Algorithms for quantum molecular dynamics: from matrix product states to path integrals

by

Dmitri Iouchtchenko

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2021

Examining committee membership

The following served on the examining committee for this thesis. The decision of the examining committee is by majority vote.

External Examiner	Artur Izmaylov Associate Professor University of Toronto, Toronto, Ontario, Canada
Supervisor	Pierre-Nicholas Roy Professor University of Waterloo, Waterloo, Ontario, Canada
Internal Member	Marcel Nooijen Professor University of Waterloo, Waterloo, Ontario, Canada
Internal-external Member	Jeff Chen Professor University of Waterloo, Waterloo, Ontario, Canada
Other Member	Scott Hopkins Associate Professor University of Waterloo, Waterloo, Ontario, Canada

Author's declaration

This thesis consists of material all of which I authored or co-authored: see the statement of contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of contributions

With the exception of Chapters 1 and 6, this thesis is composed of manuscripts (published or still in preparation) to which I have contributed substantially, but not exclusively. Contributions from my co-authors are as follows:

- Chapter 2: P.-N.R. suggested changes to the manuscript text. Based on Dmitri Iouchtchenko and Pierre-Nicholas Roy, "Ground states of linear rotor chains via the density matrix renormalization group", The Journal of Chemical Physics **148**, 134115 (2018).[1]

- Chapter 3: N.R., P.-N.R., and M.N. made recommendations to improve the manuscript. Based on Dmitri Iouchtchenko, Neil Raymond, Pierre-Nicholas Roy, and Marcel Nooijen, "Deterministic and quasi-random sampling of optimized Gaussian mixture distributions for vibronic Monte Carlo", arXiv:1912.11594v1.[2]

- Chapter 4: K.P.B. wrote parts of the Introduction and Conclusion; P.-N.R. proposed additions to the manuscript text. Based on Dmitri Iouchtchenko, Kevin P. Bishop, and Pierre-Nicholas Roy, "On the quantum mechanical potential of mean force. I. A path integral perspective", arXiv:2101.00761v1.[3]

- Chapter 5: K.P.B. and P.-N.R. engaged in fruitful discussions. Based on Dmitri Iouchtchenko, Kevin P. Bishop, and Pierre-Nicholas Roy, "On the quantum mechanical potential of mean force. II. Constrained path integral molecular dynamics integrators", arXiv:2101.00762v1.[4]

In addition to the above, I have collaborated on several manuscripts during the course of my PhD research which I have chosen to exclude from this thesis, as I was not the primary author:

- Roger G. Melko, Chris M. Herdman, Dmitri Iouchtchenko, Pierre-Nicholas Roy, and Adrian Del Maestro, "Entangling qubit registers via many-body states of ultracold atoms", Physical Review A **93**, 042336 (2016).[5]

- Tom Halverson, Dmitri Iouchtchenko, and Pierre-Nicholas Roy, "Quantifying entanglement of rotor chains using basis truncation: Application to dipolar endofullerene peapods", The Journal of Chemical Physics **148**, 074112 (2018).[6]

- Neil Raymond, Dmitri Iouchtchenko, Pierre-Nicholas Roy, and Marcel Nooijen, "A path integral methodology for obtaining thermodynamic properties of nonadiabatic systems using Gaussian mixture distributions", The Journal of Chemical Physics **148**, 194110 (2018).[7]

- Tapas Sahoo, Dmitri Iouchtchenko, Chris M. Herdman, and Pierre-Nicholas Roy, "A path integral ground state replica trick approach for the computation of entanglement entropy of dipolar linear rotors", The Journal of Chemical Physics **152**, 184113 (2020).[8]

- Isaac J. S. De Vlugt, Dmitri Iouchtchenko, Ejaaz Merali, Pierre-Nicholas Roy, and Roger G. Melko, "Reconstructing quantum molecular rotor ground states", Physical Review B **102**, 035108 (2020).[9]

- Samrit Mainali, Dmitri Iouchtchenko, Pierre-Nicholas Roy, Hans-Dieter Meyer, and Fabien Gatti, "Comparison of the Multi-Layer Multi-Configuration Time-Dependent Hartree (ML-MCTDH) and the Density Matrix Renormalization Group (DMRG) methods when computing Ground State Properties of Linear Rotor Chains", manuscript in preparation, to be submitted in 2021.

Abstract

This thesis describes several novel approaches in quantum molecular dynamics for obtaining properties of molecular systems in different regimes. We investigate ground state properties of chains of linear rotors with dipole–dipole interactions via the density matrix renormalization group (DMRG), by deriving the appropriate form of the interaction operator and implementing it in ITensor. This provides us with further evidence of a quantum phase transition in this system. We also improve the sampling of Gaussian mixture distributions for finite temperature path integral Monte Carlo (PIMC) of vibronic Hamiltonians. To do this, we replace random sampling by quasi-random sampling, and improve sampling distributions by optimizing their parameters. Finally, we introduce estimators and integrators for constrained free energy simulations in path integral molecular dynamics (PIMD). This method is applied to the study of a water dimer, for which we obtain a quantum potential of mean force.

Acknowledgements

All of the work contained in this thesis would not be possible without the generous support of my supervisor, Pierre-Nicholas Roy. The members of my advisory committee, Marcel Nooijen, Roger Melko, and Scott Hopkins have kindly provided me with probing questions and useful advice over the years, for which I am grateful.

During my time in the Theoretical Chemistry group at the University of Waterloo, I've had the pleasure of being surrounded by thoughtful, creative, and insightful people. I would like to thank the following group members, past and present, and in no particular order, for making me never want to graduate: Kevin Bishop, Joshua Cantin, Isaac De Vlugt, Nabil Faruk, Tanya Gatsak, Prateek Goel, Tom Halverson, Chris Herdman, Lee Huntington, Alexander Ibrahim, Yulia Kalugina, Mike Lecours, Xichen Lou, Adam Marr, Lindsay Orr, Neil Raymond, Tapas Sahoo, Matthew Schmidt, Jianying Sheng, Austin Tripp, Lecheng Wang, Siyuan Wu, Spencer Yim, Emily Yip, Mark Zanon, Toby Zeng, and Xiaolong Zhang. If I've collaborated with you, I look forward to it again; otherwise, we should try it some time!

University staff outside of the research group have also helped to make my graduate experience a frictionless one. Many thanks to: Jake Fisher, Rick Loney, Stephen Markan, Terry McKay, Paul Miskovsky, Krista Parsons, Cathy Van Esch, and Kate Wood.

I don't understand how my family puts up with me, but I'm very glad that they do.

Памяти Алексея Григорьевича Смирнова.

(1941–2020)

In memory of Alexey Grigorievich Smirnov.

Table of contents

Chapter 1

Introduction

In 1929, Dirac famously claimed[10] that

> The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.

Nearly a century later, these equations are still too complex to be solved exactly, but the fields of theoretical physics and chemistry are thriving, and full of creative approximations.

Much of theoretical chemical physics is concerned with the solution of the time-independent Schrödinger equation (TISE)

$$\hat{H}\,|\psi_n\rangle = E_n\,|\psi_n\rangle \tag{1.1}$$

for the molecular Hamiltonian $\hat{H}$, which contains the kinetic and potential energy terms of the nuclei ($\hat{K}_\mathrm{n}$, $\hat{V}_\mathrm{n}$) and electrons ($\hat{K}_\mathrm{e}$, $\hat{V}_\mathrm{e}$) in a chemical system:

$$\hat{H} = \hat{K}_\mathrm{n} + \hat{K}_\mathrm{e} + \hat{V}_\mathrm{n} + \hat{V}_\mathrm{e} + \hat{V}_\mathrm{ne}. \tag{1.2}$$

The coupling term $\hat{V}_\mathrm{ne}$ is responsible for the interactions between nuclei and electrons.

In the position representation, the TISE becomes a differential equation, as the Hamiltonian takes on the form

$$\langle \mathbf{R}\,\mathbf{r}|\,\hat{H} = \left[-\sum_i \frac{\hbar^2}{2M_i}\frac{\partial^2}{\partial R_i^2} - \sum_j \frac{\hbar^2}{2m_\mathrm{e}}\frac{\partial^2}{\partial r_j^2} + V(\mathbf{R},\mathbf{r})\right]\langle \mathbf{R}\,\mathbf{r}|\,, \tag{1.3}$$

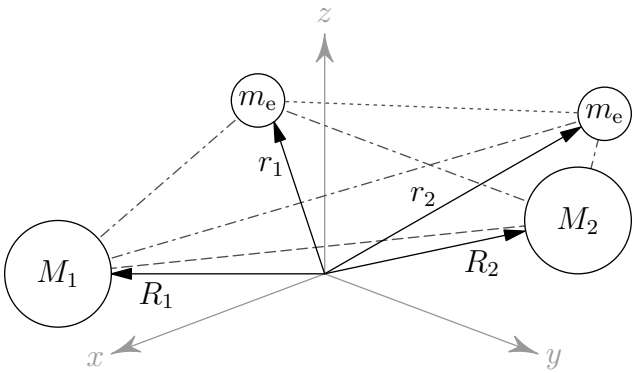

Figure 1.1: Schematic of the nuclei (R_1, R_2) and electrons (r_1, r_2) in the hydrogen molecule placed at arbitrary positions. The dashed line indicates nucleus–nucleus repulsion due to $\hat{V}_\mathrm{n}$; the dotted line indicates electron–electron repulsion due to $\hat{V}_\mathrm{e}$; the dash-dotted lines indicate nucleus–electron attraction due to $\hat{V}_\mathrm{ne}$. The figure is not to scale.

where R_i and M_i are the coordinates and masses of the nuclei, r_j are the coordinates of the electrons (which all share the same mass m_e), and $V(\mathbf{R}, \mathbf{r})$ is the full Coulombic potential arising from $\hat{V}_\mathrm{n}$, $\hat{V}_\mathrm{e}$, and $\hat{V}_\mathrm{ne}$. For example, the potential terms involved in the H_2 molecule are depicted in Fig. 1.1. The eigenstates

$$\psi_n(\mathbf{R}, \mathbf{r}) = \langle \mathbf{R}\, \mathbf{r} | \psi_n \rangle \tag{1.4}$$

along with the corresponding eigenenergies E_n make up the solutions to the TISE, but they are not simple to obtain directly.

Since the lightest nucleus (composed of just a single proton) is over 1800 times heavier than an electron,[11] a simplification is possible even for the fairly general form in Eq. (1.3). By combining all the terms except $\hat{K}_\mathrm{n}$, one arrives at the electronic Hamiltonian $\hat{H}_\mathrm{e}(\mathbf{R})$, which is parameterized by the nuclear geometry. Its eigenstates $\chi_k(\mathbf{r}; \mathbf{R})$ and eigenenergies $E_k(\mathbf{R})$ therefore also depend parametrically on the positions of the nuclei, and the total wavefunctions may be written as

$$\psi_n(\mathbf{R}, \mathbf{r}) = \sum_k \varphi_k^{(n)}(\mathbf{R}) \chi_k(\mathbf{r}; \mathbf{R}), \tag{1.5}$$

with nuclear coefficients $\varphi_k^{(n)}(\mathbf{R})$. Relying on the vast difference in masses (and hence time scales) between the nuclei and electrons, one arrives at the Born–Oppenheimer approximation (BOA).[12] In this approximation, the electrons are treated as if they are able to instantaneously adapt to the motion of the nuclei, while the nuclei only experience interactions with an effective electronic force field. As a result of this adiabatic separation, it is valid in many cases to take only solutions of

$\hat{H}_\mathrm{e}(\mathbf{R})$ with the lowest energy ($k = 0$) and keep just the first term of the wavefunction expansion:

$$\psi_n(\mathbf{R}, \mathbf{r}) \approx \varphi_0^{(n)}(\mathbf{R})\chi_0(\mathbf{r}; \mathbf{R}). \tag{1.6}$$

Although finding solutions of the electronic Hamiltonian is a noble endeavor, the focus of this thesis is the complementary problem: quantum molecular dynamics, which typically studies the motion of nuclei subject to an effective potential energy surface (PES) as dictated by the electronic structure. The Hamiltonian is reduced to just the nuclear motions, described by the kinetic and potential energy terms:

$$\hat{H}_\mathrm{n} = \hat{K}_\mathrm{n} + V_\mathrm{PES}(\hat{\mathbf{R}}). \tag{1.7}$$

In *ab initio* methods, the potential function $V_\mathrm{PES}(\mathbf{R})$ is taken to be $E_0(\mathbf{R})$ from the electronic problem, but it may also be approximated by a phenomenological model or an empirical force field. It should be noted that even though the word "dynamics" evokes notions of time evolution, quantum molecular dynamics encompasses more than just real-time phenomena; it does not necessarily imply quantum dynamics, which concerns itself with things like correlation functions and wavepacket propagation.

While the ideal solution to the TISE contains all the eigenstates, as obtained through diagonalization of the nuclear Hamiltonian matrix, this is usually neither possible nor required. Real systems are commonly found to be at some specific temperature, and it is often just thermodynamic properties that are of interest. These properties may be found using statistical mechanics, which for closed systems at constant temperature has as its central object the canonical partition function

$$Z = \operatorname{Tr} e^{-\beta \hat{H}_\mathrm{n}}, \tag{1.8}$$

where $\beta = 1/k_\mathrm{B}T$ is the reciprocal temperature.[13] Various derivatives of the partition function yield quantities such as internal energy U, pressure p, and entropy S; for example:

$$U = -\frac{\partial \log Z}{\partial \beta} = \frac{1}{Z} \operatorname{Tr} e^{-\beta \hat{H}_\mathrm{n}} \hat{H}_\mathrm{n} = \left\langle \hat{H}_\mathrm{n} \right\rangle. \tag{1.9}$$

Thus, we may apply the rules of statistical mechanics to combine the eigenstates $|\varphi_i\rangle$ of $\hat{H}_\mathrm{n}$ (with eigenvalues E_i) into thermal averages of quantum operators:

$$\langle \hat{O} \rangle = \frac{1}{Z} \operatorname{Tr} e^{-\beta \hat{H}_\mathrm{n}} \hat{O} = \frac{1}{Z} \sum_i e^{-\beta E_i} O_{ii}, \tag{1.10}$$

where $\hat{O}$ is an arbitrary operator with matrix elements $O_{ij} = \langle \varphi_i | \hat{O} | \varphi_j \rangle$.

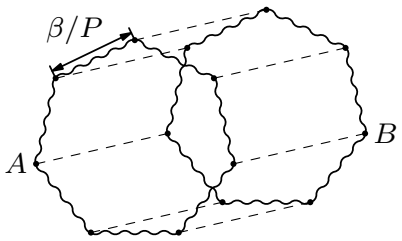

Figure 1.2: Two particles A and B expressed using path integrals with $P = 7$ beads. The dashed lines correspond to inter-particle interactions, while the wavy lines represent the coupling between the replicas of each particle.

Some finite temperature schemes, such as imaginary time path integrals, can provide these averages without explicitly computing any eigenstates.[14] The resolution of the identity in the position representation,

$$\hat{\mathbb{1}} = \int \mathrm{d}\mathbf{R}\, |\mathbf{R}\rangle\langle\mathbf{R}|\,, \tag{1.11}$$

may be inserted at evenly spaced intervals into the integral form of the partition function,

$$Z = \int \mathrm{d}\mathbf{R}\ \langle\mathbf{R}|e^{-\beta\hat{H}_\mathrm{n}}|\mathbf{R}\rangle\,, \tag{1.12}$$

to construct a discretized imaginary time path integral. This results in a sequence of coupled identical replicas of the system:

$$\begin{aligned} Z = \int\int\cdots\int \mathrm{d}\mathbf{R}_1\,\mathrm{d}\mathbf{R}_2\cdots\mathrm{d}\mathbf{R}_P\ \langle\mathbf{R}_1|e^{-\frac{\beta}{P}\hat{H}_\mathrm{n}}|\mathbf{R}_2\rangle\ \langle\mathbf{R}_2|e^{-\frac{\beta}{P}\hat{H}_\mathrm{n}}|\mathbf{R}_3\rangle\cdots \\ \times\ \langle\mathbf{R}_{P-1}|e^{-\frac{\beta}{P}\hat{H}_\mathrm{n}}|\mathbf{R}_P\rangle\ \langle\mathbf{R}_P|e^{-\frac{\beta}{P}\hat{H}_\mathrm{n}}|\mathbf{R}_1\rangle\,. \end{aligned} \tag{1.13}$$

In general, the high-temperature propagator $\langle\mathbf{R}'|e^{-\frac{\beta}{P}\hat{H}_\mathrm{n}}|\mathbf{R}\rangle$ is impossible to express exactly, but it may often be approximated. A common approximation stems from the Trotter decomposition

$$e^{-\beta\hat{H}_\mathrm{n}} = \lim_{P\to\infty}\left(e^{-\frac{\beta}{P}\hat{K}_\mathrm{n}}e^{-\frac{\beta}{P}\hat{V}_\mathrm{PES}}\right)^P, \tag{1.14}$$

in which case increasing the number of replicas in the path improves the quality of the approximation. A discretized path integral of this kind may be viewed as a necklace of P beads, such as in Fig. 1.2.

Once the high-temperature propagators are appropriately decomposed, the entire integrand in Eq. (1.13) may be written as a function, $\pi(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_P)$, which we condense to just $\pi(\mathbf{q})$ for brevity by consolidating all the coordinate vectors $\mathbf{R}_j$ into $\mathbf{q}$. The partition function may then be

approximated as

$$Z \approx \int \mathrm{d}\mathbf{q}\ \pi(\mathbf{q}), \tag{1.15}$$

which has the advantage of being entirely free of operators. Likewise, for some observable $\hat{O}$, the quantum expectation value

$$\langle \hat{O} \rangle = \frac{\mathrm{Tr}\, e^{-\beta \hat{H}_\mathrm{n}} \hat{O}}{\mathrm{Tr}\, e^{-\beta \hat{H}_\mathrm{n}}} \tag{1.16}$$

may be written as

$$\langle \hat{O} \rangle \approx \frac{\int \mathrm{d}\mathbf{q}\ \pi(\mathbf{q})\, \mathcal{E}_{\hat{O}}(\mathbf{q})}{\int \mathrm{d}\mathbf{q}\ \pi(\mathbf{q})}, \tag{1.17}$$

where $\mathcal{E}_{\hat{O}}(\mathbf{q})$ is an estimator function for $\hat{O}$. Given the ability to sample the configurations $\{\mathbf{q}_{[i]}\}_{i=1}^{N_\mathrm{MC}}$ from $\pi(\mathbf{q})$, one may apply Monte Carlo techniques to the evaluation of expectation values, estimating the above path integral expression as

$$\langle \hat{O} \rangle \approx \frac{1}{N_\mathrm{MC}} \sum_{i=1}^{N_\mathrm{MC}} \mathcal{E}_{\hat{O}}(\mathbf{q}_{[i]}). \tag{1.18}$$

This approach to computing finite temperature quantities is referred to as path integral Monte Carlo (PIMC).

While PIMC is incredibly versatile, its flexibility comes at a cost: efficiently sampling from $\pi(\mathbf{q})$ is not always straightforward. However, in some circumstances, it is possible to dynamically evolve the full replicated system using classical mechanics. This is achieved by augmenting the position coordinates of the system by fictitious momenta.[15] The result is known as path integral molecular dynamics (PIMD), and the equations of motion are typically evaluated using a specialized integrator, such as the path integral Langevin equation (PILE).[16] In this thesis, both PIMC and PIMD are used to calculate finite-temperature properties of small systems.

In 2018, a PIMC method was introduced that allows the partition function of a nonadiabatic system to be computed from a vibronic Hamiltonian.[7] This sort of Hamiltonian contains couplings between the vibrational and electronic degrees of freedom, which are neglected in the BOA, making this an example of a method that does not reduce the behavior of the electrons to a single function $V_\mathrm{PES}(\mathbf{R})$. This extension beyond the BOA can be important for nuclear configurations $\mathbf{R}$ that have some eigenvalues $E_k(\mathbf{R})$ close to each other in energy, and which may therefore have nonadiabatic coupling terms that are not negligible.

Unlike prior nonadiabatic path integral methods in which the coordinates hop between surfaces,[17,18] in the novel PIMC method, all the coordinates are evaluated on all surfaces simultaneously by virtue of the path being composed of matrices. However, this method relies on the Gaussian mixture distribution (GMD) that is used for sampling having good coverage of the vibrational regions of interest, which can be difficult to achieve in practice. In this thesis, we propose a way to improve the sampling GMD to ensure that it is as similar as possible to the integrand in the partition function integral. To that end, we construct a loss function and minimize it using a standard optimization algorithm.[19,20] Additionally, we provide some general sampling improvements, including the replacement of Monte Carlo by quasi-Monte Carlo, which uses low discrepancy sequences in place of pseudo-random numbers.[21]

Even within the confines of the BOA, the Helmholtz free energy

$$A = -\frac{1}{\beta} \log Z \tag{1.19}$$

is important for connecting theoretical simulations with experimental results, such as relative isomer populations and equilibrium constants. However, it is challenging to compute A with Monte Carlo and molecular dynamics, as it has no corresponding observable.[22] When constructing a free energy profile $A(\xi)$ (also known as a potential of mean force, or PMF) along the reaction coordinate ξ, the chief issue is that values of the reaction coordinate corresponding to energies much greater than the PES minima are extremely rare in a regular simulation. This makes it impossible to properly examine the regions of interest when studying properties such as transition rates, which inherently involve exploration of high-energy states.

To circumvent this, augmented dynamics must be used, such as umbrella sampling with subsequent histogram unbiasing.[23] It has been previously demonstrated that umbrella sampling may be implemented alongside PIMD by adding post-quantization restraints in order to study the quantum PMF of molecular clusters.[24] In our work, we propose an alternative approach that uses constraints rather than restraints, and we introduce specialized estimators and integrators for this purpose. We show that these estimators and integrators are simple to implement, and that they give correct results for the water dimer system at low and high temperatures.

For very cold systems, finite temperature methods become inefficient. For example, in the case of PIMC as described above, the required number of replicas becomes intolerable. Since only the lowest energy eigenstate, the ground state, contributes appreciably when the temperature is sufficiently low, it is more appropriate to use a method that is designed to obtain only the ground state wavefunction. Because wavefunctions in the naive tensor representation ($C_{\mathbf{n}} = \langle \mathbf{n} | \psi \rangle$; see Fig. 1.3) have a tendency to grow exponentially with system size, it is beneficial to also find a suitable ansatz to encode the wavefunction.

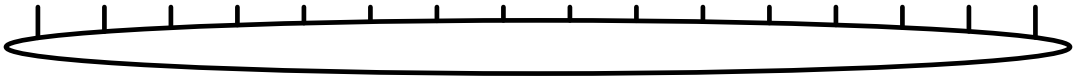

Figure 1.3: Pictorial representation of a rank-16 tensor of wavefunction coefficients $C_{\mathbf{n}} = C_{n_1 n_2 \cdots n_{16}}$. The vertical line segments correspond to the site indices.

Figure 1.4: Pictorial representation of a matrix product state decomposition of a rank-16 tensor into rank-2 and rank-3 tensors. The vertical line segments correspond to the site indices. The contraction performed by the horizontal line segments may be interpreted as matrix multiplication.

One possibility is the matrix product state (MPS)

$$\langle \mathbf{n} | \psi \rangle \approx \underline{\mathbf{A}}^{(1),n_1} \underline{\mathbf{A}}^{(2),n_2} \cdots \underline{\mathbf{A}}^{(N),n_N}, \tag{1.20}$$

which is written here as a product of matrices over N sites, and a 16-site example is depicted in Fig. 1.4. The power of the MPS originates from the ability to systematically reduce the common size between adjacent matrices $\underline{\mathbf{A}}^{(k),n_k}$ and $\underline{\mathbf{A}}^{(k+1),n_{k+1}}$, called the "bond dimension".[25] Reducing a bond dimension shrinks the state, but at the cost of cutting down the amount of entanglement that can be expressed across that bond and worsening the quality of the approximation. Many ground states of physical Hamiltonians are observed to obey an entanglement area law: the entanglement entropy of the state is proportional to the surface area of the boundary between subsystems, rather than their volume.[26] In one spatial dimension, the boundary must be of a fixed size, so the entanglement across every bond is constant, irrespective of the number of sites; the MPS of such a state will grow linearly with system size. Thus, states of long one-dimensional systems can often be represented compactly using an MPS.

