055:012003, July 2018
[<https://iopscience.iop.org/article/10.1088/1742-6596/1055/1/012003/meta>]

SOFTWARES AND TUTORIALS

1. Lead developer of i-PI, a python wrapper code for performing advanced atomistic simulations using force evaluations from prominent electronic structure softwares
[<http://ipi-code.org/>]
2. Contributor to Quantum Mobile, an open-source virtual machine containing state-of-the-art softwares for running simulations governed by laws of quantum mechanics
[<https://www.materialscloud.org/work/quantum-mobile>]
3. Developer of a tutorial on i-PI which has been used in several schools and workshops
[<https://github.com/venkatkapil24/pimd-tutorial>]