The density matrix renormalization group (DMRG) takes advantage of the structure of an MPS by sweeping back and forth across the sites and diagonalizing an effective Hamiltonian for only two sites at a time.[25,27] For some problems, DMRG is so much more efficient than exact diagonalization of the entire Hamiltonian, that system sizes larger by several orders of magnitude become possible. This is why have chosen to use DMRG in order to find the ground state properties of a chain of linear rotors with dipole–dipole interactions. Although DMRG has seen use in electronic structure[28] and for the study of vibrations in molecules,[29] to our knowledge, our work shows its first application to the many-body rotational problem.

The model system in question is composed of quantum rotors arranged on a line, with a fixed, regular spacing between them. This may be implemented in a laboratory in the form of an endofullerene peapod: a nanomolecular assembly made up of polar linear molecules encased in fullerenes, which are then wrapped up into a nanotube (see Fig. 1.5). From a theoretical standpoint, the various

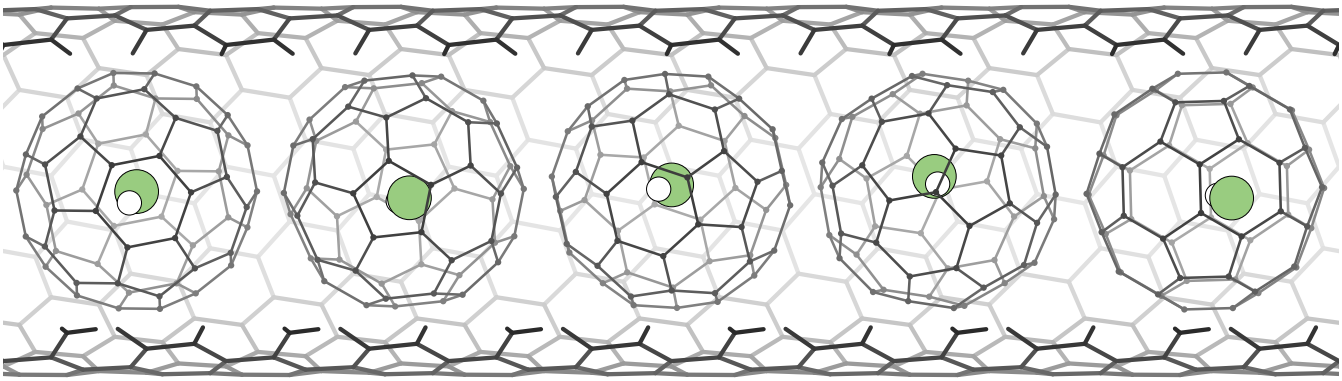

Figure 1.5: Artist's rendering of an (HF@C_{60})@SWNT nanomolecular assembly. The carbon nanotube and C_{60} fullerenes are in shades of grey, F atoms are green, and H atoms are white. A section of the nanotube has been omitted to reveal the interior.

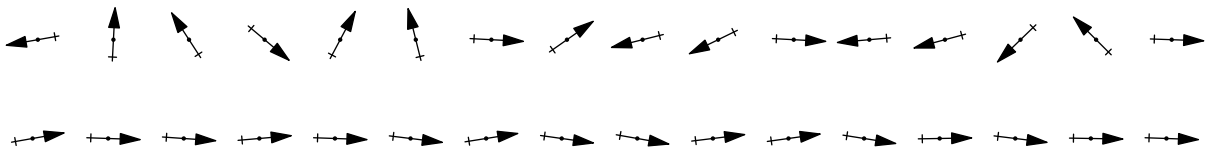

Figure 1.6: Example configurations of a chain of 16 dipolar rotors. Disordered rotors (top, $R \gg 1$) point in unrelated directions. Ordered rotors (bottom, $R \ll 1$) are aligned in a single direction, breaking left–right symmetry.

experimental parameters (such as the dipole moment and rotational constant) may be combined into a single dimensionless quantity R, allowing all possible realizations to be accessed by changing a Hamiltonian parameter.

This system of dipolar rotors has been hypothesized to undergo a quantum phase transition as the parameter R is decreased and the rotors orient themselves to point in one direction (as in Fig. 1.6),[30] and we provide further evidence to support this claim. In particular, we are able to compute the von Neumann entanglement entropy for a symmetric splitting of the system. We find that it converges fairly quickly with system size for both disordered and ordered chains, but near $R = 1$, there is instead a pronounced peak. This peak arises due to a divergence in the entanglement entropy, which is consistent with a violation of the area law that is expected at the critical point of a phase transition.

When the rotor DMRG calculations were first performed, the model that we had in mind of endofullerene peapod nanomolecular assemblies containing linear rotors inside them was only a dream. Since then, (HF@C_{60})@SWNT has been successfully synthesized in a laboratory setting.[31] Although this recent experiment was only performed at a single value of $R = 2.86$ (calculated from a lattice spacing of 1.05 nm, dipole moment of 0.45 D, and rotational constant of 20.561 cm^{-1})[31–33], it has paved the way for a physical realization of this quantum phase transition.

In all, the following chapters approach quantum molecular dynamics from many different perspectives,

and make an attempt to advance various facets of the field. Nevertheless, the diverse collection of methods mentioned above make up only a minuscule portion of the wide assortment of approaches to theoretical chemical physics, and one wonders whether Dirac envisioned such a breadth of approximations and algorithms in physics and chemistry.

The remainder of this thesis is structured as follows: in Chapter 2, we find ground states of the many-body linear rotor system using DMRG; in Chapter 3, we improve upon a PIMC method for vibronic Hamiltonians; in Chapters 4 and 5, we derive path integral estimators and molecular dynamics integrators for free energy calculations; and in Chapter 6, we reflect on what we've done.

Chapter 2

Ground states of linear rotor chains via the density matrix renormalization group

This chapter is reproduced from Dmitri Iouchtchenko and Pierre-Nicholas Roy, "Ground states of linear rotor chains via the density matrix renormalization group", The Journal of Chemical Physics **148**, 134115 (2018), with the permission of AIP Publishing.[1]

2.1 Introduction

The ability to produce endofullerenes by molecular surgery[34] has resulted in a number of exciting results, both experimental[32,35] and theoretical.[36–41] The generation of carbon nanotube "peapods" has also recently been shown to be possible.[42–47] The combination of these ideas leads to endofullerene peapods: carbon nanotubes which contain fullerene cages with atoms or molecules trapped inside.[48–54]

By treating these nanomolecular assemblies (NMAs) as fixed and rigid, we may study the motion of the atoms and molecules enclosed therein. The resulting model is similar in many respects to that obtained from placing ultracold particles in an optical lattice,[55,56] and has previously been studied in that context.[30,57] Nevertheless, there are some fundamental differences: in an NMA, the imprisoned entities may not move between sites as they do in an optical lattice, so there cannot be double occupation of a site; the spacing between adjacent sites is much smaller in an NMA (on the order of 1 nm)[47] than in a typical optical lattice (on the order of 100 nm);[55] and the carbon walls of the fullerene cages shield the interactions between the captive particles.[32]

For an endofullerene peapod NMA model as described above, one is in principle left with translational, vibrational, and rotational degrees of freedom for the confined particles. At very low temperatures, the translations and vibrations are restricted to their respective ground states, and only the rotational motion remains relevant, such that one can approximate the low-lying energy spectrum with an

effective rotor Hamiltonian. In the following, we therefore focus on the rotational degrees of freedom of molecules arranged in a linear chain. Specifically, we choose dipolar linear rotor molecules (such as HF, LiCl, or CsI) which interact pairwise through the (dimensionless) dipole–dipole potential

$$V_{ij}(\mathbf{e}_i, \mathbf{e}_j; \mathbf{r}_{ij}) = \mathbf{e}_i \cdot \mathbf{e}_j - 3(\mathbf{e}_i \cdot \mathbf{r}_{ij})(\mathbf{e}_j \cdot \mathbf{r}_{ij}), \tag{2.1}$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are unit vectors describing the orientation of two rotors and $\mathbf{r}_{ij}$ is the unit vector in the direction from one rotor to the other.

In this article, we propose a method for the calculation of ground state energies and wavefunctions for long one-dimensional systems of dipolar rotors using the density matrix renormalization group (DMRG). Originally introduced by White in 1992,[58] the approach of DMRG has proven fruitful in a number of applications ranging from condensed matter physics[25,27] to quantum chemistry.[28,59,60] Although it has been extended to the study of two-dimensional systems, finite temperature systems, and real-time evolution, DMRG excels at finding ground states of strongly-correlated one-dimensional systems.[61]

For small systems of this kind (up to around 10 rotors), sparse iterative methods for Hamiltonian diagonalization are sufficient to obtain a handful of low-lying eigenstates.[6] As the systems grow, the size of the many-body basis increases exponentially, and the problem quickly becomes intractable. Hence, we turn to DMRG in order to grow the rotor chain under study to 100 rotors, which is made feasible by the matrix product state (MPS) wavefunction ansatz inherent to DMRG. We accomplish this with the ITensor package, which allows us to efficiently formulate the Hamiltonian as a matrix product operator (MPO) and which contains an implementation of DMRG.[62] While existing publications have also examined many-body quantum systems with dipole–dipole interactions using DMRG,[57,63,64] due to their use of different geometries and focus on mapping to other model systems, they do not capture the full physics of interacting molecules under quantum rotation.

This paper is organized as follows: in Sec. 2.2, we provide a brief introduction to matrix product states; in Sec. 2.3 we derive an expression for the dipole–dipole interaction potential in terms of angular momentum ladder operators; in Sec. 2.4, we show the results of our DMRG calculations for the rotor system; in Sec. 2.5, we end with some concluding remarks.

2.2 Matrix product states

Despite originally being developed as a renormalization group technique, DMRG is now typically understood in terms of MPSs.[25] When a wavefunction $|\psi\rangle$, which is an abstract element of a Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$, is represented in a finite (perhaps truncated) basis, it may be treated as a vector $C_{\mathbf{n}} = \langle \mathbf{n}|\psi\rangle$ of coefficients indexed by the multi-index $\mathbf{n} = (n_1, n_2, \ldots, n_N)$. Such

a vector may always be expanded exactly into a product of matrices

$$\langle \mathbf{n}|\psi\rangle = \underline{\mathbf{A}}^{(1),n_1}\underline{\mathbf{A}}^{(2),n_2}\cdots\underline{\mathbf{A}}^{(N),n_N}. \tag{2.2}$$

Even though they do not necessarily correspond to physical sites on a lattice, the entities living in the N one-body Hilbert spaces $\mathcal{H}_k$ are referred to as "sites." The collection $\{\underline{\mathbf{A}}^{(k)}\}$ of rank-3 tensors forms an MPS representation of $|\psi\rangle$, where n_k is the *physical* index for the site and the two matrix indices (implied by the matrix multiplication) are *bond* indices with "bond dimension" M_k. [Each $\underline{\mathbf{A}}^{(1),n_1}$ is treated as a row vector and each $\underline{\mathbf{A}}^{(N),n_N}$ as a column vector, so that the product in Eq. (2.2) results in a scalar.] In practice, this expansion will not be made exact, as that leads to exponential scaling with system size; instead, the matrices are truncated to keep the bond dimension small.

We may arbitrarily group the sites into regions A and B, and write the many-body Hilbert space as the product $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. At the core of DMRG is the Schmidt decomposition

$$|\psi\rangle = \sum_i \sqrt{\lambda_i}\left|\varphi_i^A\right\rangle \otimes \left|\varphi_i^B\right\rangle, \tag{2.3}$$

where the λ_i are non-negative real numbers, and $\{\left|\varphi_i^A\right\rangle\}$ and $\{\left|\varphi_i^B\right\rangle\}$ are orthonormal sets in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Truncation of the terms with the smallest coefficients results in an optimal approximation $|\tilde{\psi}\rangle$ in the sense of the 2-norm: $\||\psi\rangle - |\tilde{\psi}\rangle\|_2$ is minimized compared to other truncation schemes that retain the same number of terms. This truncation is typically performed so that

$$\sum_i \lambda_i \leq \varepsilon, \tag{2.4}$$

where the sum is over the discarded values and $\varepsilon \ll 1$. Hence, this decomposition provides a systematic way to generate an efficient MPS. DMRG makes repeated use of this decomposition (along with iterative diagonalization, for example using the Lanczos algorithm) by "sweeping" the boundary between A and B from one end of the system to the other in order to obtain the ground state of a Hamiltonian.

The number of terms that remain in the sum dictates the bond dimension and has a direct impact on the difficulty of the calculation. The bond dimension also bounds the amount of entanglement that may be present across the corresponding boundary between the regions connected by the bond. The standard measure of bipartite entanglement for pure states is the von Neumann entropy

$$S_{\mathrm{vN}} = -\operatorname{Tr}(\hat{\varrho}_A \log \hat{\varrho}_A), \tag{2.5}$$

where $\hat{\varrho}_A = \mathrm{Tr}_B\,|\psi\rangle\langle\psi|$ is the reduced density operator for region A.[65] The entanglement entropy cannot exceed $\log M$, where M is the dimension of the Hilbert space $\mathcal{H}_A$. When the boundary is at bond k, the dimension of the subspaces spanned by $\{|\varphi_i^A\rangle\}$ and $\{|\varphi_i^B\rangle\}$ is M_k, so the maximum possible von Neumann entropy is $S_{\mathrm{vN}} = \log M_k$. Conversely, the minimum bond dimension required to faithfully represent a state with entanglement entropy S_{vN} across bond k is $M_k = e^{S_{\mathrm{vN}}}$. As these bounds are saturated only for maximally-entangled states, the bond dimension will in general be larger than $e^{S_{\mathrm{vN}}}$.

In one spatial dimension and away from criticality, ground states of gapped Hamiltonians with short-range interactions are known to obey an entanglement area law.[66] That is, the amount of entanglement between regions is proportional to the size of the boundary between them and is independent of the sizes of the regions themselves. Because a single boundary in one dimension must always have a constant size, the entanglement entropy of such states cannot depend on the size of the system and therefore the bond dimension will not increase when the system is made larger.

For long-range interactions, the picture is less clear. Recent work has shown that, under appropriate conditions, ground states of one-dimensional Hamiltonians with interactions that decay faster than $1/r^4$ must satisfy an entanglement area law.[67] For systems with even longer range interactions, such as the $1/r^3$ considered in the present work, we are not aware of any proofs regarding entanglement area laws away from criticality. However, near a critical point, where conformal field theory takes over, the entanglement entropy scales logarithmically with system size in one dimension,[68] which causes the bond dimension to grow linearly, increasing the computational cost of DMRG.

2.3 Dipolar rotors

For N identical rotors with rotational constant B and dipole moment μ, the general Hamiltonian is

$$\hat{H} = \frac{B}{\hbar^2}\sum_{i=1}^{N}\hat{\ell}_i^2 + \frac{\mu^2}{4\pi\varepsilon_0}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\frac{\hat{V}_{ij}}{r_{ij}^3}, \tag{2.6}$$

where r_{ij} is the distance between rotors i and j. Since a peapod NMA is inherently linear, without loss of generality, we may place the rotors along the z axis and express the potential operator compactly as

$$\hat{V}_{ij}^{(z)} = \hat{x}_i\hat{x}_j + \hat{y}_i\hat{y}_j - 2\hat{z}_i\hat{z}_j. \tag{2.7}$$

Because of the regular structure of a peapod NMA, we space the rotors evenly and write the Hamiltonian as

$$\frac{\hat{H}}{B} = \sum_{i=1}^{N} \frac{\hat{\ell}_i^2}{\hbar^2} + \frac{1}{R^3} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \frac{\hat{V}_{ij}^{(z)}}{(i-j)^3}, \tag{2.8}$$

where

$$R = r \left(\frac{4\pi\varepsilon_0 B}{\mu^2} \right)^{\frac{1}{3}}, \tag{2.9}$$

r is the distance between adjacent rotors (the lattice spacing), and we have taken this opportunity to non-dimensionalize the Hamiltonian. Since all the physical properties appear only in R, the one-parameter form of the Hamiltonian allows us to explore the entire realm of physical realizations of this model by scanning a single parameter.

As the effective rotor Hamiltonian lacks the microscopic details describing the confinement of the rotors, it is general enough to apply to a variety of experimental situations in addition to NMAs, such as molecules trapped in a sufficiently deep optical lattice in which tunnelling between minima is suppressed. The specifics of the physical configuration are encoded in R; for example, in an NMA, the dipole moment μ is screened by the fullerene cages,[32] which causes an increase in R.

A natural one-body basis for this problem is that of the spherical harmonics $|\ell_i\, m_i\rangle$, in which the squared angular momentum operator $\hat{\ell}_i^2$ is diagonal

$$\hat{\ell}_i^2 \,|\ell_i\, m_i\rangle = \hbar^2 \ell_i(\ell_i+1)\,|\ell_i\, m_i\rangle\,. \tag{2.10}$$

Although in principle this basis is infinite, in order to carry out any calculations, it will need to be truncated at a finite $\ell_{\max}$ so that it is large enough to accurately represent the quantities in question, but no larger.

Thanks to the form of the potential operator, the Hamiltonian in Eq. (2.8) conserves the total ℓ parity

$$\ell_{\mathrm{p}} \equiv \sum_{i=1}^{N} \ell_i \pmod 2 \tag{2.11a}$$

and the total m value

$$m = \sum_{i=1}^{N} m_i. \tag{2.11b}$$

In order to exploit the block-diagonal structure of the Hamiltonian in DMRG, we must make explicit use of these good quantum numbers. That is, we need to express the potential operator in terms of one-body operators that only change the quantum numbers ℓ_p and m by a definite amount, termed the "flux." This makes it possible to construct both the wavefunction MPS and Hamiltonian MPO as sparse objects, reducing the amount of storage required and significantly accelerating the calculation.[25] Terms like $\hat{x}_i\hat{x}_j$ do not suffice, because the position operators ($\hat{x}_i$, $\hat{y}_i$, $\hat{z}_i$) do not have a well-defined flux. The action of one of these operators on a state with definite ℓ_p and m quantum numbers does not result in a state with definite ℓ'_p and m' values.

The ladder operators $\hat{\ell}_i^\pm$ and $\hat{m}_i^\pm$, which raise and lower ℓ_i and m_i, are obvious candidates for building blocks, as their flux is immediately evident. The latter operators have the well-known form

$$\hat{m}_i^\pm = \hat{\ell}_{i,x} \pm i\hat{\ell}_{i,y} \tag{2.12}$$

and they act as[69]

$$\hat{m}_i^\pm \left|\ell_i\, m_i\right\rangle = \hbar\sqrt{(\ell_i \pm m_i + 1)(\ell_i \mp m_i)}\left|\ell_i, m_i \pm 1\right\rangle. \tag{2.13}$$

On the other hand, the ladder operators for ℓ_i do not appear to have been as deeply analyzed. There exist the definitions[70]

$$\hbar\hat{R}_{i,z} = i(\hat{x}_i\hat{\ell}_{i,y} - \hat{y}_i\hat{\ell}_{i,x}) + \frac{\hat{z}_i}{2}\left(\hbar + \sqrt{4\hat{\ell}_i^2 + \hbar^2}\right) \tag{2.14a}$$

and

$$\hbar\hat{Q}_{i,z} = i(\hat{x}_i\hat{\ell}_{i,y} - \hat{y}_i\hat{\ell}_{i,x}) + \frac{\hat{z}_i}{2}\left(\hbar - \sqrt{4\hat{\ell}_i^2 + \hbar^2}\right), \tag{2.14b}$$

but unfortunately $\hat{R}_{i,z}^\dagger \neq \hat{Q}_{i,z}$. We instead introduce the operators

$$\hat{\ell}_i^\pm = \frac{\hbar}{2}\hat{z}_i\left(1 \pm \frac{\hbar}{\sqrt{4\hat{\ell}_i^2 + \hbar^2}}\right) \pm i(\hat{x}_i\hat{\ell}_{i,y} - \hat{y}_i\hat{\ell}_{i,x})\frac{\hbar}{\sqrt{4\hat{\ell}_i^2 + \hbar^2}}, \tag{2.15}$$

which are intimately related to $\hat{R}_{i,z}$ and $\hat{Q}_{i,z}$, but satisfy $(\hat{\ell}_i^+)^\dagger = \hat{\ell}_i^-$.

From this definition, it follows that

$$\hat{x}_i = \frac{1}{2\hbar^2}\Big[(\hat{\ell}_i^+ + \hat{\ell}_i^-), (\hat{m}_i^+ - \hat{m}_i^-)\Big], \tag{2.16a}$$

$$\hat{y}_i = \frac{1}{2i\hbar^2}\Big[(\hat{\ell}_i^+ + \hat{\ell}_i^-), (\hat{m}_i^+ + \hat{m}_i^-)\Big], \tag{2.16b}$$

and

$$\hat{z}_i = \frac{1}{\hbar}(\hat{\ell}_i^+ + \hat{\ell}_i^-). \tag{2.16c}$$

The clean and concise form of these expressions suggests that our choice of the ladder operators $\hat{\ell}_i^\pm$ is an appropriate one. The potential from Eq. (2.1) may then be written as

$$\hat{V}_{ij} = (1 - 3r_{ij,z}^2)\hat{B}_i^0\hat{B}_j^0 - \frac{1}{4}\left[(1 - 3r_{ij,z}^2)\hat{B}_i^-\hat{B}_j^+ + 3r_{ij,\perp}^2\hat{B}_i^-\hat{B}_j^- + 6r_{ij,\perp}r_{ij,z}(\hat{B}_i^-\hat{B}_j^0 + \hat{B}_i^0\hat{B}_j^-) + \text{h.c.}\right], \tag{2.17}$$

where

$$\hat{B}_i^\pm = \pm\frac{1}{\hbar^2}\left(\left[\hat{\ell}_i^+, \hat{m}_i^\pm\right] + \left[\hat{\ell}_i^-, \hat{m}_i^\pm\right]\right), \tag{2.18a}$$

$$\hat{B}_i^0 = \frac{1}{\hbar}(\hat{\ell}_i^+ + \hat{\ell}_i^-), \tag{2.18b}$$

and

$$r_{ij,\perp} = r_{ij,x} + ir_{ij,y}. \tag{2.19}$$

The simplified form for rotors aligned along the z axis is

$$\hat{V}_{ij}^{(z)} = -2\hat{B}_i^0\hat{B}_j^0 + \frac{1}{2}\left[\hat{B}_i^-\hat{B}_j^+ + \text{h.c.}\right]. \tag{2.20}$$

When written in this form, the potential operator may be constructed as a sparse MPO.

2.4 Results

In this section, we give some results for the ground state properties of the rotor system, as well as transition dipole moments (TDMs) for an excitation spectrum. The results were computed using DMRG with the MPS truncation parameter ε from Eq. (2.4) set to 10^{-10}.

2.4.1 Ground state properties

The primary result of the DMRG routine is the ground state energy E_0. As these energies are expected to decrease with increasing system size N, we present them in the form of chemical potentials in Fig. 2.1. It is evident that a smaller value of R (stronger interactions) requires a larger $\ell_{\max}$ (more basis states), as expected. For sufficiently large systems, we expect on physical grounds that the addition of a single particle will result in a constant decrease in the energy of the

system, regardless of the system size. In other words, because a newly added rotor should only be substantially correlated with finitely many rotors on the end of the system, the chemical potential should tend to a constant in the large N limit. For $R = 0.5$ and $R = 2$, this limit is reached by 25 rotors, but for $R = 1$, the chemical potential continues to change even at 50 rotors, indicating longer-ranged correlations.

The maximum bond dimension $M_{\max}$, shown in Fig. 2.2, is the largest bond dimension across the entire MPS, and it is indicative of the amount of long-range correlations in the state. That $M_{\max}$ plateaus quickly for $R = 0.5$ and $R = 2$ implies the presence of only short-range correlations, but the same cannot be said for $R = 1$. The von Neumann entanglement entropy S_{vN} for the partitioning of the system into halves behaves similarly to the bond dimension, as shown in Fig. 2.3. Indeed, we see what appears to be area law scaling at $R = 0.5$ and $R = 2$, as the entanglement does not change with system size once the system is large enough to make finite size effects negligible. At $R = 1$, we instead notice what looks like an area law violation, possibly signifying a phase transition.

The quantum rotor model, which resembles the model used in the present article, but lacks the anisotropic term in Eq. (2.1), is known to have no ordered phase in one dimension and therefore no phase transition.[71] In light of this, the observed anomalies at $R = 1$ are peculiar, but it has been suggested that the breaking of rotational symmetry in the anisotropic model is responsible for a second-order phase transition between ordered and disordered phases.[30] This is corroborated by the sudden change in both the expectation value of the orientational correlation operator

$$\frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j, \tag{2.21}$$

and the von Neumann entanglement entropy S_{vN} near $R = 1$, as demonstrated in Fig. 2.4. Of the two, it seems that the latter is a sharper indicator of the apparent phase transition. For larger system sizes, the entropy even peaks at $R = 1$, clearly delineating the boundary between strongly and weakly interacting systems.

Though the maximum MPS bond dimension $M_{\max}$ is not a physical parameter, it contains valuable information about the effectiveness of DMRG for the system in question. As can be seen in Fig. 2.5, away from $R = 1$, the maximum bond dimension converges very quickly, and we expect linear scaling of computational time. On the contrary, the peak at $R = 1$ indicates that the scaling will not be as favorable, which is to be expected for DMRG near a second order phase transition.

At such a phase transition, the spatial correlation length should also diverge.[72] Thus, in the near future we plan to examine the behaviour of the correlation length around $R = 1$ to confirm the existence of the transition and identify the value of the critical parameter R_{c}. We then hope to extract the central charge of the relevant conformal field theory for the critical system.[68]

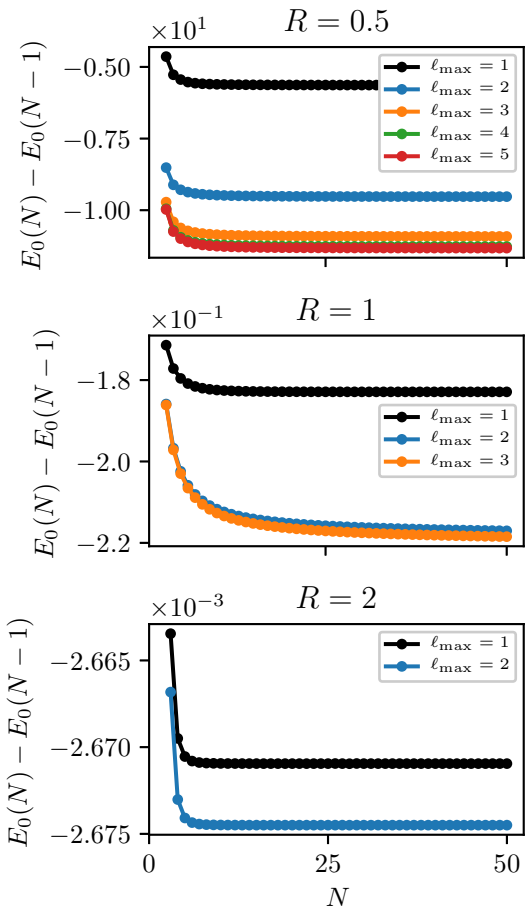

Figure 2.1: Chemical potential of rotor chains of length N. Several curves with different $\ell_{\max}$ are shown to demonstrate the effect of basis truncation. The chemical potential at $R = 1$ takes longer than the others to plateau.

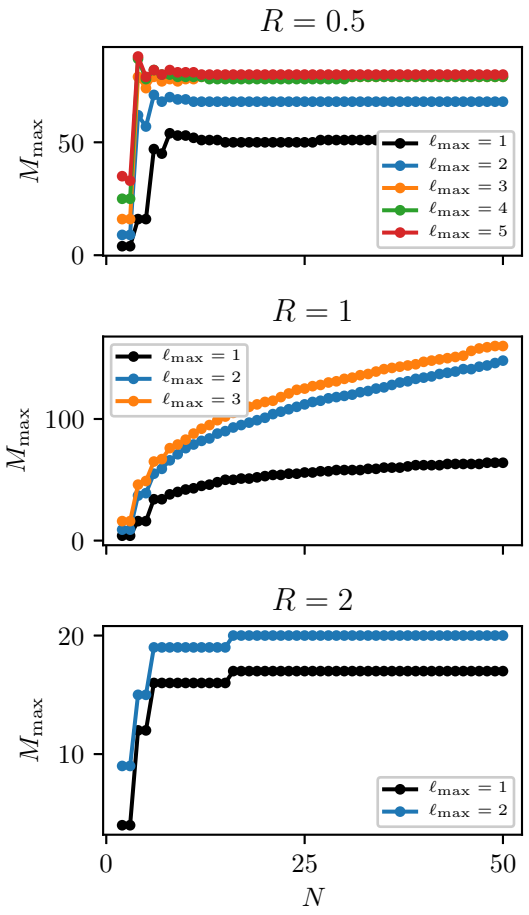

Figure 2.2: Maximum MPS bond dimension for rotor chains of length N. Several curves with different $\ell_{\max}$ are shown to demonstrate the effect of basis truncation. The bond dimension at $R = 1$ is larger than the others and not constant by $N = 50$.

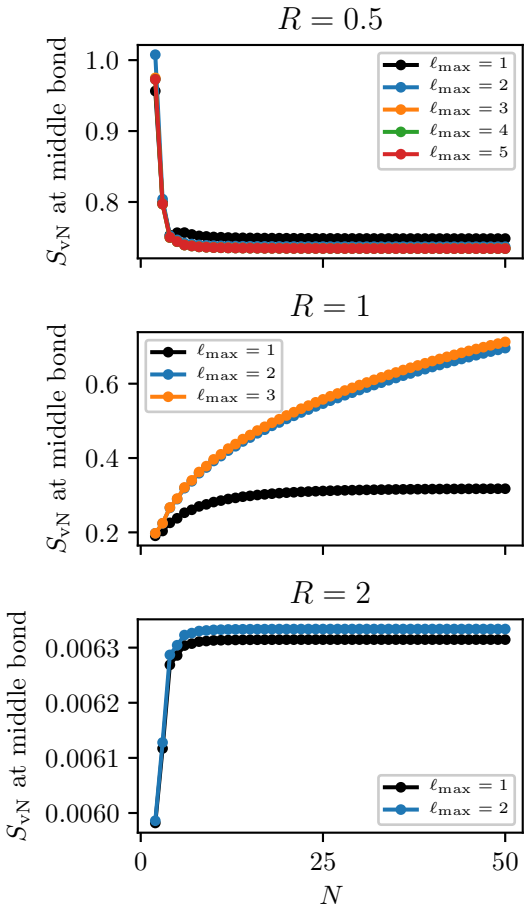

Figure 2.3: von Neumann entanglement entropy of rotor chains of length N. Several curves with different ℓ_{max} are shown to demonstrate the effect of basis truncation. The entropy at $R = 1$ is still increasing by $N = 50$.

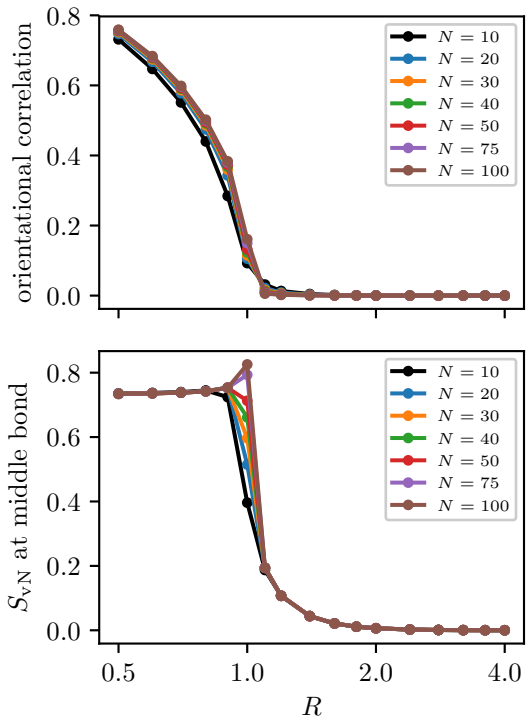

Figure 2.4: Comparison of the orientational correlation (top panel) and the entanglement entropy (bottom panel) for rotor chains of size N across the apparent phase transition around $R = 1$.

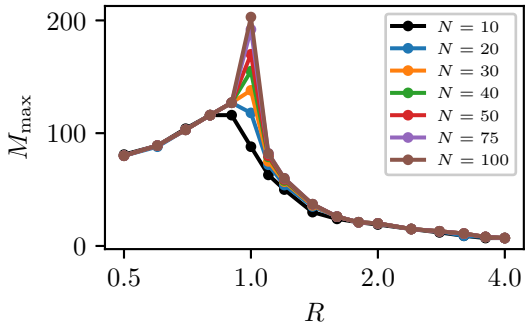

Figure 2.5: Maximum MPS bond dimension for rotor chains of size N across the apparent phase transition around $R = 1$. There is a clear divergence with system size near $R = 1$.

Table 2.1: Energy differences and transition dipole moments for rotor chains of length N at $R = 2$. Each value is computed between the ground state of the entire Hamiltonian and the ground state of the specified block at $\ell_{\max} = 12$ and with 30 DMRG sweeps.

N	ℓ_p	m	ΔE_0	TDM
5	1	0	1.856	5.935×10^{-2}
	1	± 1	1.950	3.117×10^{-4}
10	1	0	1.843	2.296×10^{-2}
	1	± 1	1.950	7.340×10^{-5}
15	1	0	1.824	1.611×10^{-2}
	1	± 1	1.948	3.504×10^{-5}

2.4.2 Excitation spectrum

Because the Hamiltonian in Eq. (2.8) is block-diagonal, we can also use DMRG to target the ground state of any symmetry block. The $\ell_\mathrm{p} = 0, m = 0$ block contains the ground state of the entire Hamiltonian, but the $\ell_\mathrm{p} = 1, m = 0$ and $\ell_\mathrm{p} = 1, m = \pm 1$ blocks are also of interest, because their states $\left|n_{\ell_\mathrm{p},m}\right\rangle$ are reachable from the ground state $\left|0_{\ell_\mathrm{p}=0,m=0}\right\rangle$ by application of the $\hat{x}_i$ and $\hat{z}_i$ operators. That is, while most transition dipole moments (TDMs) are forbidden by symmetry, the moments

$$\left| \left\langle n_{\ell_\mathrm{p}=1,m=0} \right| \frac{1}{N} \sum_{i=1}^{N} \hat{z}_i \left| 0_{\ell_\mathrm{p}=0,m=0} \right\rangle \right|^2 \tag{2.22a}$$

and

$$\left| \left\langle n_{\ell_\mathrm{p}=1,m=\pm 1} \right| \frac{1}{N} \sum_{i=1}^{N} \hat{x}_i \left| 0_{\ell_\mathrm{p}=0,m=0} \right\rangle \right|^2 \tag{2.22b}$$

do not necessarily vanish.

Computed energy differences ΔE_0 and TDMs for $R = 2$ are listed in Table 2.1. Despite this calculation providing only two peaks of a dipole excitation spectrum for each system size, it lays the foundation for a series of more involved calculations which can reveal more information from the spectrum. We propose two complementary approaches for this. The first involves the direct calculation of excited states within the identified symmetry blocks. From these, more energy differences and TDMs can be computed, gradually populating a stick spectrum. The second requires time evolution of the ground state to obtain a correlation function, followed by a Fourier transform which yields a spectrum with finite resolution. Both approaches are presently possible using standard extensions to DMRG for excited states and real-time evolution.[25]

2.5 Conclusions

We have shown that DMRG can be applied to systems of dipolar linear rigid rotors which are pinned to the sites of a one-dimensional lattice. The method can be used to obtain ground state properties, such as the energy and von Neumann entanglement entropy. Excited states which are ground states of their respective symmetry blocks may also be reached by this method.

In order to perform the DMRG calculations efficiently, we have expressed the dipole–dipole potential in terms of angular momentum ladder operators $\hat{\ell}_i^{\pm}$ and $\hat{m}_i^{\pm}$. These operators have a well-defined quantum number flux and they permit the use of an algorithm that conserves the total quantum numbers of the MPS. By making use of the symmetries of the Hamiltonian, which conserve ℓ_p and m, we are able to apply our approach to a chain of 100 rotors.

Away from $R = 1$, the computational time of the method scales linearly with system size and ground states of much longer chains with several hundred rotors may be computed relatively quickly. However, near $R = 1$, the scaling is worse than linear and computing properties of larger systems becomes challenging. We attribute the favorable growth to area law scaling of the entanglement and its breakdown to a violation of the area law that is expected for a critical system.

An important extension to the above model is the addition of translational motion for the rotor molecules.[38] This would take into account the interactions between the rotor molecule and the fullerene cage and would greatly enhance the applicability of the model to NMA experiments. The primary difficulty in implementing this change is the dynamical nature of the intermolecular separations: the classical parameter R must be augmented by quantum mechanical operators which describe the deviations from the cage centers. Although it is not currently clear how such an implementation would look, it is likely to involve a model for the rotation–translation coupling which can be expressed in terms of raising and lowering operators for the site-local translational states.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Ministry of Research and Innovation (MRI), the Canada Research Chair program, the Canada Foundation for Innovation (CFI), and the Canada First Research Excellence Fund (CFREF). We thank Martin Ganahl and Roger Melko for informative discussions.

Chapter 3

Deterministic and quasi-random sampling of optimized Gaussian mixture distributions for vibronic Monte Carlo

This chapter is reproduced from Dmitri Iouchtchenko, Neil Raymond, Pierre-Nicholas Roy, and Marcel Nooijen, "Deterministic and quasi-random sampling of optimized Gaussian mixture distributions for vibronic Monte Carlo", arXiv:1912.11594v1.[2]

3.1 Introduction

Although the Born–Oppenheimer approximation (BOA) has proven itself to be an invaluable tool in computational chemistry, certain classes of systems cannot be accurately described under its restrictions.[12] Systems with coupling between vibrational and electronic degrees of freedom (vibronic coupling) have regions of nonadiabaticity, where adiabatic surfaces approach each other in energy, possibly crossing; this situation is often observed in photochemical reactions.[73–76] Examples in biologically relevant molecules include the photoisomerization of retinal and the photostability of DNA base pairs.[77,78]

When vibronic coupling is present, it is necessary to use methods that are not bound by the BOA.

Several approaches use a vibronic description of the form

$$\hat{H}^{aa'} = \delta_{aa'} \sum_{j=1}^{N} \frac{\omega_j}{2} (\hat{p}_j^2 + \hat{q}_j^2) + E^{(0)aa'} + \sum_{j=1}^{N} E_j^{(1)aa'} \hat{q}_j + \sum_{j=1}^{N} \sum_{j'=1}^{N} E_{jj'}^{(2)aa'} \hat{q}_j \hat{q}_{j'} + \sum_{j=1}^{N} \sum_{j'=1}^{N} \sum_{j''=1}^{N} E_{jj'j''}^{(3)aa'} \hat{q}_j \hat{q}_{j'} \hat{q}_{j''} + \cdots , \tag{3.1}$$

where A diabatic surfaces (labelled by $a = 1, \ldots, A$) describing N vibrational degrees of freedom are coupled via the off-diagonal ($a \neq a'$) elements of $\hat{H}$.[75,76,79] For convenience, all the quantities in Eq. (3.1) (i.e. the position and momentum coordinates, as well as the various parameters, and hence also the resulting Hamiltonian operator) are taken to be dimensionless, with $\hbar = 1$.

In recent years, multiple proposals have been put forward to find properties of such vibronic Hamiltonians using path integrals, and in particular path integral Monte Carlo (PIMC).[7,80–83] While most of these focus primarily on dynamical properties, in Ref. [7] it was shown that PIMC may be used to compute the canonical partition function Z of a system with vibronic coupling.

In the present work, we build on the approach established in Ref. [7], offering several improvements. To reduce the overall stochastic error, we augment the random sampling of a Gaussian mixture distribution (GMD) with two deterministic techniques. The first is a way to select the component of the GMD for optimal reduction of the variance of the mean. The other is quasi-Monte Carlo, which uses low-discrepancy sequences (quasi-random numbers) instead of pseudo-random numbers, and has seen success in physical, chemical, and financial applications.[84–88]

Furthermore, we describe a generic approach for optimizing the parameters of the GMD used in importance sampling to ensure that regions of high integrand magnitude can be sufficiently well explored, even when their locations are not already known. This is accomplished by means of the simultaneous perturbation stochastic approximation (SPSA) algorithm, which repeatedly varies the GMD parameters in order to minimize a loss function.[19,20]

The remainder of this article is structured as follows: in Sec. 3.2, we provide the theoretical background for the subsequent sections; in Sec. 3.3, we explain the proposed enhancements; in Sec. 3.4, we apply these enhancements to model systems; and in Sec. 3.5, we give some concluding remarks.

3.2 Background

To bring the reader up to speed with the variant of vibronic PIMC discussed in this work, we include a brief derivation in Sec. 3.2.1. Since randomized quasi-Monte Carlo is not frequently seen

in computational chemistry literature, we give a short overview in Sec. 3.2.2. In Sec. 3.2.3, we review the simultaneous perturbation stochastic approximation algorithm.

3.2.1 Vibronic path integral Monte Carlo

The goal of Ref. [7] is to compute the partition function

$$Z = \operatorname{Tr} e^{-\beta \hat{H}} \tag{3.2}$$

of the vibronic Hamiltonian $\hat{H}$ in Eq. (3.1) at reciprocal temperature $\beta = 1/k_{\mathrm{B}}T$. This is accomplished by inserting resolutions of the identity

$$\hat{\mathbb{1}} = \sum_{a=1}^{A} \int \mathrm{d}\mathbf{q} \; |a\,\mathbf{q}\rangle\langle a\,\mathbf{q}| \tag{3.3}$$

in the combined diabatic basis and (normal mode) position representation to arrive at the expression

$$Z = \sum_{a_1=1}^{A} \cdots \sum_{a_P=1}^{A} \int \mathrm{d}\mathbf{q}_1 \cdots \int \mathrm{d}\mathbf{q}_P \; \langle a_P\,\mathbf{q}_P|e^{-\tau\hat{H}}|a_1\,\mathbf{q}_1\rangle \prod_{i=1}^{P-1} \left\langle a_i\,\mathbf{q}_i\middle|e^{-\tau\hat{H}}\middle|a_{i+1}\,\mathbf{q}_{i+1}\right\rangle, \tag{3.4}$$

which has the form of a discretized imaginary time path integral with P beads and time step $\tau = \beta/P$. The Trotter factorization is then applied to obtain an approximation that is exact in the $P \to \infty$ limit. It takes on the form

$$Z = \int \mathrm{d}\mathbf{R} \; g(\mathbf{R}), \tag{3.5}$$

in which: we use $\mathbf{R}$ to mean the vector containing all the continuous path coordinates $\mathbf{q}_1$, …, $\mathbf{q}_P$; the integrand is

$$g(\mathbf{R}) = \operatorname{Tr}\left[\mathbb{O}(\mathbf{q}_P, \mathbf{q}_1)\mathbb{M}(\mathbf{q}_1) \prod_{i=1}^{P-1} \mathbb{O}(\mathbf{q}_i, \mathbf{q}_{i+1})\mathbb{M}(\mathbf{q}_{i+1})\right]; \tag{3.6}$$

and the matrix-valued functions $\mathbb{O}$ and $\mathbb{M}$ have the elements

$$\mathbb{O}(\mathbf{q}, \mathbf{q}')_{aa'} = \delta_{aa'} \langle \mathbf{q}|e^{-\tau\hat{h}^a}|\mathbf{q}'\rangle, \tag{3.7a}$$

$$\mathbb{M}(\mathbf{q})_{aa'} = \langle a|e^{-\tau\hat{V}(\mathbf{q})}|a'\rangle. \tag{3.7b}$$

The operator $\hat{h}$ is diagonal in the diabatic basis and has the form

$$\hat{h}^a = \sum_{j=1}^{N} \frac{\omega_j}{2}(\hat{p}_j^2 + \hat{q}_j^2) + E^{(0)aa} + \sum_{j=1}^{N} E_j^{(1)aa}\hat{q}_j, \tag{3.8}$$

whereas the operator

$$\hat{V} = \hat{H} - \hat{h} \tag{3.9}$$

is diagonal in the position representation. The shorthand notation

$$\hat{h}_\mathrm{o} = \sum_{j=1}^{N} \frac{\omega_j}{2}(\hat{p}_j^2 + \hat{q}_j^2) \tag{3.10}$$

is sometimes used for the harmonic oscillator terms.

The integral in Eq. (3.5) is then approximated by N_MC steps of Monte Carlo with importance sampling from the probability density function (pdf) $\pi(\mathbf{R})$:

$$Z = \langle f \rangle_\pi = \int \mathrm{d}\mathbf{R}\ \pi(\mathbf{R}) f(\mathbf{R}) \approx \frac{1}{N_\mathrm{MC}} \sum_{i=1}^{N_\mathrm{MC}} f(\mathbf{R}_i) = \bar{f}, \tag{3.11}$$

where

$$f(\mathbf{R}) = \frac{g(\mathbf{R})}{\pi(\mathbf{R})}. \tag{3.12}$$

In principle, $\pi(\mathbf{R})$ may be any normalized pdf that does not vanish on the support of $g(\mathbf{R})$.[‡] In practice, the distribution π must be chosen to allow efficient sampling and to have significant overlap with g. If the latter condition is not fulfilled, the result is spectacular failure of the method, as shown in Ref. [7].

The ease and efficiency with which one can sample from a Gaussian mixture distribution (GMD), combined with the GMD-like form obtained for g when $\hat{V} = 0$, make a GMD a natural choice for π. The general form of a GMD pdf is the convex combination

$$\pi(\mathbf{R}) = \sum_{b=1}^{B} w^b \pi^b(\mathbf{R}) \tag{3.13}$$

of multivariate Gaussian pdfs $\pi^b(\mathbf{R})$. The straightforward approach for randomly sampling a point $\mathbf{R}$ from π is to first choose a component π^b, and then sample from π^b according to its mean vector

[‡]The unnormalized function $\varrho(\mathbf{R})$ that appears in Ref. [7] is related to $\pi(\mathbf{R})$ by the normalization: $\varrho(\mathbf{R}) = Z_\varrho \pi(\mathbf{R})$.

$\mathbf{d}^b$ and covariance matrix $\mathbf{\Sigma}^b$. Although each component pdf

$$\pi^b(\mathbf{R}) = \frac{1}{\sqrt{(2\pi)^{PN} \det \mathbf{\Sigma}^b}} e^{-\frac{1}{2}(\mathbf{R}-\mathbf{d}^b)^{\mathrm{T}}(\mathbf{\Sigma}^b)^{-1}(\mathbf{R}-\mathbf{d}^b)} \tag{3.14}$$

has a fixed number of parameters (the mean vector $\mathbf{d}^b$ has length PN and the covariance matrix $\mathbf{\Sigma}^b$ is $PN \times PN$), the total number of parameters is arbitrary, scaling linearly with B.

Broadly speaking, there are three issues to be addressed when using the GMD π as a sampling distribution: how to select a component π^b, how to sample from π^b, and how to choose the parameters of each component. In this work, we consider some improvements to all of these areas.

3.2.2 Randomized quasi-Monte Carlo

It is often more important for Monte Carlo (MC) that sampled points be consistently distributed rather than randomly distributed. To take advantage of this, low-discrepancy sequences (LDSs) may be used as a smoother substitute for pseudo-random sequences, such as those output by a pseudo-random number generator (RNG). The aim of an LDS is to produce values which are as evenly spaced as possible within some volume (typically the D-dimensional hypercube $[0,1)^D$).[21] In this context, the values forming an LDS are called "quasi-random numbers" and their application to MC gives rise to quasi-Monte Carlo (qMC[‡]).[89] The chief argument in favor of qMC is that the estimates it provides have an error whose asymptotic scaling with the number of samples N_{MC} is expected to be better than the $N_{\mathrm{MC}}^{-\frac{1}{2}}$ scaling seen in plain MC.[85,90]

In the present work, we employ Sobol sequences to produce quasi-random numbers.[91] Other sequences may also be used, such as those of Halton or Faure, but Sobol sequences have been observed to perform better for high-dimensional problems[92] and there exists a convenient software package to generate them.[93] Some authors suggest skipping the initial points of a low-discrepancy sequence,[94] but we have noticed no change when doing so, and therefore choose not to skip any points. We have included a demonstration of qMC in Appendix 3.A, where we show its application to the classic MC problem of estimating the number π, and a very brief discussion in Appendix 3.B, where we mention some associated difficulties with Gaussian distributions and correlations in many dimensions. For more information about qMC, the reader is directed to Refs. [85, 89, 95].

The primary concern with qMC is its inability to provide error estimates. Some theoretical bounds are known, but they are not useful in practice.[96] Instead, one may use randomized quasi-Monte Carlo (RqMC), which reintroduces pseudo-random numbers in order to compute statistical error bars. The idea is simple: several low-discrepancy sequences are run in tandem, each with a different

[‡]We use a lower-case "q" in "qMC" to avoid confusion with quantum Monte Carlo, which is frequently abbreviated as "QMC".

shift $\mathbf{v}$ generated by an RNG.[96,97] An LDS where every point $\mathbf{u}$ is shifted by the same displacement vector $\mathbf{v}$ to produce

$$\mathbf{w}' = \mathbf{u} + \mathbf{v} \tag{3.15}$$

is still an LDS in the new shifted hypercube $[0,1)^D + \mathbf{v}$. If we take the elements of $\mathbf{w}'$ that stick out of the original hypercube $[0,1)^D$ and wrap them back in as if using periodic boundary conditions, we end up with the vector $\mathbf{w}$, whose elements (for $i = 1, \ldots, D$) are

$$w_i = w_i' \,\%\, 1 = w_i' - \lfloor w_i' \rfloor. \tag{3.16}$$

Here, $x \,\%\, y$ is the unique value in $[0, y)$ such that $x - (x \,\%\, y)$ is an integer multiple of $y > 0$ (commonly referred to as the remainder), and $\lfloor x \rfloor$ is the largest integer not exceeding x (commonly called the floor).

After being shifted and wrapped, every individual sequence in RqMC should still be evenly distributed in the original hypercube, and will generate a single estimate for the quantity in question. Because the estimates depend on the random shifts, they are themselves random variables and may be combined in the customary ways to obtain not only a sample mean, but also its standard error. This adds another parameter into the calculation: the number of shifted sequences N_S must be chosen carefully to strike a balance between ensuring a sufficient sample size for valid estimation of the error, and conserving the smooth results provided by low-discrepancy sequences. An example of this trade-off is shown in Appendix 3.C.

In addition to drawing uniform samples from a hypercube, we require the ability to sample quasi-randomly from the multivariate Gaussian pdf in Eq. (3.14). To do this, we first generate a point in the D-dimensional hypercube $[0,1)^D$ using a D-dimensional Sobol sequence and shift this point to randomize it. Then each of the D elements is treated as a cumulative distribution function (cdf) value (i.e. as a probability) and inverted to produce a one-dimensional standard Gaussian sample.[‡] Finally, these are combined to generate the appropriate multi-dimensional sample. In full, the randomized low-discrepancy multivariate Gaussian sampling routine SAMPLE_Q(b) may be implemented as in Alg. 1, where $b = 1, \ldots, B$ determines the distribution π^b.

3.2.3 Simultaneous perturbation stochastic approximation

The goal of the simultaneous perturbation stochastic approximation (SPSA) algorithm is to find the global minimum of a loss function $\mathcal{L}(\boldsymbol{\Theta})$ by varying the G-dimensional parameter vector $\boldsymbol{\Theta}$.[19] Although similar to the finite difference stochastic approximation (FDSA) algorithm, in which each

[‡]Some programming languages and libraries include a function called ERFCINV, which computes the inverse of the complementary error function and may be used for Gaussian cdf inversion: $\Phi^{-1}(\varphi) = -\sqrt{2}\,\mathrm{erfcinv}(2\varphi)$.

Algorithm 1 Multivariate Gaussian sampling for randomized quasi-Monte Carlo.

Require: $N_\mathrm{S} \geq 1$
Require: mean vector $\mathbf{d}^b$
Require: covariance matrix $\mathbf{\Sigma}^b$ whose inverse $(\mathbf{\Sigma}^b)^{-1}$ has eigenvalues $\mathbf{\Lambda}^b$ and eigenvectors $\mathbf{S}^b$
Require: shift vectors $\mathbf{v}_s^b$
Require: NEXT(b) returns the next point in the D-dimensional LDS labelled by b
Require: CDFINV(w) performs Gaussian cdf inversion

function SAMPLE_Q(b)
  $\mathbf{u} \leftarrow$ NEXT(b)

  for $s \leftarrow 1 : N_\mathrm{S}$ **do**
    for $i \leftarrow 1 : D$ **do**
      $w_{si} \leftarrow (u_i + v_{si}^b) \,\%\, 1$
      $y_{si} \leftarrow \mathrm{CDFINV}(w_{si}) \,/\sqrt{\Lambda_i^b}$
    end for
    $\mathbf{x}_s \leftarrow \mathbf{S}^b \mathbf{y}_s + \mathbf{d}^b$
  end for

  RETURN($\mathbf{x}$)
end function

Algorithm 2 One iteration of SPSA.

Require: FLIP() returns ± 1, each with probability $1/2$

function SPSA(LOSS, k, $\boldsymbol{\Theta}$)
  for $i \leftarrow 1 : G$ **do** ▷ Perturb parameters.
    $\Delta_i \leftarrow \text{FLIP}()$
    $\Theta_i^+ \leftarrow \Theta_i + c_k \Delta_i$
    $\Theta_i^- \leftarrow \Theta_i - c_k \Delta_i$
  end for

  $\ell^+, \sigma^+ \leftarrow \text{LOSS}(\boldsymbol{\Theta}^+)$ ▷ Approximate gradient.
  $\ell^-, \sigma^- \leftarrow \text{LOSS}(\boldsymbol{\Theta}^-)$
  $\mathbf{g} \leftarrow \frac{\ell^+ - \ell^-}{2c_k} \boldsymbol{\Delta}$

  $\boldsymbol{\Theta}' \leftarrow \boldsymbol{\Theta} - a_k \mathbf{g}$ ▷ Update parameters.
  RETURN($\boldsymbol{\Theta}'$)
end function

dimension requires a finite difference evaluation at every step, SPSA perturbs all G dimensions simultaneously, accelerating convergence. The "SA" part of the name refers to the approximation made in computing the gradient of $\mathcal{L}$ using samples of

$$\ell(\boldsymbol{\Theta}) = \mathcal{L}(\boldsymbol{\Theta}) + \text{noise}, \tag{3.17}$$

in contrast to many other methods that require exact gradients.[20]

This iterative algorithm is straightforward to describe and implement. The function SPSA(LOSS, k, $\boldsymbol{\Theta}$) in Alg. 2 is repeatedly called with increasing integer $k \geq 1$ until convergence or a maximum iteration threshold N_{SPSA}. The argument LOSS is a function that provides an estimate of the loss, together with the standard error of the estimate (which is not used here, but will be necessary elsewhere). Implicit in the presented algorithm is our choice to use a symmetric Bernoulli distribution to generate the perturbation vector $\boldsymbol{\Delta}$.

Besides N_{SPSA}, the algorithm requires five parameters to be chosen: A, a, c, α, and γ. These are used to compute the gain sequences

$$a_k = \frac{a}{(A+k)^\alpha} \tag{3.18a}$$

and

$$c_k = \frac{c}{k^\gamma}. \tag{3.18b}$$

For α and γ, we use the recommended values $\alpha = 0.602$ and $\gamma = 0.101$, and we set A to 10 % of N_{SPSA} and c to be equal to the standard error of $\ell(\boldsymbol{\Theta})$ before the initial iteration.[20] For a, we resort to manual adjustment; the optimization will be inefficient if it is too small, and erratic if it is too large.

3.3 Methods

In this section, we present our proposed improvements to GMD sampling for purposes of estimating partition functions of vibronic Hamiltonians using PIMC. In Sec. 3.3.1 and Sec. 3.3.2, we describe how to make the sampling more efficient. In Sec. 3.3.3 and Sec. 3.3.4, we show how to optimize the GMD parameters to better describe the integrand.

3.3.1 Deterministic component selection

The intuitive way to sample from a mixture distribution of the form in Eq. (3.13) is to first choose a component π^b using an RNG and then sample from that component. Repeating this N_{MC} times results in the point set $\{\mathbf{R}_i\}_{i=1}^{N_{\text{MC}}}$, where it is assumed that the points are not correlated. If the components are chosen in accordance with the weights w^b, these points can be used in the usual way[98] to find the sample mean

$$\bar{f} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} f(\mathbf{R}_i), \tag{3.19}$$

sample variance

$$s_f^2 = \frac{1}{N_{\text{MC}} - 1} \sum_{i=1}^{N_{\text{MC}}} (f(\mathbf{R}_i) - \bar{f})^2, \tag{3.20}$$

and sample error of the mean (also known as standard error)

$$s_{\bar{f}} = \sqrt{\frac{s_f^2}{N_{\text{MC}}}} = \sqrt{\frac{1}{N_{\text{MC}}(N_{\text{MC}} - 1)} \sum_{i=1}^{N_{\text{MC}}} (f(\mathbf{R}_i) - \bar{f})^2}. \tag{3.21}$$

Although quite straightforward (see Alg. 3 and Fig. 3.1), this approach struggles with small component weights. For example, consider $N_{\text{MC}} = 10^6$ and $w^{b^\star} = 10^{-6}$ for some $b^\star$. From a simple

Algorithm 3 GMD sampling with stochastic component selection.

Require: $N_{\mathrm{MC}} \geq 1$
Require: CHOOSE($\mathbf{w}$) returns index b with probability w^b
Require: SAMPLE(b) returns a sample from π^b

for $i \leftarrow 1 : N_{\mathrm{MC}}$ **do**
  $b \leftarrow \textsc{choose}(\mathbf{w})$ ▷ Component selection.
  $\mathbf{R}_i \leftarrow \textsc{sample}(b)$ ▷ Component sampling.
  $f_i \leftarrow f(\mathbf{R}_i)$
end for

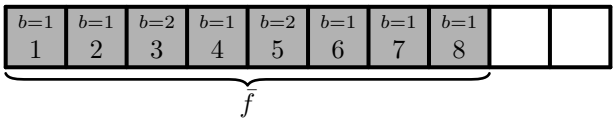

Figure 3.1: Example of $N_{\mathrm{MC}} = 8$ samples f_i drawn using Alg. 3, with GMD components stochastically chosen in the order 1, 1, 2, 1, 2, 1, 1, 1. The numbers along the bottom indicate the order in which the samples were obtained, and the b labels show the GMD components from which they were sampled. The brace shows the values used to compute the average $\bar{f}$.

binomial distribution analysis, we expect to find exactly one sample for this component only

$$\binom{10^6}{1}(10^{-6})^1(1-10^{-6})^{10^6-1} \approx 37\,\% \tag{3.22}$$

of the time, with it being under-represented (zero samples) 37 % of the time, and over-represented (two or more samples) the remaining 26 %.[‡] In any event, it is unlikely that just a handful of samples is sufficient to properly explore the component.

The formulas in Eqs. (3.19)–(3.21) continue to hold so long as the effective distribution of components is consistent with the weights, even if the components are chosen in some other way, such as by a deterministic procedure. This suggests that we could sample from each component exactly $N^b = w^b N_{\mathrm{MC}}$ times, in any order. The obvious flaw with this approach is that N^b must be an integer, so a very small w^b implies a very large N_{MC}, which is inconvenient. In the remainder of this section, we demonstrate how to adjust the standard formulas to compensate for point sets where the distribution of components used for sampling does not match their weights, and provide a strategy for choosing the components during sampling.

Suppose we sample from each component π^b an arbitrary number of times N^b, obtaining the point

[‡]The asymptotic distribution as $N_{\mathrm{MC}} \to \infty$ and $w^b N_{\mathrm{MC}} = 1$ is a Poisson distribution with mean 1, resulting in the following probabilities: $1/e$ for zero samples, $1/e$ for one sample, and $1 - 2/e$ for two or more samples.

set $\bigcup_{b=1}^{B}\{\mathbf{R}_i^b\}_{i=1}^{N^b}$. From these samples, we may compute a collection of sample averages

$$\bar{f}^b = \frac{1}{N^b}\sum_{i=1}^{N^b} f(\mathbf{R}_i^b). \tag{3.23}$$

The combined quantity

$$\bar{f} = \sum_{b=1}^{B} w^b \bar{f}^b \tag{3.24}$$

is an unbiased estimator for $\langle f\rangle_\pi$:

$$\left\langle \bar{f}\right\rangle = \sum_{b=1}^{B} w^b \frac{1}{N^b}\sum_{i=1}^{N^b}\langle f\rangle_{\pi^b} = \sum_{b=1}^{B} w^b \int \mathrm{d}\mathbf{R}\ \pi^b(\mathbf{R}) f(\mathbf{R}) = \int \mathrm{d}\mathbf{R}\ \pi(\mathbf{R}) f(\mathbf{R}) = \langle f\rangle_\pi\,. \tag{3.25}$$

Each $\bar{f}^b$ is itself an independent random variable with population variance $\sigma^2_{\bar{f}^b}$. As shown in Appendix 3.D, the overall population variance of $\bar{f}$ is then given by

$$\sigma^2_{\bar{f}} = \sum_{b=1}^{B}(w^b)^2\sigma^2_{\bar{f}^b}. \tag{3.26}$$

For the variance $\sigma^2_{\bar{f}^b}$, we have the usual unbiased estimate

$$s^2_{\bar{f}^b} = \frac{s^2_{f^b}}{N^b} = \frac{1}{N^b(N^b-1)}\sum_{i=1}^{N^b}(f(\mathbf{R}_i^b)-\bar{f}^b)^2, \tag{3.27}$$

from which we get

$$s^2_{\bar{f}} = \sum_{b=1}^{B}(w^b)^2 s^2_{\bar{f}^b}. \tag{3.28}$$

Thus, the standard error may be estimated by

$$s_{\bar{f}} = \sqrt{\sum_{b=1}^{B}(w^b)^2 s^2_{\bar{f}^b}} = \sqrt{\sum_{b=1}^{B}\frac{(w^b)^2 s^2_{f^b}}{N^b}}. \tag{3.29}$$

Although it is possible to fix N^b ahead of time, it is not necessary to do so. Instead, we may choose each component during sampling in a way that attempts to optimally reduce the overall statistical error. We observe that after sampling N^b points from each component, an additional sample from

the component labelled by $b^\star$ changes the sample variance of the mean $s^2_{\bar{f}}$ to

$$(s'_{\bar{f}})^2 = \frac{(w^{b^\star})^2 (s'_{fb^\star})^2}{N^{b^\star}+1} + \sum_{\substack{b=1 \\ (b\neq b^\star)}}^{B} \frac{(w^b)^2 s^2_{fb}}{N^b}. \tag{3.30}$$

Ideally, the estimate of the component variance will not be greatly affected by a single sample, so we make the simplifying assumption that

$$(s'_{fb^\star})^2 \approx s^2_{fb^\star}. \tag{3.31}$$

This allows us to write

$$(s'_{\bar{f}})^2 \approx s^2_{\bar{f}} - (w^{b^\star})^2 s^2_{fb^\star} \left[\frac{1}{N^{b^\star}} - \frac{1}{N^{b^\star}+1}\right], \tag{3.32}$$

from which it follows that

$$(s'_{\bar{f}})^2 \approx s^2_{\bar{f}} - \frac{(w^{b^\star})^2 s^2_{\bar{f}b^\star}}{N^{b^\star}+1}, \tag{3.33}$$

and we see that to lower the standard error most quickly, it is beneficial to choose the component with the largest

$$\delta^b = \frac{(w^b)^2 s^2_{\bar{f}b}}{N^b+1}. \tag{3.34}$$

This criterion makes intuitive sense, as it targets components that have larger weights (because they are more important to sample well), larger errors of the mean (because they have a lot of remaining uncertainty), and fewer samples (because they have not been explored as thoroughly). Since the computation of δ^b requires an existing estimate of the variance in that component, it is necessary to bootstrap this method by taking N_{boot} points from each component.

When implementing the above prescription, one must be careful to avoid the common formula in Eq. (3.20) for estimating variance, since it scales linearly with the number of samples. Using it after sampling each point leads to quadratic scaling of the total sampling algorithm. To avoid this undesirable behavior, one may use an efficient "online" variance update scheme. For example, the scheme of Youngs and Cramer for the L-point sample variance $s^2_f(L)$ is

$$s^2_f(L) = \frac{U_f(L)}{L-1}, \tag{3.35a}$$

where

$$T_f(L) = T_f(L-1) + f(\mathbf{R}_L), \tag{3.35b}$$

$$U_f(L) = U_f(L-1) + \frac{(Lf(\mathbf{R}_L) - T_f(L))^2}{L(L-1)}, \tag{3.35c}$$

subject to $T_f(0) = U_f(1) = 0$.[99,100] These updates can be evaluated in constant time at each step of the calculation. The scheme is made more effective with a conditioning step where the data are shifted by the mean:[100]

$$f(\mathbf{R}_L) \to \Delta f_L = f(\mathbf{R}_L) - \bar{f}. \tag{3.36}$$

Although the final sample mean will not be known until the end of the calculation, uniformly shifting the data has no impact on its variance. Thus, the precise value of the shift is not crucial, and the estimates $\bar{f}^b_{\text{boot}}$ obtained from the bootstrap points may be used. The overall algorithm for deterministic component selection is shown in Alg. 4, with an example in Fig. 3.2.

3.3.2 Randomized quasi-Monte Carlo

In order to incorporate quasi-random numbers into the PIMC study of vibronic Hamiltonians, the essential change is the substitution of an RNG by N_{S} randomly shifted Sobol sequences for each GMD component. Instead of drawing N_{MC} pseudo-random points from an RNG, one then obtains N_{qMC} quasi-random ones from the low-discrepancy sequences. The main consequence of this change can be seen by comparing the examples in Fig. 3.2 and Fig. 3.3: the component averages $\bar{f}^b$ are not computed from N^b values, but from N_{S} of them. This does not affect the prior formulas in Eq. (3.24) and Eq. (3.28) for combining these averages into the overall mean $\bar{f}$ and variance $s^2_{\bar{f}}$.

However, the component selection criterion in Eq. (3.34) and the online variance update formula in Eq. (3.35) are not directly applicable when RqMC is in use. They are based on the premise that each additional sample increases the number of entities being considered. Since the number of statistically independent values used for computing each component mean $\bar{f}^b$ is now always fixed to be N_{S}, the reasoning behind these equations is no longer valid. Conveniently, the online variance update is not necessary anymore, as the conventional variance formula becomes independent of the number of samples and can therefore be computed in constant time.

Component selection, on the other hand, becomes highly non-trivial, since Eq. (3.33) changes to

$$(s'_{\bar{f}})^2 = s^2_{\bar{f}} - \frac{(w^{b\star})^2}{N_{\text{S}}} \left(s^2_{f^{b\star}} - (s'_{f^{b\star}})^2 \right), \tag{3.37}$$

which requires us to be able to predict the change in variance for a component when a single

Algorithm 4 GMD sampling with deterministic component selection.

Require: $N_{\text{boot}} \geq 2$
Require: $N_{\text{MC}} \geq BN_{\text{boot}}$
Require: SAMPLE(b) returns a sample from π^b

for $b \leftarrow 1 : B$ **do** ▷ Bootstrap.
  for $i \leftarrow 1 : N_{\text{boot}}$ **do**
    $\mathbf{R}_i^b \leftarrow \text{SAMPLE}(b)$
    $f_i^b \leftarrow f(\mathbf{R}_i^b)$
  end for

  $N^b \leftarrow N_{\text{boot}}$
  $\bar{f}_{\text{boot}}^b \leftarrow \frac{1}{N^b} \sum_{i=1}^{N^b} f_i^b$

  $T^b \leftarrow 0$
  $U^b \leftarrow \sum_{i=1}^{N^b} (f_i^b - \bar{f}_{\text{boot}}^b)^2$
  $s_{\bar{f}^b}^2 \leftarrow \frac{U^b}{N^b(N^b-1)}$
end for

for $i \leftarrow 1 : (N_{\text{MC}} - BN_{\text{boot}})$ **do** ▷ Sampling.
  for $b \leftarrow 1 : B$ **do** ▷ Component selection.
    $\delta^b \leftarrow \frac{(w^b)^2 s_{\bar{f}^b}^2}{N^b+1}$
  end for
  $b \leftarrow \text{ARGMAX}(\boldsymbol{\delta})$
  $N^b \leftarrow N^b + 1$

  $\mathbf{R}_{N^b}^b \leftarrow \text{SAMPLE}(b)$ ▷ Component sampling.
  $f_{N^b}^b \leftarrow f(\mathbf{R}_{N^b}^b)$

  $\Delta f \leftarrow f_{N^b}^b - \bar{f}_{\text{boot}}^b$ ▷ Online variance update.
  $T^b \leftarrow T^b + \Delta f$
  $U^b \leftarrow U^b + \frac{(N^b \Delta f - T^b)^2}{N^b(N^b-1)}$
  $s_{\bar{f}^b}^2 \leftarrow \frac{U^b}{N^b(N^b-1)}$
end for

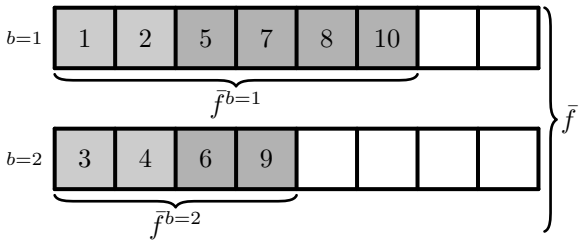

Figure 3.2: Example of $N_{\mathrm{MC}} = 10$ samples f_i^b drawn using Alg. 4, with GMD components deterministically chosen in the order 1, 1, 2, 2 for the bootstrap ($N_{\mathrm{boot}} = 2$), followed by 1, 2, 1, 1, 2, 1. The component-wise counts are $N^{b=1} = 6$ and $N^{b=2} = 4$. The numbers indicate the order in which the samples were obtained, and the b labels on the left identify the GMD components. The braces show the way in which the samples are combined: those for each GMD component are first grouped into $\bar{f}^b$, and these averages are subsequently used to form $\bar{f}$.

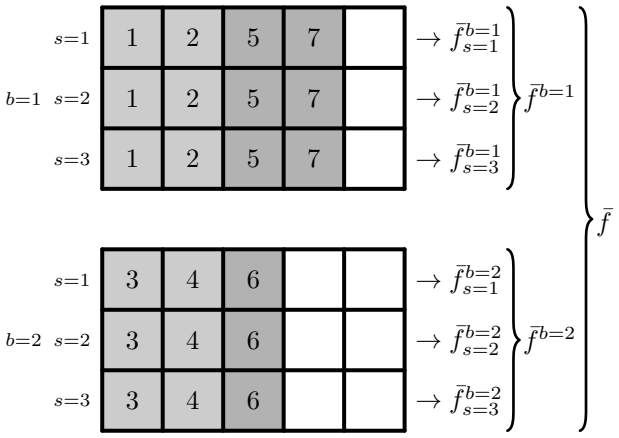

Figure 3.3: Example of $N_{\mathrm{MC}} = 21$ samples f_{is}^b across $N_{\mathrm{S}} = 3$ sequences drawn using Alg. 5 from $N_{\mathrm{qMC}} = 7$ low-discrepancy points, with GMD components deterministically chosen in the order 1, 1, 2, 2 for the bootstrap ($N_{\mathrm{boot}} = 2$), followed by 1, 2, 1. The component-wise counts are $N^{b=1} = 4$ and $N^{b=2} = 3$. The numbers indicate the order in which the samples were obtained, and the b and s labels on the left identify the GMD components and randomized sequences. The arrows and braces show the way in which the samples are combined: each sequence is first folded into the pseudo-random quantity $\bar{f}_s^b$, then these are averaged for each GMD component to find $\bar{f}^b$, and finally the GMD component averages are combined into $\bar{f}$.

quasi-random sample is added. To side-step this issue altogether, we argue that the intuition behind Eq. (3.34) is still legitimate and we continue to choose the component with the largest

$$\delta^b = \frac{(w^b)^2 s^2_{\bar{f}^b}}{N^b + 1}, \tag{3.38}$$

where N^b is the total number of quasi-random points drawn so far for that component. The overall procedure for RqMC with deterministic component selection is detailed in Alg. 5, with an example in Fig. 3.3. Note that N_{qMC} is the number of quasi-random points obtained in total from the LDSs; for a fair comparison with plain Monte Carlo methods, we also define $N_{\mathrm{MC}} = N_{\mathrm{S}} N_{\mathrm{qMC}}$, which is the total number of times that $f(\mathbf{R})$ is evaluated.

3.3.3 Parameter optimization

A major concern with the sampling scheme presented in Ref. [7] is the need to choose a sampling GMD π that closely approximates the true path density g. Neglecting to do so results in catastrophic failure when estimating $\langle f \rangle_\pi$, as the sampled points are chosen to lie in irrelevant locations that contribute nothing of substance to the integral in Eq. (3.5). Thus, we need a way to refine π, making it more similar to g, and this requires a quantitative measurement of the difference between them.

A common metric to determine the similarity between two distributions is the relative entropy, also known as the Kullback–Leibler divergence.[101] It is not symmetric in the distributions, so given pdfs $g(\mathbf{R})/Z$ and $\pi(\mathbf{R})$, there are two possible quantities:

$$D(\pi || g/Z) = \int \mathrm{d}\mathbf{R}\ \pi(\mathbf{R}) \log \frac{\pi(\mathbf{R}) Z}{g(\mathbf{R})} = \log \langle f \rangle_\pi - \langle \log f \rangle_\pi \tag{3.39a}$$

and

$$D(g/Z || \pi) = \int \mathrm{d}\mathbf{R}\ \frac{g(\mathbf{R})}{Z} \log \frac{g(\mathbf{R})}{\pi(\mathbf{R}) Z} = \log \left\langle \frac{1}{f} \right\rangle_{g/Z} - \left\langle \log \frac{1}{f} \right\rangle_{g/Z}. \tag{3.39b}$$

For our purposes, these expressions are both ill-defined, because $g(\mathbf{R})$ may differ in sign from Z (see Appendix 3.E for an example), and $g(\mathbf{R})/Z$ is therefore not necessarily a pdf. This may be remedied by using

$$\frac{|g(\mathbf{R})|}{Z_{||}} = \frac{|g(\mathbf{R})|}{\int \mathrm{d}\mathbf{R}' \, |g(\mathbf{R}')|} \tag{3.40}$$

instead; if the negative regions of g are small in magnitude, this will be a good approximation to g/Z. However, with no way to sample directly from $|g|/Z_{||}$, it is not clear how to efficiently evaluate $D(|g|/Z_{||} \,||\, \pi)$.

Algorithm 5 Quasi-random GMD sampling.

Require: $N_{\text{boot}} \geq 1$
Require: $N_{\text{S}} \geq 2$
Require: $N_{\text{qMC}} \geq BN_{\text{boot}}$
Require: SAMPLE_Q(b) is implemented as in Alg. 1 in Sec. 3.2.2

for $b \leftarrow 1 : B$ **do** ▷ Bootstrap.
  for $i \leftarrow 1 : N_{\text{boot}}$ **do**
    $\mathbf{R}_i^b \leftarrow$ SAMPLE_Q(b)
    for $s \leftarrow 1 : N_{\text{S}}$ **do**
      $f_{is}^b \leftarrow f(\mathbf{R}_{is}^b)$
    end for
  end for
  $N^b \leftarrow N_{\text{boot}}$

  for $s \leftarrow 1 : N_{\text{S}}$ **do**
    $T_s^b \leftarrow \sum_{i=1}^{N^b} f_{is}^b$
    $\bar{f}_s^b \leftarrow \frac{T_s^b}{N^b}$
  end for

  $\bar{f}^b \leftarrow \frac{1}{N_{\text{S}}} \sum_{s=1}^{N_{\text{S}}} \bar{f}_s^b$
  $s^2_{\bar{f}^b} \leftarrow \frac{1}{N_{\text{S}}(N_{\text{S}}-1)} \sum_{s=1}^{N_{\text{S}}} (\bar{f}_s^b - \bar{f}^b)^2$
end for

for $i \leftarrow 1 : (N_{\text{qMC}} - BN_{\text{boot}})$ **do** ▷ Sampling.
  for $b \leftarrow 1 : B$ **do** ▷ Component selection.
    $\delta^b \leftarrow \frac{(w^b)^2 s^2_{\bar{f}^b}}{N^b+1}$
  end for
  $b \leftarrow$ ARGMAX($\boldsymbol{\delta}$)
  $N^b \leftarrow N^b + 1$

  $\mathbf{R}_{N^b}^b \leftarrow$ SAMPLE_Q(b) ▷ Component sampling.
  for $s \leftarrow 1 : N_{\text{S}}$ **do**
    $f_{N^b s}^b \leftarrow f(\mathbf{R}_{N^b s}^b)$
    $T_s^b \leftarrow T_s^b + f_{N^b s}^b$
    $\bar{f}_s^b \leftarrow \frac{T_s^b}{N^b}$
  end for

  $\bar{f}^b \leftarrow \frac{1}{N_{\text{S}}} \sum_{s=1}^{N_{\text{S}}} \bar{f}_s^b$ ▷ Variance update.
  $s^2_{\bar{f}^b} \leftarrow \frac{1}{N_{\text{S}}(N_{\text{S}}-1)} \sum_{s=1}^{N_{\text{S}}} (\bar{f}_s^b - \bar{f}^b)^2$
end for

Thus, the best measure of quality that we can get is

$$D(\pi \,||\, |g|/Z_{||}) = \int \mathrm{d}\mathbf{R}\; \pi(\mathbf{R}) \log \frac{\pi(\mathbf{R}) Z_{||}}{|g(\mathbf{R})|} = \log \langle |f| \rangle_\pi - \langle \log |f| \rangle_\pi \,. \tag{3.41}$$

We can then minimize this quantity by modifying π using any of a number of optimization techniques, yielding

$$\underset{\mathbf{E}}{\operatorname{argmin}}\, D(\pi_{\mathbf{E}} \,||\, |g|/Z_{||}), \tag{3.42}$$

where the sampling distribution $\pi_{\mathbf{E}}$ is parameterized by the tensor $\mathbf{E}$. In most cases, attempting this would not be fruitful: if first we need to optimize π in order to be able to compute $\langle f \rangle_\pi$, then we can't use Eq. (3.41) directly. However, for cases where π is already a reasonably good approximation of g/Z, this could be used to improve it further.

Though each $\pi_{\mathbf{E}}$ may be any GMD, as specified by its weights, mean vectors, and covariance matrices, we choose to construct it from the diagonal Hamiltonian $\hat{h}_{\mathbf{E}}$ with the elements

$$\hat{h}^b_{\mathbf{E}} = \sum_{j=1}^{N} \frac{\omega_j}{2} (\hat{p}_j^2 + \hat{q}_j^2) + E^{(0)b} + \sum_{j=1}^{N} E_j^{(1)b} \hat{q}_j, \tag{3.43}$$

where $b = 1, \ldots, B$, and B must be large enough that $\pi_{\mathbf{E}}$ is sufficiently flexible to describe g, but not so large that the optimization becomes too costly. This is a natural form for this problem and allows us to easily use the same frequencies ω_j as in the full Hamiltonian $\hat{H}$ in Eq. (3.1). Additionally, because $\pi_{\mathbf{E}}$ may be generated from $\hat{h}_{\mathbf{E}}$ for any number of beads P, the same optimized parameters may be used for a range of P values.

The primary constituent of the optimization loss function $\mathcal{L}(\mathbf{E})$ should be the relative entropy $D(\pi_{\mathbf{E}} \,||\, |g|/Z_{||})$, but we can avoid some pitfalls in the optimization by adding extra terms. One concern is with components $\pi^b_{\mathbf{E}}$ and $\pi^{b'}_{\mathbf{E}}$ that come too close to each other (that is, whose mean vectors $\mathbf{d}^b_{\mathbf{E}}$ and $\mathbf{d}^{b'}_{\mathbf{E}}$ become nearly the same). When this happens, they are likely to follow the same trajectory during optimization, which is wasteful. To encourage the components to steer clear of one another, we introduce a penalty term proportional to

$$\frac{2}{B(B-1)} \sum_{b=1}^{B-1} \sum_{b'=b+1}^{B} e^{-\frac{\left|\mathbf{d}^b_{\mathbf{E}} - \mathbf{d}^{b'}_{\mathbf{E}}\right|^2}{2\varepsilon^2}}, \tag{3.44}$$

where ε determines the stiffness of the repulsion.

Another potential issue is that some components $\pi^b_{\mathbf{E}}$ can fall out of favor and stop being considered important due to their low weights $w^b_{\mathbf{E}}$. When this occurs, they cease being purposefully optimized

Algorithm 6 GMD parameter optimization.

Require: initial parameter tensor $\mathbf{E}$
Require: $N_{\text{SPSA}} \geq 1$
Require: LOSS($\mathbf{E}$) estimates the loss function $\mathcal{L}$ in Eq. (3.46), returning a value and standard error
Require: SPSA(LOSS, k, $\mathbf{E}$) is implemented as in Alg. 2 in Sec. 3.2.3

$\mathbf{E}_0 \leftarrow \mathbf{E}$

for $k \leftarrow 1 : N_{\text{SPSA}}$ **do**
  $\mathbf{E}_k \leftarrow$ SPSA(LOSS, k, $\mathbf{E}_{k-1}$)
end for

$\mathbf{E} \leftarrow \mathbf{E}_{N_{\text{SPSA}}}$

and begin to aimlessly wander. While this does allow them to explore highly unfavorable regions and therefore possibly cross barriers to find deeper minima, in our experience this is futile and should be suppressed. To support those components which might find themselves at risk of becoming irrelevant, we penalize mixtures that have a low Shannon entropy[102] of the weights using a term proportional to

$$1 + \frac{\sum_{b=1}^{B} w_{\mathbf{E}}^{b} \log w_{\mathbf{E}}^{b}}{\log B}. \tag{3.45}$$

This vanishes when all the weights are equal and grows to unity when a single component dominates the mixture.

The complete loss function that we employ is

$$\mathcal{L}(\mathbf{E}) = D(\pi_{\mathbf{E}} \,||\, |g|/Z_{||}) + C_1 \frac{2}{B(B-1)} \sum_{b=1}^{B-1} \sum_{b'=b+1}^{B} e^{-\frac{\left|\mathbf{d}_{\mathbf{E}}^{b} - \mathbf{d}_{\mathbf{E}}^{b'}\right|^2}{2\varepsilon^2}} + C_2 \left(1 + \frac{\sum_{b=1}^{B} w_{\mathbf{E}}^{b} \log w_{\mathbf{E}}^{b}}{\log B}\right), \tag{3.46}$$

where ε, C_1, and C_2 are tunable parameters, and only the first term has an associated statistical error. To optimize the parameter tensor $\mathbf{E}$, we use SPSA, as it is both efficient and very simple to implement.[19,20] The overall optimization algorithm is shown in Alg. 6.

Estimation of the relative entropy from Eq. (3.41) is performed using the RqMC method with deterministic component selection outlined in Sec. 3.3.2.[‡] When a single RqMC calculation is used to compute both terms of Eq. (3.41), evaluating the standard error of the mean is not as simple

[‡]When the RqMC algorithm selects components, it tries to reduce the overall error in the estimate of the partition function. Ideally, it would be modified to attempt to reduce the error in the estimate of Eq. (3.41) instead, but we find that this is not necessary.

as for the partition function, because of the correlations between $\overline{|f|}$ and $\overline{\log|f|}$. We may use the jackknife technique,[98] but since there are several layers to RqMC, we must be careful about what exactly our statistical sample means are. Recall that each of $\overline{|f|}$ and $\overline{\log|f|}$ is computed separately using the weights and component means, so a simple estimate of the sample mean is given by

$$D\left(\overline{|f^1|},\ldots,\overline{|f^B|},\overline{\log|f^1|},\ldots,\overline{\log|f^B|}\right) = \log\left[\sum_{b=1}^{B} w^b\overline{|f^b|}\right] - \sum_{b=1}^{B} w^b\overline{\log|f^b|}. \tag{3.47}$$

Because each component mean $\overline{|f^b|}$ and $\overline{\log|f^b|}$ is computed as a statistical average over N_S independent random variables, we may apply jackknife to this function as usual.

3.3.4 Parameter optimization with deformation

To get around the circularity in Sec. 3.3.3 that arises from needing π in order to determine π in the general case, we propose an iterative scheme. Recall that the function g is derived from the vibronic Hamiltonian $\hat{H}$ in Eq. (3.1) as described in Sec. 3.2.1. We define the parameterized function g_ν (with $0 \le \nu \le 1$) to be similar to g, but instead derived from the parameterized Hamiltonian $\hat{H}_\nu$, where all the terms except the pure harmonic oscillator are scaled by ν:

$$\hat{H}_\nu^{aa'} = \nu\hat{H}^{aa'} + (1-\nu)\delta_{aa'}\hat{h}_\mathrm{o}. \tag{3.48}$$

When $\nu = 1$, we recover the original Hamiltonian $\hat{H} = \hat{H}_{\nu=1}$, so $g = g_{\nu=1}$; when $\nu = 0$, all off-diagonal terms vanish and the Hamiltonian $\hat{H}_{\nu=0}$ is trivial, so $g_{\nu=0}$ is proportional to a GMD. This allows us to smoothly interpolate between easily solvable models and chemically interesting ones. Thus, we can use g_ν to guide the optimization of π towards g. For each g_ν, we introduce the normalized and sign-free distribution

$$\frac{|g_\nu(\mathbf{R})|}{Z_{\nu||}} = \frac{|g_\nu(\mathbf{R})|}{\int \mathrm{d}\mathbf{R}\,|g_\nu(\mathbf{R})|}. \tag{3.49}$$

In practice, we must choose a monotonically increasing sequence ν_n, where $\nu_1 > 0$ and $\nu_{N_\nu} = 1$, such that each step $\Delta\nu_n = \nu_n - \nu_{n-1}$ is small enough to obtain accurate estimates, yet large enough to make rapid progress toward the desired distribution. To select the ν_n, we use the function STEP_NU in Alg. 7, which is described in more detail in Appendix 3.F. At the final iteration $n = N_\nu$, the target distribution is $|g|/Z_{||}$ itself, so the optimized GMD should be adequate for evaluating $Z = \langle f\rangle_\pi$.

To initialize the coefficients $E^{(0)b}$ and $E_j^{(1)b}$ making up the tensor $\mathbf{E}$ in Eq. (3.43), we set them all

to zero; this leaves us with only the pure harmonic oscillator terms:

$$\hat{h}^b_{\mathbf{E}=\mathbf{0}} = \sum_{j=1}^{N} \frac{\omega_j}{2}(\hat{p}_j^2 + \hat{q}_j^2) = \hat{h}_\mathrm{o}. \tag{3.50}$$

Hence, the initial distribution $\pi_{\mathbf{E}=\mathbf{0}}$ resembles $g_{\nu=0}$, but potentially with a different number of components.

For each $n = 1, \ldots, N_\nu$, we perturb $\mathbf{E}$ using the optimization routine from Alg. 6 to minimize the loss function

$$\mathcal{L}_\nu(\mathbf{E}) = D(\pi_\mathbf{E} \,||\, |g_\nu|/Z_{\nu||}) + C_1 \frac{2}{B(B-1)} \sum_{b=1}^{B-1} \sum_{b'=b+1}^{B} e^{-\frac{\left|\mathbf{d}^b_\mathbf{E} - \mathbf{d}^{b'}_\mathbf{E}\right|^2}{2\varepsilon^2}} + C_2 \left(1 + \frac{\sum_{b=1}^{B} w^b_\mathbf{E} \log w^b_\mathbf{E}}{\log B}\right), \tag{3.51}$$

where the second and third terms are the same as in Eq. (3.46). We take the best of these perturbed distributions $\pi_\mathbf{E}$ and promote it to be the starting distribution for the next iteration. The overall algorithm for optimization with deformation can be seen in Alg. 8. Note that we use N_W independent "walkers" and take the result from the best performing one (in the sense of minimizing the loss function); because they are independent, they may be executed in parallel.

3.4 Results

In this section, we apply the methods from Sec. 3.3 to model systems. All the systems that we consider have two diabatic surfaces and two spatial degrees of freedom, with the Hamiltonian

$$\hat{H} = \begin{pmatrix} E^1 + \hat{h}_\mathrm{o} + \lambda\hat{q}_1 & 0 \\ 0 & E^2 + \hat{h}_\mathrm{o} - \lambda\hat{q}_1 \end{pmatrix} + \gamma \begin{pmatrix} 0 & \hat{q}_2 \\ \hat{q}_2 & 0 \end{pmatrix}, \tag{3.52}$$

which has six parameters. We use an inverse temperature of $\beta = 38.7$, which is approximately 300 K if the parameter values are interpreted in eV. Where applicable, we use $N_\mathrm{S} = 100$ to ensure sufficiently many randomized sequences for accurate error bars, and $N_\mathrm{boot} = 64$ to obtain reasonable initial estimates. The sole exception to this is during GMD parameter optimization, when quality estimates are not required, in which case we use $N_\mathrm{S} = 50$ and $N_\mathrm{boot} = 4$. The algorithms are implemented using VibronicToolkit.[103]

3.4.1 Deterministic component selection

We start with a contrived example that highlights the ability of the deterministic component selection algorithm described in Sec. 3.3.1 to reduce the impact of outliers. We use the model

Algorithm 7 Step size selection for ν.

Require: $\text{LOSS}_\nu(\mathbf{E})$ estimates the loss function $\mathcal{L}_\nu$ in Eq. (3.51), returning a value and standard error

function STEP_NU(ν_{prev}, $\Delta\nu_{\text{prev}}$, $\mathbf{E}$)
  $\Delta\nu_{\text{min}} \leftarrow 10^{-3}$
  if $\nu_{\text{prev}} > 1 - 10^{-1} - 10^{-2}$ **then**
    $\Delta\nu_{\text{max}} \leftarrow 10^{-2}$
  else
    $\Delta\nu_{\text{max}} \leftarrow 10^{-1}$
  end if

  $\Delta\nu \leftarrow \text{MIN}(\Delta\nu_{\text{max}}, 2\Delta\nu_{\text{prev}})$

  for $i \leftarrow 1 : 8$ **do**
    $\nu_{\text{new}} \leftarrow \text{MIN}(1, \nu_{\text{prev}} + \Delta\nu)$
    $\ell, \sigma \leftarrow \text{LOSS}_{\nu_{\text{new}}}(\mathbf{E})$

    if $\ell < 1$ **then**
      $\Delta\nu_{\text{min}} \leftarrow \Delta\nu$
    else if $\ell > 2$ **then**
      $\Delta\nu_{\text{max}} \leftarrow \Delta\nu$
    else
      BREAK
    end if

    $\Delta\nu \leftarrow (\Delta\nu_{\text{min}} + \Delta\nu_{\text{max}})/2$
  end for

  RETURN($\Delta\nu$)
end function

Algorithm 8 GMD parameter optimization with deformation.

Require: $N_{\text{SPSA}} \geq 1$
Require: $N_{\text{W}} \geq 1$
Require: $\text{LOSS}_\nu(\mathbf{E})$ estimates the loss function $\mathcal{L}_\nu$ in Eq. (3.51), returning a value and standard error
Require: $\text{SPSA}(\text{LOSS}, k, \mathbf{E})$ is implemented as in Alg. 2 in Sec. 3.2.3
Require: $\text{STEP\_NU}(\nu_{\text{prev}}, \Delta\nu_{\text{prev}}, \mathbf{E})$ is implemented as in Alg. 7

for $b \leftarrow 1 : B$ **do** ▷ Initialization.
  $E^{(0)b} \leftarrow 0$
  for $j \leftarrow 1 : N$ **do**
    $E_j^{(1)b} \leftarrow 0$
  end for
end for

$\nu \leftarrow 0$
$\Delta\nu \leftarrow 10^{-1}$

while $\nu < 1$ **do** ▷ Iteration.
  $\Delta\nu \leftarrow \text{STEP\_NU}(\nu, \Delta\nu, \mathbf{E})$ ▷ ν selection.
  $\nu \leftarrow \text{MIN}(1, \nu + \Delta\nu)$

  for $i \leftarrow 1 : N_{\text{W}}$ **do** ▷ Parameter update.
    $\mathbf{E}_{i0} \leftarrow \mathbf{E}$
    for $k \leftarrow 1 : N_{\text{SPSA}}$ **do**
      $\mathbf{E}_{ik} \leftarrow \text{SPSA}(\text{LOSS}_\nu, k, \mathbf{E}_{i,k-1})$
      $\ell_{ik}, \sigma_{ik} \leftarrow \text{LOSS}_\nu(\mathbf{E}_{ik})$
    end for
  end for

  $i, k \leftarrow \text{ARGMIN}(\boldsymbol{\ell} + \boldsymbol{\sigma})$
  $\mathbf{E} \leftarrow \mathbf{E}_{ik}$
end while

Table 3.1: Parameters of the model Hamiltonian in Eq. (3.52) for the modified Displaced γ_2 system.

Parameter	Value
E^1	0.1
E^2	0.275

Parameter	Value
ω_1	0.02
ω_2	0.04

Parameter	Value
λ	0.075
γ	0.05

Hamiltonian Eq. (3.52) with the parameters given in Tab. 3.1, which are similar to those of the Displaced model of Ref. [7] with the γ_2 parameter.

The coupling between the surfaces is not very strong, so we expect the basic sampling approach to work well, and we use the diagonal portion of $\hat{H}$ to construct the two-component sampling distribution π_0. At a temperature of $\beta = 38.7$, the component weights are $w^{b=1} \approx 0.998\,86$ and $w^{b=2} \approx 0.001\,14$. Thus, out of every 1000 samples, approximately one should be drawn from the second component.

What makes this example particularly artificial is that we force the second point drawn from the second component to be an extreme outlier. Recall from Ref. [7] that each path is sampled from the component π_0^b using uncoupled coordinates $y_{j\lambda}$ (where $j = 1, \ldots, N$ and $\lambda = 1, \ldots, P$), each with a standard deviation $\sigma_{j\lambda}^b$. For the outlier point, we choose the coordinates

$$y_{j\lambda} = \begin{cases} 6\sigma_{j\lambda}^{b=2} & \text{if } j = 2 \text{ and } \lambda = 1 \\ 0.01\sigma_{j\lambda}^{b=2} & \text{otherwise,} \end{cases} \tag{3.53}$$

which place the centroid mode of the path quite far from its mean in the second spatial coordinate, as shown in Fig. 3.4. Since the cdf of a univariate Gaussian is (with erf being the error function)

$$\Phi_\sigma(x) = \frac{1}{2}\left(1 + \operatorname{erf}\frac{x}{\sqrt{2\sigma^2}}\right), \tag{3.54}$$

the total probability of either $x < -n\sigma$ or $x > n\sigma$ is

$$1 - \operatorname{erf}\frac{n}{\sqrt{2}}. \tag{3.55}$$

Hence, when $P = 16$, the probability of sampling a point that's at least as unlikely in all directions is

$$\left(1 - \operatorname{erf}\frac{6}{\sqrt{2}}\right)\left(1 - \operatorname{erf}\frac{0.01}{\sqrt{2}}\right)^{31} \approx 1.539\,276 \times 10^{-9}. \tag{3.56}$$

Admittedly, this is not a very probable event. However, when taking huge numbers of samples over many calculations, very unlikely things are bound to happen occasionally.

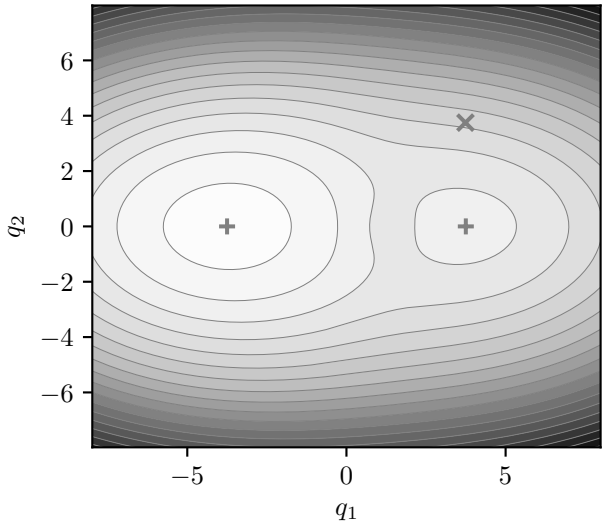

Figure 3.4: Ground state potential energy surface of the modified Displaced γ_2 model. The locations of the GMD component minima are marked with + signs. The position of the forced outlier from Eq. (3.53) is marked with a $\times$ sign.

We compare the convergence of $Z \approx \bar{f}$ between four scenarios. In Fig. 3.5(a), we have the simplest case: familiar stochastic component selection, and no forced outlier; here, the mean stabilizes relatively quickly, and the second component is sampled twice, as expected. The forced outlier is clearly visible in Fig. 3.5(b), where the stochastic algorithm accepts it as yet another point, greatly changing the mean and increasing the standard error. Fig. 3.5(c) features the deterministic algorithm and looks very similar to (a), except the second component is sampled multiple times during the bootstrap phase. Finally, in Fig. 3.5(d), the outlier is found to have a strong impact during bootstrapping, but this is quickly quelled; additionally, in the remainder of the calculation, the outlier causes more samples to be drawn from the second component. The final result is similar in all scenarios, except the stochastic algorithm with a forced outlier in Fig. 3.5(b), which fares poorly.

We also consider a more legitimate example, which doesn't require any elaborate setup of the sampling. The same model system and sampling distribution are used, but without any forced outliers. We run 1000 calculations with the stochastic component selection algorithm, and 1000 more with the deterministic one. These calculations are stopped when the standard error of the mean reaches 10^{-3}, and the number of samples (including the $N_{\text{boot}} = 64$ bootstrap samples per component for the deterministic version) is recorded. The numbers of samples are then binned to produce the histograms in Fig. 3.6, where it is clear that the deterministic method tends to require slightly fewer samples to achieve the same level of error. The reduction in the mean number of samples is about 1.7 %.

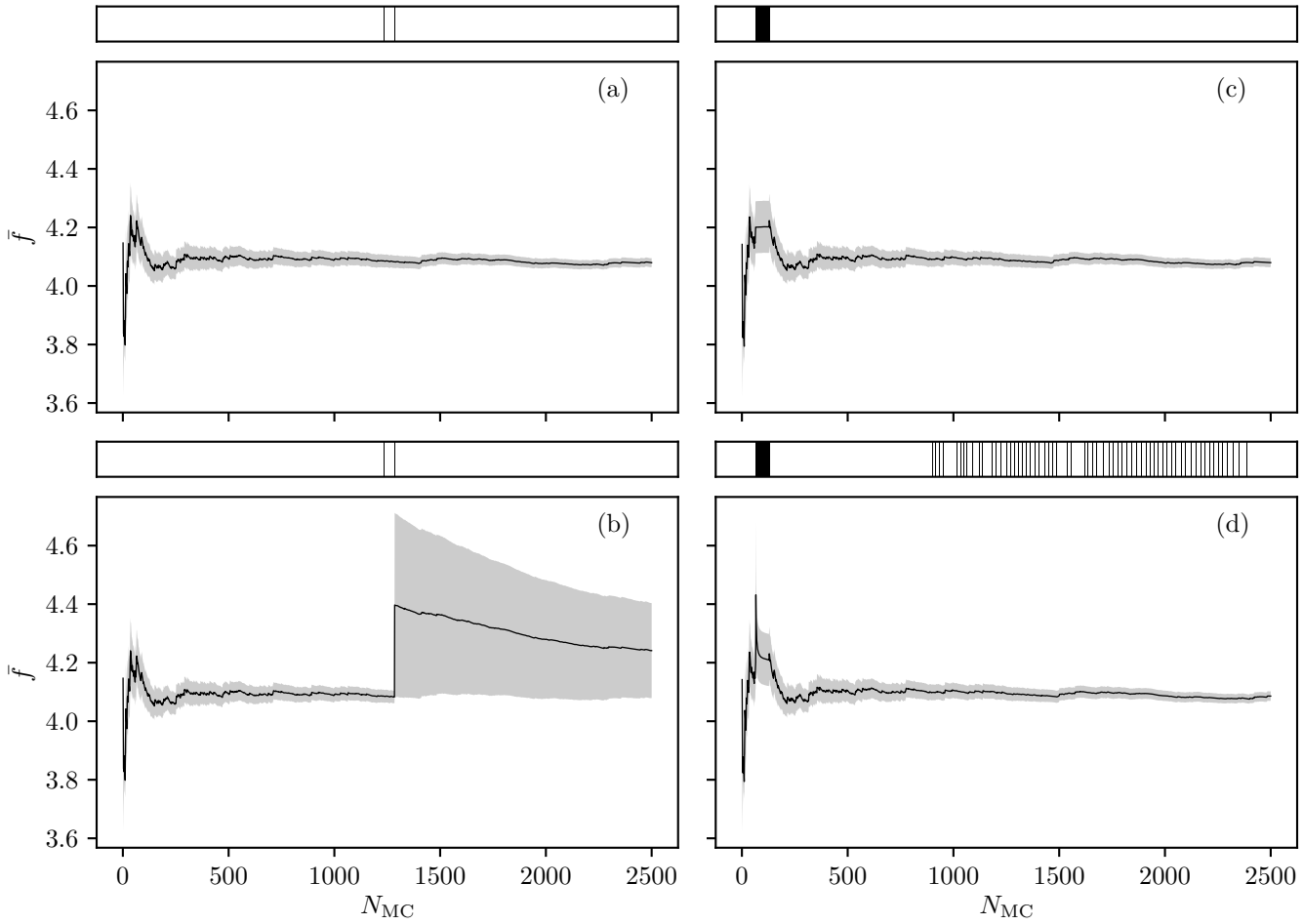

Figure 3.5: Convergence of $\bar{f}$ for the modified Displaced γ_2 model with $P = 16$ and (a) stochastic component selection with no outlier; (b) stochastic component selection with an outlier; (c) deterministic component selection with no outlier; (d) deterministic component selection with an outlier. The shaded area reflects the instantaneous error estimates. The vertical marks above the plots indicate steps at which the sample was drawn from the second component; in (c) and (d), this occurs N_{boot} times during bootstrapping.

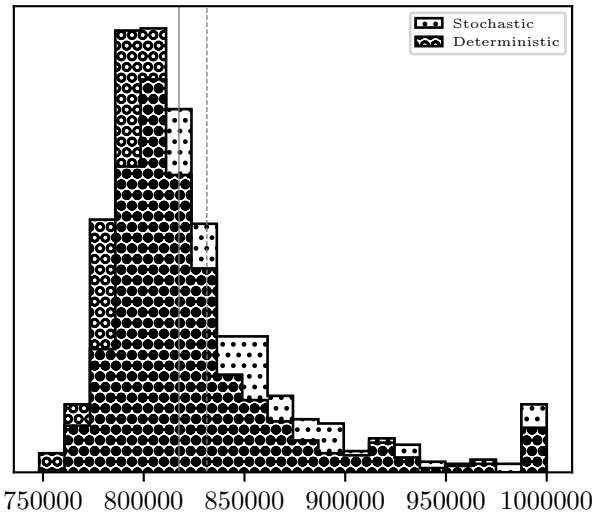

Figure 3.6: Distributions of the numbers of samples required to obtain a standard error of the mean of 10^{-3} for the modified Displaced γ_2 model. To avoid very wide histograms, the right-most bins collect all the points with more than 10^6 samples. The vertical lines indicate the means, dashed for stochastic and solid for deterministic.

3.4.2 Randomized quasi-Monte Carlo

In order to evaluate the efficiency of RqMC compared to MC, we use the same modified Displaced γ_2 model system in Tab. 3.1 as in Sec. 3.4.1, along with the same sampling distribution π_0. The approach for the comparison is simple: for several values of τ, we estimate $Z \approx \bar{f}$ from a fixed number of samples ($N_{\mathrm{MC}} = 10^5$), and compare it to the exact result Z_{Trotter}, which includes the systematic error due to the Trotter factorization. It is clear from Fig. 3.7 that RqMC tends to result in smaller error bars than MC for this system, even though we use the deterministic component selection algorithm for MC.

We also present in Fig. 3.8 a comparison of the error scaling with the number of samples. We compare the scaling for several P values (see Tab. 3.2), finding it extremely consistent for MC, where the slope is always near $-1/2$, indicating an asymptotic error scaling proportional to $N_{\mathrm{MC}}^{-\frac{1}{2}}$, which is precisely what one expects for Monte Carlo. In the RqMC case, the raw data are not as smooth (on account of being computed from just N_S points instead of N_{MC}), so the variability in the slopes is greater; still, we find that they are more negative than their MC counterparts, implying faster error reduction.

3.4.3 Parameter optimization

The GMD parameter optimization algorithm described in Sec. 3.3.3 is tested here using a simple model system derived from the Displaced model of Ref. [7] with the γ_6 parameter by reducing the

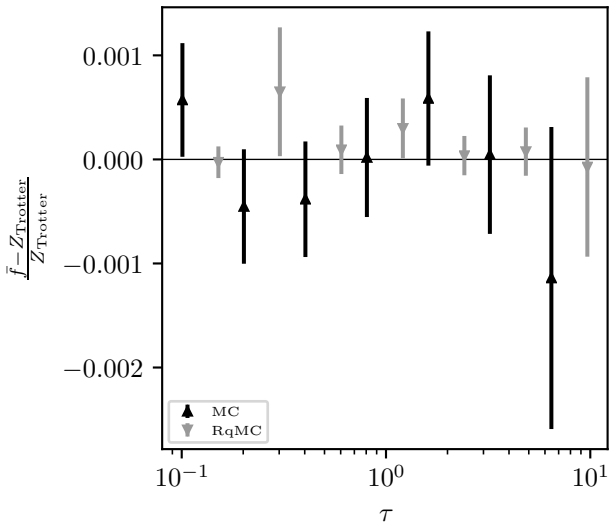

Figure 3.7: Comparison of RqMC with MC for the convergence of $\bar{f}$ with imaginary time step size τ. The data are plotted relative to the exact result Z_{Trotter}. Values of τ are staggered between the two methods for better visibility ($P = 6, 12, 24, 48, 96, 192, 384$ for MC and $P = 4, 8, 16, 32, 64, 128, 256$ for RqMC).

Table 3.2: Slopes of linear fits in Fig. 3.8.

P	Slope (MC)	Slope (RqMC)
16	−0.496 47	−0.558 76
32	−0.501 14	−0.587 58
64	−0.501 90	−0.640 40
128	−0.494 53	−0.652 34

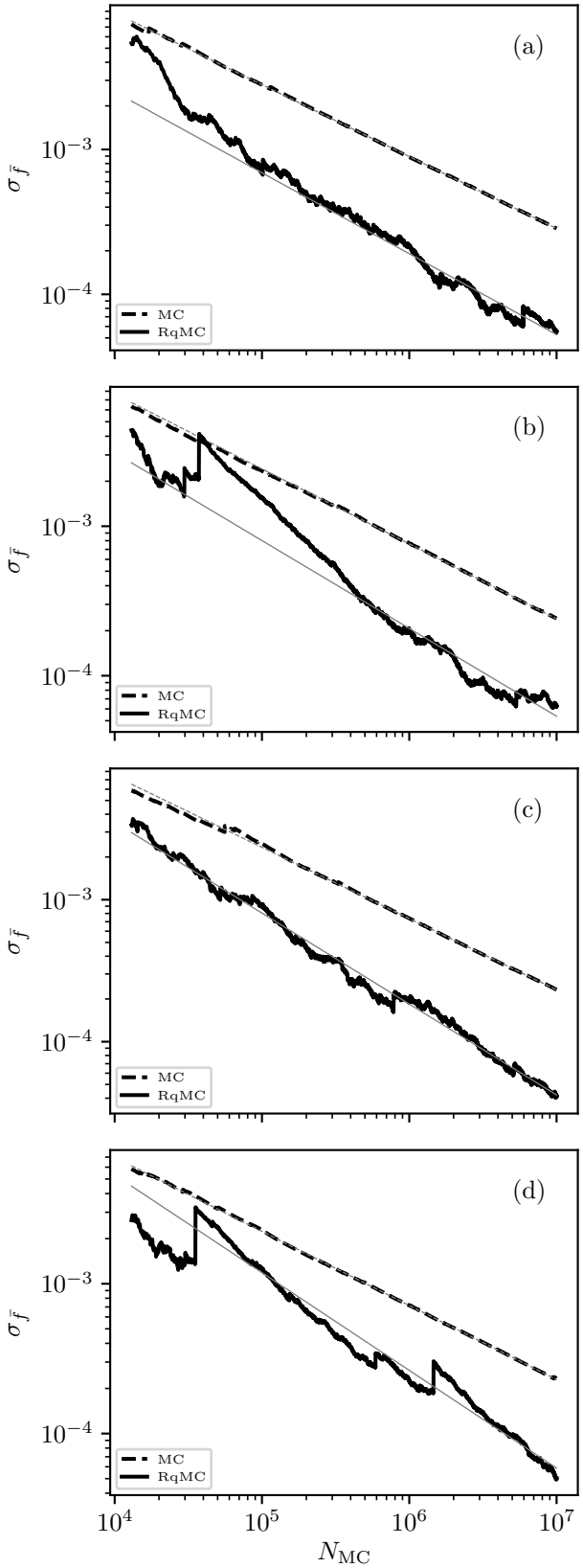

Figure 3.8: Scaling of the estimated error in $\bar{f}$ with the number of samples for $P =$ (a) 16, (b) 32, (c) 64, (d) 128. In grey are linear least squares fits whose slopes are given in Tab. 3.2.

Table 3.3: Parameters of the model Hamiltonian in Eq. (3.52) for the weakened Displaced γ_6 system.

Parameter	Value
E^1	0.02
E^2	0.04
ω_1	0.02
ω_2	0.04

Parameter	Value
λ	0.01
γ	0.05

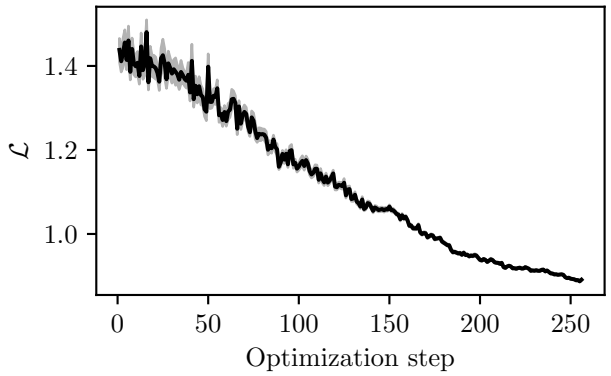

Figure 3.9: The loss function $\mathcal{L}$ at each step of the GMD optimization for the weakened Displaced γ_6 model. The uncertainty of the loss function estimates is indicated by the shaded area.

magnitude of the non-frequency parameters. The resulting system parameters are given in Tab. 3.3, while those for the algorithm are in Tab. 3.4.

The sampling distribution is initially determined from the diagonal portion of the Hamiltonian, but with each GMD component replicated to make up a total of $B = 8$ components. It is then optimized using Alg. 6 to better match the true distribution arising from the entire Hamiltonian. The progress of this optimization can be seen in Fig. 3.9 and Fig. 3.10; the former shows the loss function, while the latter displays the motion of the GMD components. As the oscillators find their way to the minima of the potential energy surface, the loss function steadily decreases. The standard error of the loss function estimates also lessens as the optimized distribution gets closer to the desired distribution. The components' journey is a fairly boring one: about half-way through, they get near the mimima, and towards the end, the ones closest to the minima have the largest weights.

To demonstrate that the optimized distribution π_{opt} is an improvement over the simple diagonal sampling distribution π_0, we compare estimates of the partition function $Z \approx \bar{f}$ at various values of P from $N_{\text{MC}} = 10^5$ samples. As evident in Fig. 3.11, π_0 already results in relatively small deviations from the exact values, but π_{opt} is undeniably the better choice.

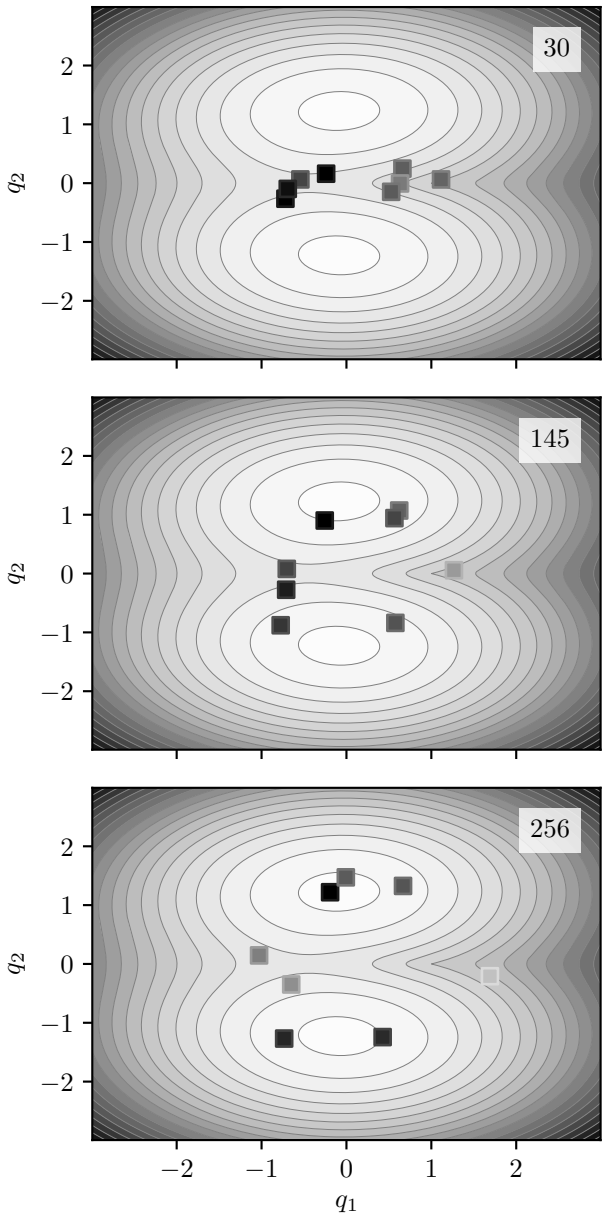

Figure 3.10: Snapshots of the GMD optimization for the weakened Displaced γ_6 model. The number in each panel indicates the optimization step. The background is the ground state potential energy surface of $\hat{H}$. The minimum of each GMD component is represented by a square, colored according to the relative component weights.

Table 3.4: Parameters for GMD optimization.

Parameter	Value	Description
P	64	# of beads
B	8	# of components
a	0.002	SPSA gain factor
N_{SPSA}	256	# of SPSA steps
N_{MC}	2000	# of samples per $\mathcal{L}$ estimation
ε	3	component repulsion stiffness
C_1	1	prefactor for repulsion term
C_2	1.5	prefactor for weight balancing term

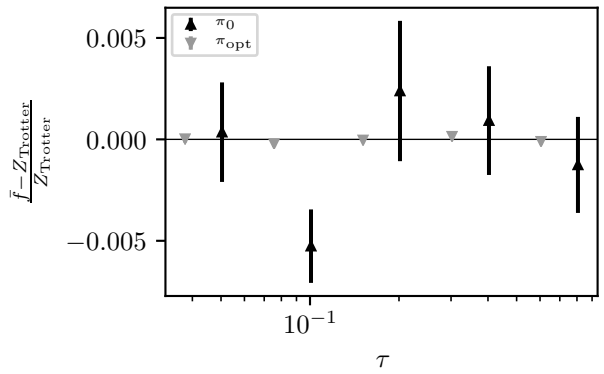

Figure 3.11: Comparison of sampling distributions for the convergence of $\bar{f}$ with imaginary time step size τ for the weakened Displaced γ_6 model. The data are plotted relative to the exact result Z_{Trotter}. Values of τ are staggered between the two distributions for better visibility ($P = 48, 96, 192, 384, 768$ for π_0 and $P = 64, 128, 256, 512, 1024$ for π_{opt}).

Table 3.5: Parameters for GMD optimization with deformation.

Parameter	Value	Description
P	64	# of beads
B	8	# of components
a	0.0005	SPSA gain factor
N_{SPSA}	64	# of SPSA steps
N_{W}	10	# of SPSA walkers
N_{MC}	2000	# of samples per $\mathcal{L}$ estimation
ε	3	component repulsion stiffness
C_1	1	prefactor for repulsion term
C_2	1.5	prefactor for weight balancing term

3.4.4 Parameter optimization with deformation

To evaluate the algorithm for parameter optimization with deformation, we use one system with weak coupling and two with strong coupling. The sampling distribution for each system is optimized using the procedure described in Sec. 3.3.4 and the parameters in Tab. 3.5.

To evaluate the quality of the optimized distribution π_{opt} for each system, we use it with the RqMC method to estimate the partition function $Z \approx \bar{f}$ at various values of P. For comparison, we also provide results computed using the best hand-picked sampling distribution of Ref. [7] for the system in question, which we call π_{best}. For both distributions, we use $N_{\mathrm{MC}} = 10^5$.

Displaced γ_2

We start with a weakly-coupled model system: the γ_2 version of the Displaced model from Ref. [7], which has the parameters given in Tab. 3.6. The optimization progress is shown in Fig. 3.12 and Fig. 3.13. Because this is a fairly simple system, the loss function never takes on large values and the optimization algorithm always makes the largest ν step possible, which makes for a quick optimization. As the Hamiltonian is deformed, the oscillators track the minima of the ground state potential energy surface. The end result has most of the components located in the deeper minimum and having larger weights, and two components in the shallower minimum with smaller weights.

As the reference distribution π_{best} for the sampling efficiency comparison, we simply use the one built from the diagonal portion of the Hamiltonian, which Ref. [7] calls ϱ_0. The results shown in Fig. 3.14(a) are quite promising: the estimates using π_{opt} are slightly better than the ones using π_{best}, which we suspect is due to the crowding of components in the deeper well.

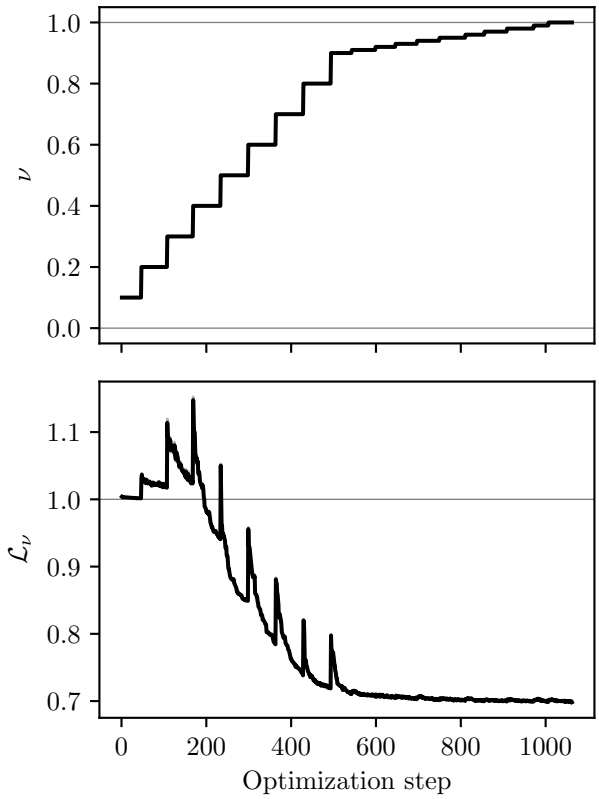

Figure 3.12: The value of ν (top panel) and the loss function $\mathcal{L}_\nu$ (bottom panel) at each step of the GMD optimization for the Displaced γ_2 model. In the top panel, the horizontal lines indicate the bounds on ν. In the bottom panel, the horizontal line marks the lower bound on the desired loss function values, and the uncertainty of the loss function estimates is indicated by the shaded area.

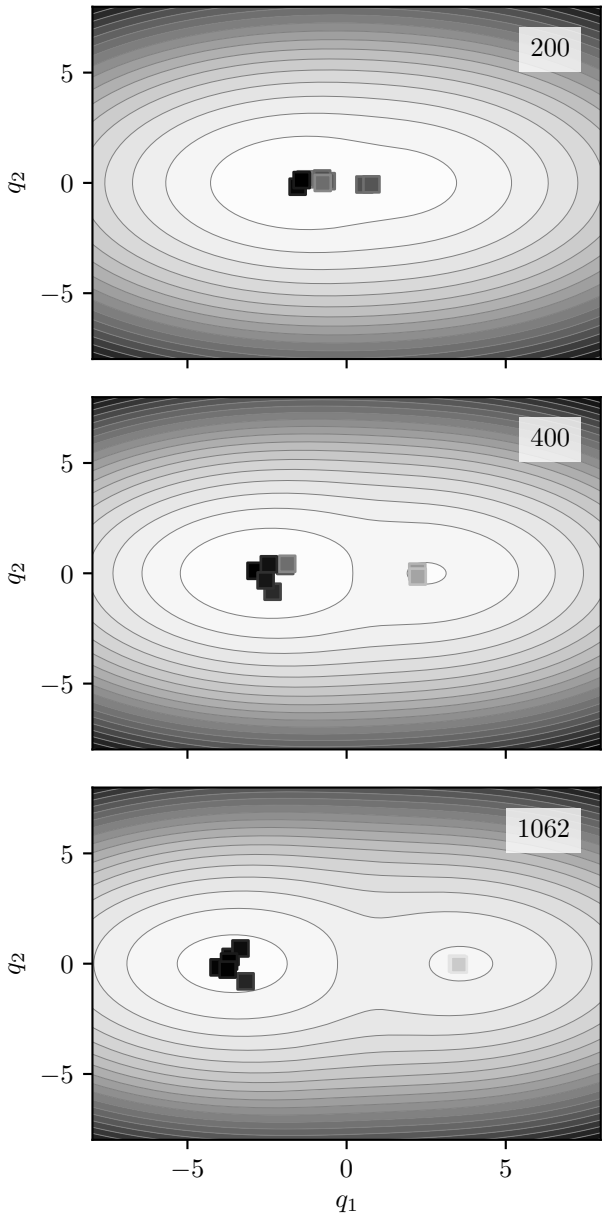

Figure 3.13: Snapshots of the GMD optimization for the Displaced γ_2 model. The number in each panel indicates the optimization step. The backgrounds are the ground state potential energy surfaces of $\hat{H}_\nu$. The minimum of each GMD component is represented by a square, colored according to the relative component weights.

Table 3.6: Parameters of the model Hamiltonian in Eq. (3.52) for the Displaced γ_2 system.

Parameter	Value
E^1	0.0996
E^2	0.1996
ω_1	0.02
ω_2	0.04

Parameter	Value
λ	0.072
γ	0.04

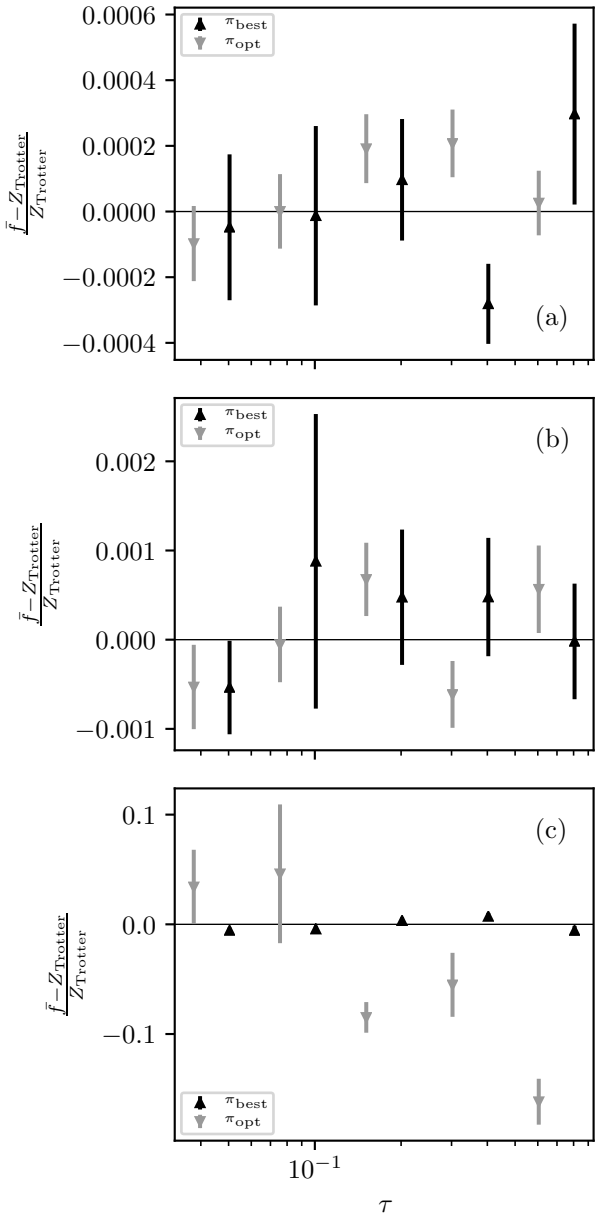

Figure 3.14: Comparison of sampling distributions for the convergence of $\bar{f}$ with imaginary time step size τ for (a) Displaced γ_2, (b) Displaced γ_6, and (c) Jahn–Teller λ_6. The data are plotted relative to the exact result Z_{Trotter}. Values of τ are staggered between the two distributions for better visibility ($P = 48, 96, 192, 384, 768$ for π_{best} and $P = 64, 128, 256, 512, 1024$ for π_{opt}).

Table 3.7: Parameters of the model Hamiltonian in Eq. (3.52) for the Displaced γ_6 system.

Parameter	Value
E^1	0.0996
E^2	0.1996

Parameter	Value
ω_1	0.02
ω_2	0.04

Parameter	Value
λ	0.072
γ	0.2

Displaced γ_6

The first strongly-coupled model system that we consider is the γ_6 variant of the Displaced model, with the parameters in Tab. 3.7, and results in Fig. 3.15 and Fig. 3.16. The optimization algorithm struggles when the minima first appear (around steps 100 to 350; top panel of Fig. 3.16), but it is eventually able to find them. This difficulty is clear from the large and noisy values of the loss function and the small ν steps. However, once the minima are found, the rest of the optimization proceeds fairly well.

Unfortunately, many of the components in the final distribution are of no consequence (bottom panel of Fig. 3.16). As they have very low weights, this does not put a damper on the estimation of the partition function seen in Fig. 3.14(b). Again, both sampling distributions (in this case π_{best} is what Ref. [7] calls ϱ_1) work well, but the optimized distribution π_{opt} ekes out a tiny decrease in the error bars, possibly due to slightly more thorough coverage of one of the wells.

Jahn–Teller λ_6

As the other strongly-coupled model system we use the λ_6 variant of the Jahn–Teller model from Ref. [7], whose parameters are given in Tab. 3.8. The results can be found in Fig. 3.17 and Fig. 3.18. The early part of the optimization is a bit uncertain, as the components search for the circular well (top panel of Fig. 3.18). Once they fall into the well around step 350, the situation temporarily improves. However, as the symmetric well grows, more regions become neglected and loss function estimation becomes more difficult. Additionally, there are significant fluctuations in the components weights; at times, only a single component will have most of the weight (middle panel of Fig. 3.18). Still, by the end of the optimization (bottom panel of Fig. 3.18), the components are reasonably well spread out with somewhat even weights, which attests to the importance of the penalty terms in Eq. (3.51) for avoiding clustering of components and domination by a single component.

Likely owing to the uneven distribution of the components in the well, the partition function estimation in Fig. 3.14(c) does not perform as well with the optimized distribution π_{opt} as it does with π_{best} (in this case, ϱ_2 of Ref. [7], which also has 8 components). This could conceivably be remedied by including more components in the GMD or adjusting the ε and C_1 parameters responsible for the repulsion penalty term of the loss function.

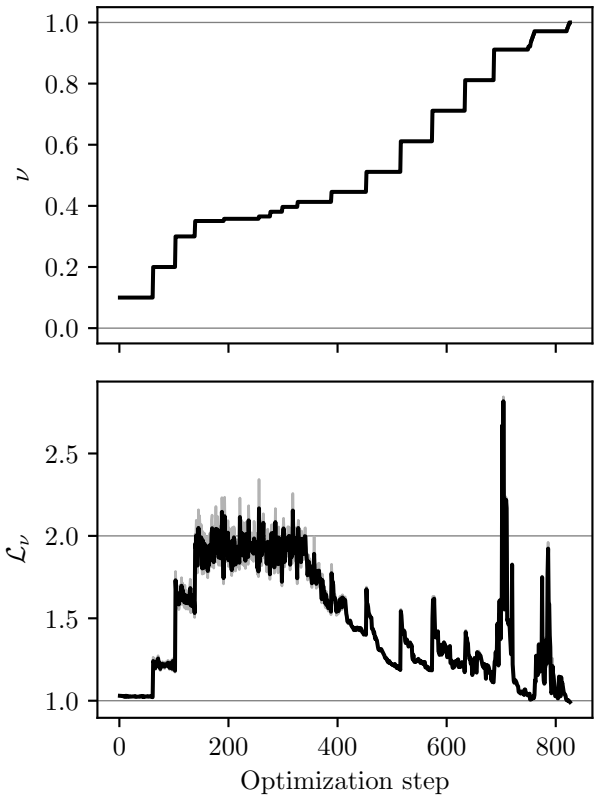

Figure 3.15: The value of ν (top panel) and the loss function $\mathcal{L}_\nu$ (bottom panel) at each step of the GMD optimization for the Displaced γ_6 model. In the top panel, the horizontal lines indicate the bounds on ν. In the bottom panel, the horizontal lines mark the bounds on the desired loss function values, and the uncertainty of the loss function estimates is indicated by the shaded area.

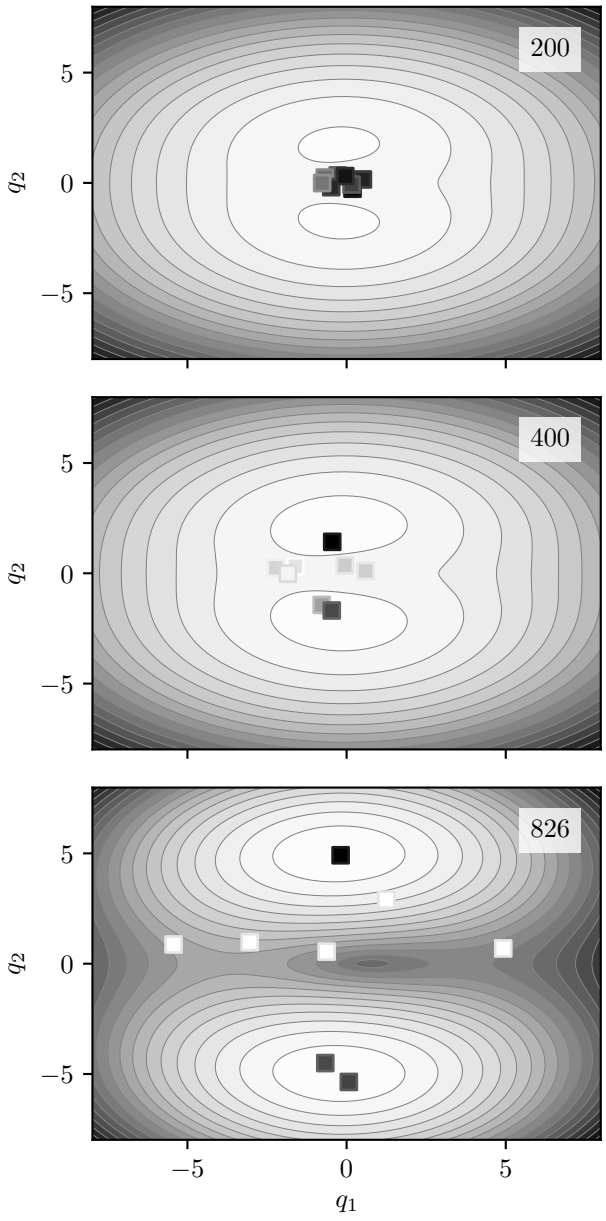

Figure 3.16: Snapshots of the GMD optimization for the Displaced γ_6 model. The number in each panel indicates the optimization step. The backgrounds are the ground state potential energy surfaces of $\hat{H}_\nu$. The minimum of each GMD component is represented by a square, colored according to the relative component weights.

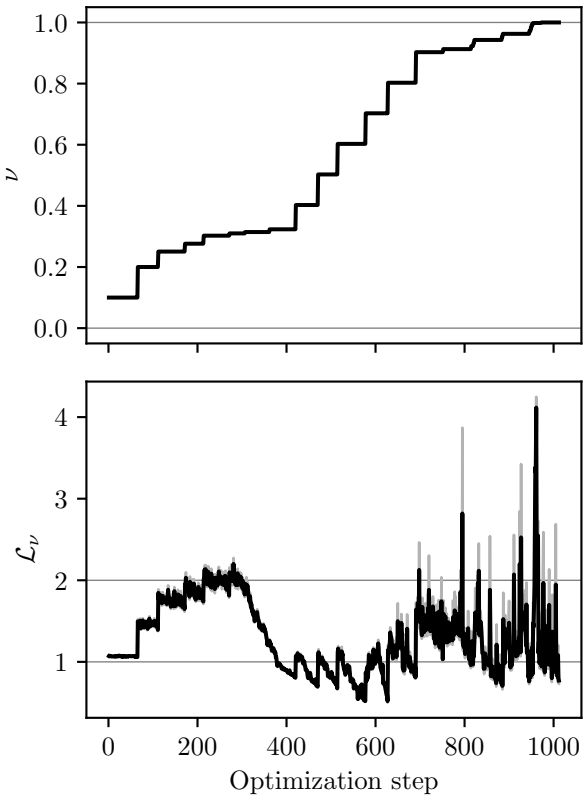

Figure 3.17: The value of ν (top panel) and the loss function $\mathcal{L}_\nu$ (bottom panel) at each step of the GMD optimization for the Jahn–Teller λ_6 model. In the top panel, the horizontal lines indicate the bounds on ν. In the bottom panel, the horizontal lines mark the bounds on the desired loss function values, and the uncertainty of the loss function estimates is indicated by the shaded area.

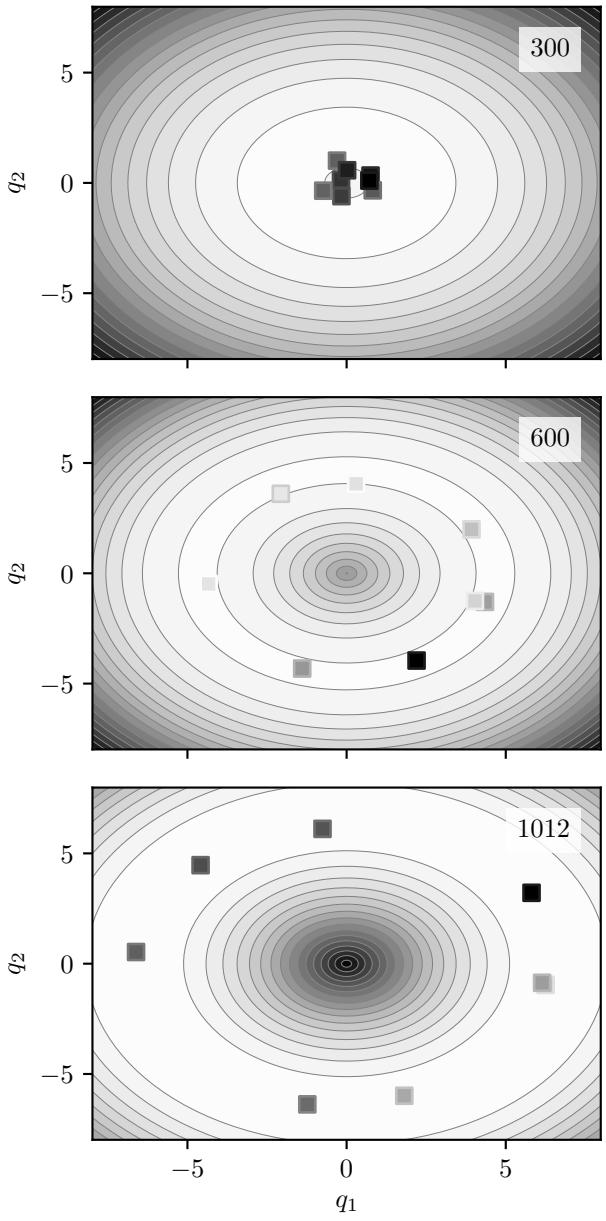

Figure 3.18: Snapshots of the GMD optimization for the Jahn–Teller λ_6 model. The number in each panel indicates the optimization step. The backgrounds are the ground state potential energy surfaces of $\hat{H}_\nu$. The minimum of each GMD component is represented by a square, colored according to the relative component weights.

Table 3.8: Parameters of the model Hamiltonian in Eq. (3.52) for the Jahn–Teller λ_6 system.

Parameter	Value
E^1	0.631 35
E^2	0.631 35
ω_1	0.03
ω_2	0.03

Parameter	Value
λ	0.2
γ	0.2

3.5 Conclusions

We have demonstrated four methods to enhance partition function estimation for vibronic Hamiltonians: deterministic component selection in GMD sampling when using MC (Sec. 3.3.1, Sec. 3.4.1); RqMC as a substitute for MC in GMD sampling (Sec. 3.3.2, Sec. 3.4.2); and optimization of the GMD parameters without (Sec. 3.3.3, Sec. 3.4.3) and with deformation (Sec. 3.3.4, Sec. 3.4.4). Each of these was shown to improve the scheme described in Ref. [7].

We have found that choosing the components of a GMD in an MC calculation deterministically as opposed to stochastically can slightly reduce the standard error of the mean with a fixed number of samples (or, conversely, reduce the number of samples for a fixed standard error). Additionally, it can diminish the impact of extreme outliers. Thus, we recommend the use of deterministic component selection for MC partition function calculations of vibronic models, and for GMD sampling in general.

However, when possible, RqMC should be used in place of MC. We have shown that employing quasi-random numbers in addition to pseudo-random numbers can help improve the rate at which the stochastic error decreases with the number of samples, making RqMC the more efficient choice. In subsequent studies, the impact of the N_S and N_boot parameters on the quality of the error estimate should be determined.

Finally, the optimization of GMD parameters was observed to improve the sampling efficiency when compared to the simplest choice of sampling distribution, and sometimes even when compared to a hand-crafted sampling distribution. For the version without deformation of the target distribution, we have presented an example of a distribution which is significantly improved in just 256 steps. Because of the way the sampling GMD is constructed from Hamiltonian parameters, even though the optimization was carried out at $P = 64$ beads, the distribution obtained for $P = 1024$ functions very well.

The iterative algorithm to optimize GMD parameters with deformation of the target distribution was seen to perform rather well for two of the three systems, but only tolerably for the third. It was able to find the regions of high density for each model system, resulting in very well-behaved sampling at several values of τ, although the highly symmetric Jahn–Teller λ_6 system was shown to

give the method some trouble. Because the algorithm parameter values were the same across all the model systems, this appears to be a fairly robust and general method, and we are interested in seeing it applied to realistic molecular systems.

Several upgrades to the optimization are possible which may improve its efficiency, and which we hope to attempt in the future. The first is to use different magnitudes for the different components of the SPSA perturbation vector $\mathbf{\Delta}$, allowing the $E^{(0)}$ parameters to evolve at a different rate than the $E^{(1)}$ parameters. Another is to allow low-weight components to be culled and replaced by clones of high-weight components. Finally, if the loss function is seen to be decreasing rapidly, it may be beneficial to continue beyond N_{SPSA} optimization steps before moving on to the next value of ν.

Acknowledgments

P.-N.R. (RGPIN-2016-04403) and M.N. (RGPIN-2018-04187) acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC). P.-N.R. acknowledges the Ontario Ministry of Research and Innovation (MRI), the Canada Research Chair program (950-231024), the Canada Foundation for Innovation (CFI) (project No. 35232), and the Canada First Research Excellence Fund (CFREF).

3.A Demonstration of quasi-Monte Carlo

At its core, quasi-Monte Carlo (qMC) is just plain Monte Carlo (MC), but with the pseudo-random number generator (RNG) replaced by a low-discrepancy sequence (LDS). Informally, the discrepancy of a sequence is its maximum local deviation from a uniform density.[89] LDSs attempt to minimize the discrepancy of the generated points and therefore cover space more evenly than RNGs.[85] The hope is that this can help reduce the stochastic error when using a statistical estimator.

The canonical introductory MC problem is that of estimating π (the ratio of a circle's circumference to its diameter) by sampling uniformly from a square of unit side length, and counting the fraction F of points lying within one unit of one of the corners. Since the ratio of areas is $\pi/4$, $4F$ should tend to π. In this section, we compare the behavior of MC to qMC for this elementary problem. For the former, we use the Mersenne Twister, an extremely popular RNG;[104] for the latter, we use a two-dimensional Sobol sequence.[91]

As seen in Fig. 3.19(a), the RNG creates a typically irregular pattern, with clumps and empty spaces. On the other hand, in Fig. 3.19(b), the quasi-random points are extremely uniform. It is this regularity that results in the faster convergence of qMC shown in Fig. 3.20. Although qMC is not a panacea for numerical integration, this example demonstrates that it can have improved performance over plain MC.

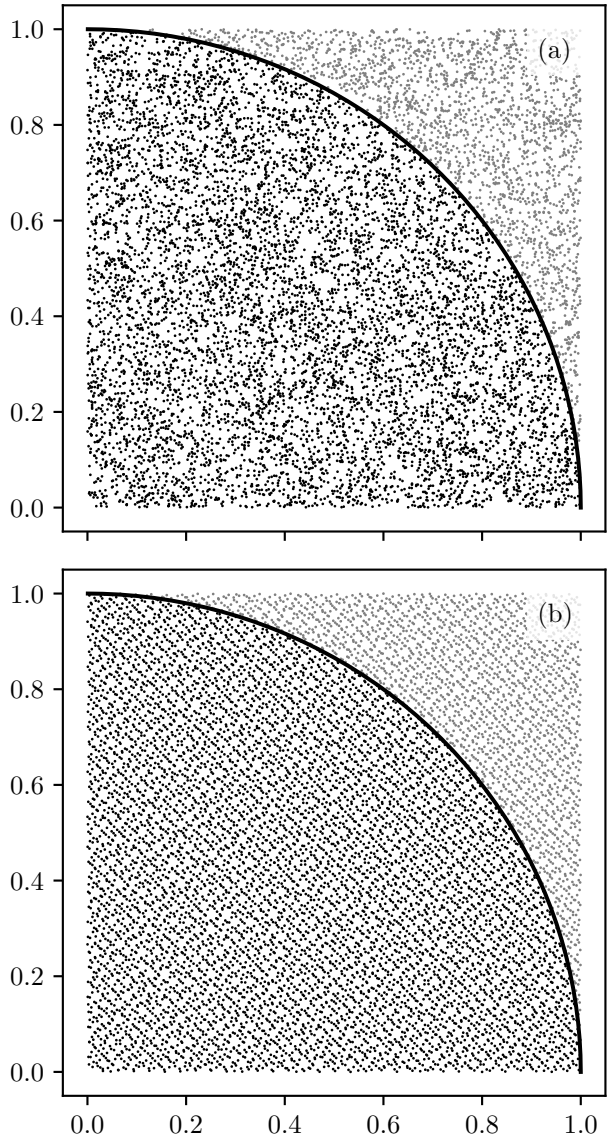

Figure 3.19: Distribution of 10^4 points sampled (a) pseudo-randomly using a Mersenne Twister RNG and (b) quasi-randomly using a Sobol sequence. The curve is a segment of the unit quarter-circle centered at the origin; points inside it are darker, and those outside are lighter.

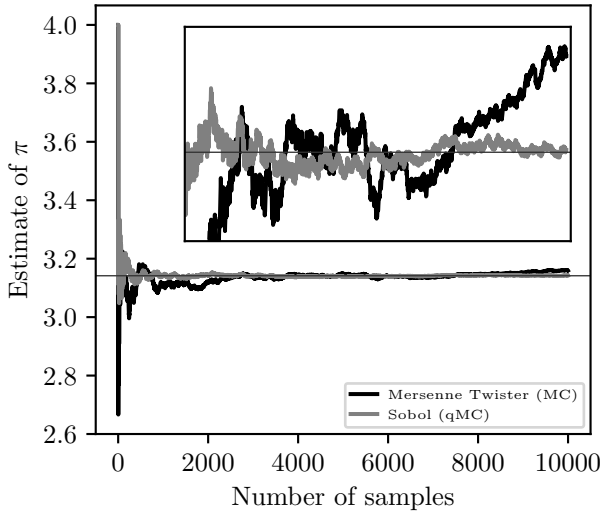

Figure 3.20: Convergence of MC and qMC methods for the estimation of π. The exact value is shown with a horizontal line. The inset shows the data directly below it.

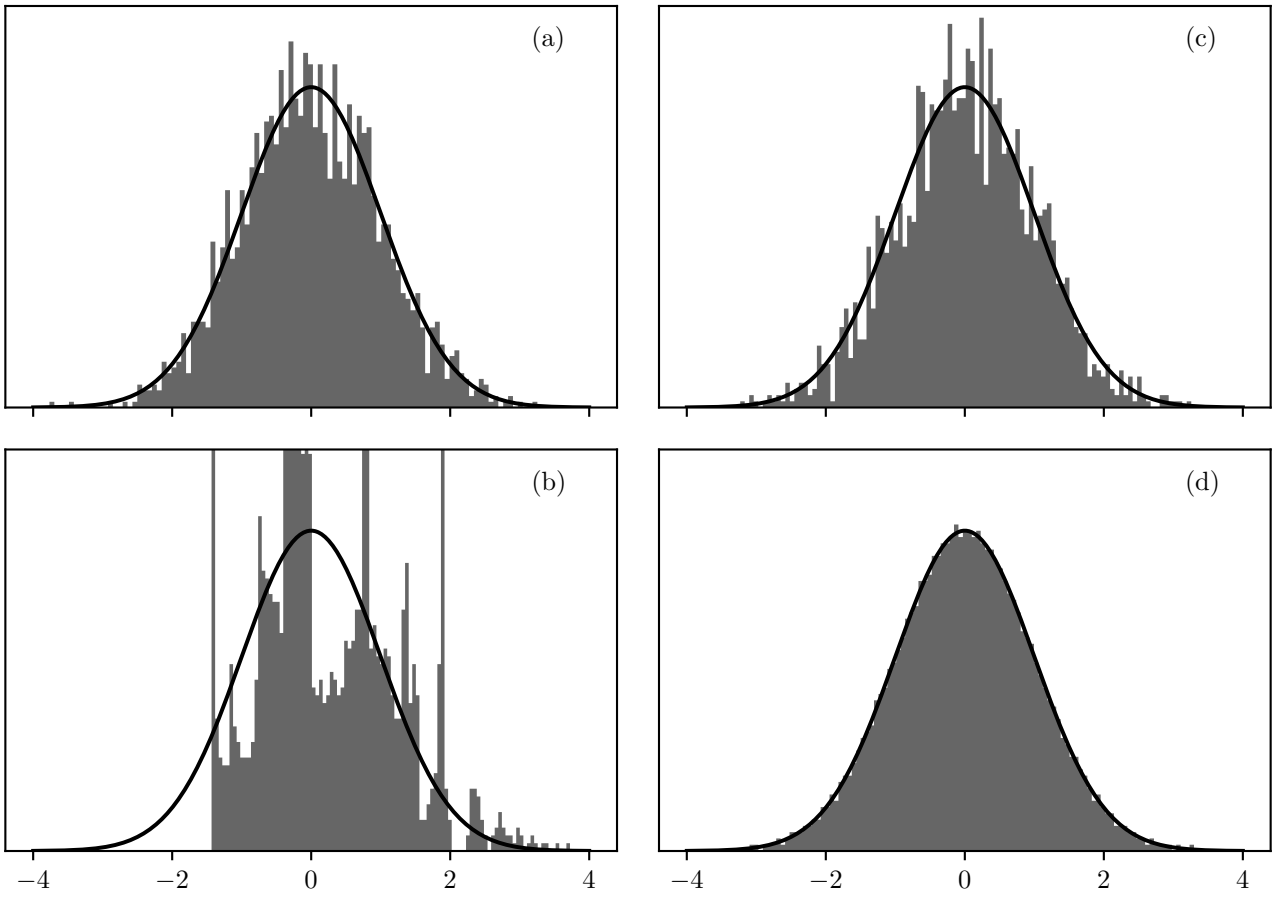

Figure 3.21: Histograms comparing pseudo-random and quasi-random Gaussian sampling methods: (a) Box–Muller with Mersenne Twister RNG; (b) Box–Muller with Sobol LDS; (c) cdf inversion with Mersenne Twister RNG; (d) cdf inversion with Sobol LDS. Each histogram is generated from 2000 samples grouped into 100 bins. In (c), some bins are cut off at the top to show more detail elsewhere. The true pdf is drawn as a solid curve.

3.B Gaussian sampling in quasi-Monte Carlo

One method for sampling from an arbitrary one-dimensional probability distribution function (pdf) uses the inverse of the corresponding cumulative distribution function (cdf). If uniformly distributed samples are input into the inverse cdf, the outputs are distributed according to the desired pdf. Unfortunately, this is often computationally taxing, so specialized methods are preferred. For the case of Gaussian random variables, two common alternatives are Box–Muller and ziggurat.[105,106]

While these methods perform well with pseudo-random numbers, they destroy the low-discrepancy nature of quasi-random numbers. Compare, for example, the histogram obtained from sampling a standard Gaussian with the Box–Muller method and the Mersenne Twister pseudo-random number generator (RNG) in Fig. 3.21(a) and the same method applied to a Sobol low-discrepancy sequence (LDS) in Fig. 3.21(b). The latter does an extremely poor job of describing the true pdf.

Despite its cost, the preferred method in this circumstance turns out to be straightforward cdf inversion.[84,87] Application of this method is shown in Fig. 3.21(c) using the Mersenne Twister RNG and in Fig. 3.21(d) using a Sobol LDS. It is evident that among the four options presented in Fig. 3.21, cdf inversion of an LDS results in by far the smoothest histogram. Thus, we have chosen this approach as the basis of low-discrepancy Gaussian sampling in the present work.

Nevertheless, even cdf inversion is imperfect with a many-dimensional LDS. Because the higher dimensions of a Sobol sequence tend to be more strongly correlated before the inversion,[85] these correlations are carried through to the generated points. In Fig. 3.22, we show two-dimensional cuts of a 32-dimensional point set created from a 32-dimensional LDS using cdf inversion separately for each dimension. While there are no obvious correlations present in Fig. 3.22(a) between the second and third dimensions, we see some glaring patterns in Fig. 3.22(b), where dimensions 27 and 32 are shown. However, these correlations do not prevent us from using quasi-random numbers as a smoother alternative to pseudo-random numbers for multidimensional Gaussian sampling.

3.C Example of randomized quasi-Monte Carlo

The flexibility allowed in randomized quasi-Monte Carlo (RqMC) by changing the number of sequences N_S creates a tension between pure qMC ($N_\mathrm{S} = 1$), which lacks the ability to accurately estimate error bars, and pure Monte Carlo ($N_\mathrm{S} = N_\mathrm{MC}$), which misses out on the low-discrepancy smoothness. In this section, we demonstrate the practical implications of this balance by estimating the 64-dimensional integral

$$I = \frac{1}{(2\pi)^{32}} \int \mathrm{d}x_1 \cdots \int \mathrm{d}x_{64}\; e^{-\sum_{j=1}^{64} \frac{x_j^2}{2}} \sum_{j=1}^{64} x_j^2. \tag{3.57}$$

Factoring and evaluating these integrals individually reveals the exact solution $I = 64$.

To compute I using Monte Carlo (MC), we sample $N_\mathrm{MC} = 2^{14}$ times from a collection of 64 uncoupled standard Gaussian distributions. For RqMC, we instead obtain $N_\mathrm{qMC} = N_\mathrm{MC}/N_\mathrm{S}$ samples from each of N_S 64-dimensional Sobol sequences, then transform them using cdf inversion, and combine them into a single point with error bars. In either case, we use the estimator

$$\sum_{j=1}^{64} x_j^2 \tag{3.58}$$

and the total number of sampled points is always the same.

The results in Fig. 3.23 make it clear that for such a well-behaved integrand, RqMC does not require all that many sequences in order to generate faithful error bars. Additionally, the error bars for

$N_S \ll N_{MC}$ are substantially smaller than the MC ones, suggesting improved performance. When $N_S \approx N_{MC}$, the RqMC error bars are of a comparable size to their MC counterpart, which confirms that RqMC matches MC in this limit.

3.D Variance of linear combination

Consider a collection of B uncorrelated random variables $\{r^b\}_{b=1}^B$, from which we form a new random variable

$$r = \sum_{b=1}^{B} w^b r^b \tag{3.59}$$

using the constant coefficients w^b. If each r^b has population variance

$$\sigma_{r^b}^2 = \left\langle \left(r^b - \langle r^b \rangle\right)^2 \right\rangle = \langle (r^b)^2 \rangle - \langle r^b \rangle^2 , \tag{3.60}$$

then

$$\sigma_r^2 = \left\langle (r - \langle r \rangle)^2 \right\rangle \tag{3.61a}$$

$$= \sum_{b=1}^{B} \sum_{b'=1}^{B} w^b w^{b'} \left\langle (r^b - \langle r^b \rangle)(r^{b'} - \langle r^{b'} \rangle) \right\rangle . \tag{3.61b}$$

For $b \neq b'$, r^b and $r^{b'}$ are uncorrelated, so

$$\left\langle (r^b - \langle r^b \rangle)(r^{b'} - \langle r^{b'} \rangle) \right\rangle = 0. \tag{3.62}$$

Thus,

$$\sigma_r^2 = \sum_{b=1}^{B} (w^b)^2 \left\langle \left(r^b - \langle r^b \rangle\right)^2 \right\rangle = \sum_{b=1}^{B} (w^b)^2 \sigma_{r^b}^2 . \tag{3.63}$$

Suppose that $s_{r^b}^2$ is an unbiased estimator of $\sigma_{r^b}^2$, meaning that

$$\left\langle s_{r^b}^2 \right\rangle = \sigma_{r^b}^2 . \tag{3.64}$$

Then

$$s_r^2 = \sum_{b=1}^{B} (w^b)^2 s_{r^b}^2 \tag{3.65}$$

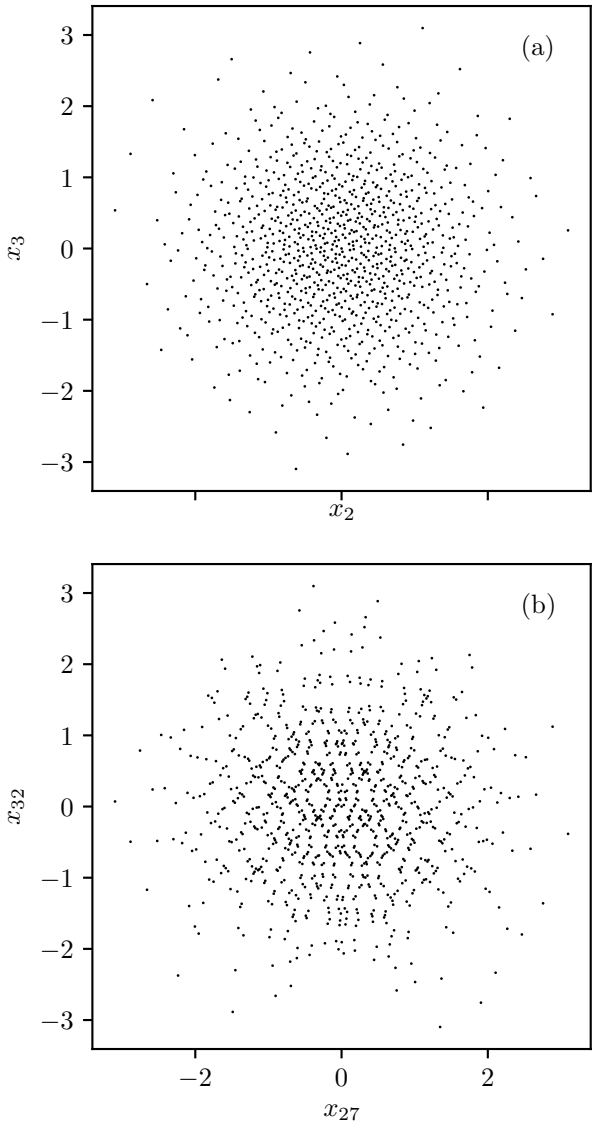

Figure 3.22: Cuts along dimensions (a) 2 and 3 and (b) 27 and 32 of a 32-dimensional Gaussian LDS point set containing 1000 points.

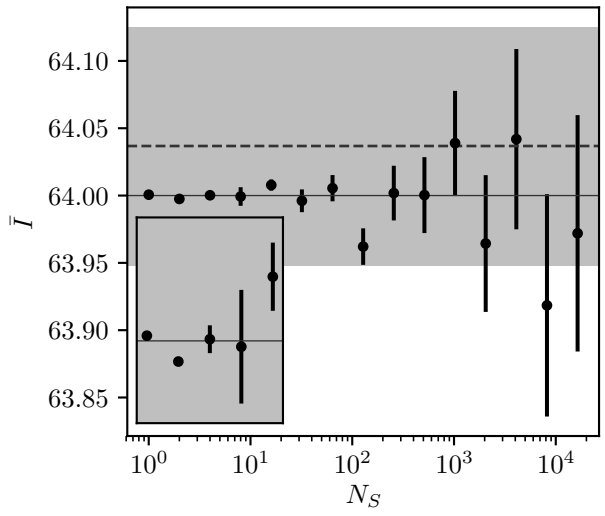

Figure 3.23: Estimation of I in Eq. (3.57) using RqMC. The exact value is shown with a solid horizontal line. The MC result is shown with a dashed horizontal line and the shaded area indicates the extent of the MC error bars. The inset shows the data directly above it.

Table 3.9: Parameters of the model Hamiltonian in Eq. (3.67).

Parameter	Value
ω_1	0.1
$\gamma_{1,1}$	1.0

Parameter	Value
$\gamma_{1,2}$	0.1
$\gamma_{2,2}$	0.01

is an unbiased estimator of σ_r^2, since

$$\left\langle s_r^2 \right\rangle = \sum_{b=1}^{B} (w^b)^2 \left\langle s_{r^b}^2 \right\rangle = \sum_{b=1}^{B} (w^b)^2 \sigma_{r^b}^2 = \sigma_r^2. \tag{3.66}$$

3.E Example of negative $g(\mathbf{R})/Z$

It was noted in Ref. [7] that $g(\mathbf{R})$, as defined in Eq. (3.6), may take on negative values. If $g(\mathbf{R}) > 0$ for some $\mathbf{R}$, but $g(\mathbf{R}) < 0$ for others, $g(\mathbf{R})$ cannot always agree in sign with any constant real number Z, so $g(\mathbf{R})/Z$ must be negative at some point and is not a pdf. We give a simple example of this using the Hamiltonian

$$\hat{H} = \begin{pmatrix} \hat{h}_\mathrm{o} & 0 \\ 0 & \hat{h}_\mathrm{o} \end{pmatrix} + \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} \\ \gamma_{1,2} & \gamma_{2,2} \end{pmatrix} \hat{q}_1, \tag{3.67}$$

with the parameters in Tab. 3.9. At an inverse temperature of $\beta = 38.7$, with $P = 3$ beads located at $\mathbf{R} = (1, 2, 0)$, we obtain

$$g(\mathbf{R}) \approx 0.003\,116\,5. \tag{3.68}$$

However, at $\mathbf{R} = (1, 2, -0.3)$, we instead find

$$g(\mathbf{R}) \approx -0.010\,741. \tag{3.69}$$

3.F Explanation of ν step size selection algorithm

The function STEP_NU for choosing the step size $\Delta\nu$ in Sec. 3.3.4 is given in Alg. 7, and is based on a number of heuristics. The core assumption is that (all else being fixed) the loss function grows monotonically with the step size. We justify this by pointing out that the relative entropy term which appears in the loss function quantifies the difference between the GMD used for sampling and the target distribution, and this difference should grow with increasing deformation of the target distribution.

We use this assumption to tune the loss function to be between 1 and 2 by varying $\Delta\nu$ with a simple binary search. If the loss function is too low (and we would not be optimizing very much this iteration), the lower bound $\Delta\nu_{\text{min}}$ is set to the current value of $\Delta\nu$; if it is too high (and we might be approaching a regime where the sampling is not trustworthy), the upper bound $\Delta\nu_{\text{max}}$ is set to the current value of $\Delta\nu$. After each update, the new trial value of $\Delta\nu$ is set to be the average of the lower and upper bounds. The values of 1 and 2 for the loss function bounds, as well as the number of binary search iterations, are entirely ad hoc, but appear to be reasonable for the systems considered in this work.

We start the process by optimistically guessing that the current step should be twice as large as the previous step. We also take steps that are no smaller than 10^{-3} so that the optimization doesn't grind to a halt in difficult regions, and no larger than 10^{-1} (or 10^{-2} close to the end) to ensure that we don't skip over any details.

Chapter 4

On the quantum mechanical potential of mean force. I. A path integral perspective

This chapter is reproduced from Dmitri Iouchtchenko, Kevin P. Bishop, and Pierre-Nicholas Roy, "On the quantum mechanical potential of mean force. I. A path integral perspective", arXiv:2101.00761v1.[3]

4.1 Introduction

Free energy calculations provide vital theoretical insights into the behaviour of chemical systems (such as equilibrium constants and stability of molecular conformations) and are commonly used to make comparisons with experimental results. The classical potential of mean force (PMF), also referred to as a free energy profile, may be obtained from classical molecular dynamics simulations in a number of ways: umbrella sampling with the weighted histogram analysis method (WHAM),[23,107,108] blue moon sampling,[109–111] metadynamics,[112,113] and potential of mean constraint force,[114,115] among others.

The quantum mechanical analogue of the PMF has not been as thoroughly studied, but there are many interesting systems where nuclear quantum effects play an integral role. These effects are essential within simulations performed at low temperature or containing light atoms. Crucially, the inclusion of nuclear quantum effects has been demonstrated to be indispensable in the accurate determination of properties of the water dimer as well as other small water clusters due to the presence of the light hydrogen atoms.[116–118] Recent work has combined the existing umbrella sampling method with path integral molecular dynamics simulations to study the free energy profile of water–water[24,119] and water–methanol dimers[119] while accounting for such nuclear quantum effects.

Many existing efforts to compute the quantum PMF[120–124] have used the path integral centroid coordinate,[125,126] but this is known to produce deviations from the exact quantum mechanical result.[24,127] Our current focus is therefore to obtain the PMF as a function of the true reaction coordinate rather than the path centroid.

In the present work, we formally derive two path integral Monte Carlo (PIMC) estimators for the derivative of the PMF, which can be integrated to determine the PMF. The first estimator is obtained by performing an analytical differentiation of the exact path integral, followed by its discretization over the path. Conversely, the second estimator is derived by initially discretizing the path integral before performing the analytical differentiation. Theoretically, these estimators should provide the same numerical results, and to verify this, they are benchmarked against known model systems.

Both estimators may be used in conjunction with the path integral Langevin equation (PILE),[16] as we show in Paper II of this series, titled "Constrained path integral molecular dynamics integrators". Together with the constrained PILE formulation, we aim to use these estimators for systems that are not as easily studied with PIMC, such as low-temperature molecular clusters.

The remainder of this article is organized as follows: in Sec. 4.2, we describe our notation; in Sec. 4.3, we develop two estimators for the derivative of the PMF; in Sec. 4.4, we apply the estimators to model systems; and in Sec. 4.5, we summarize our findings.

4.2 Background

We consider systems with f Cartesian degrees of freedom, that we label q_1, q_2, ..., q_f; commonly, there are N particles in three spatial dimensions, in which case $f = 3N$. For convenience, we group them into a single vector $\mathbf{q}$.

We restrict the Hamiltonian to have the form

$$\hat{H} = \hat{K} + \hat{V} = \sum_{i=1}^{f} \frac{\hat{p}_i^2}{2m_i} + V(\hat{\mathbf{q}}) = \frac{1}{2}\hat{\mathbf{p}} \cdot \mathbf{M}^{-1} \cdot \hat{\mathbf{p}} + V(\hat{\mathbf{q}}), \tag{4.1}$$

where $\hat{p}_i$ is the momentum operator conjugate to the position operator $\hat{q}_i$, m_i is the mass corresponding to q_i, and $\mathbf{M}$ is the diagonal mass matrix whose elements are m_i. The restriction on the kinetic energy allows us to write the exact free particle propagator

$$\langle \mathbf{q}' | e^{-\tau \hat{K}} | \mathbf{q} \rangle = \sqrt{\frac{|\mathbf{M}|}{(2\pi\hbar^2\tau)^f}}\, e^{-\frac{1}{2\hbar^2\tau}(\mathbf{q}'-\mathbf{q})\cdot\mathbf{M}\cdot(\mathbf{q}'-\mathbf{q})} \tag{4.2}$$

for an imaginary time duration τ. We require that the potential energy be diagonal in the position

representation so that

$$\langle \mathbf{q}' | e^{-\tau \hat{V}} | \mathbf{q} \rangle = \delta(\mathbf{q}' - \mathbf{q}) e^{-\tau V(\mathbf{q})}. \tag{4.3}$$

Despite these limitations, such Hamiltonians are general enough to describe many diverse systems of itinerant particles.

The partition function of a system with Hamiltonian $\hat{H}$ at reciprocal temperature $\beta = 1/k_{\mathrm{B}}T$ is

$$Z = \operatorname{Tr} e^{-\beta \hat{H}} = \int \mathrm{d}\mathbf{q} \, \langle \mathbf{q} | e^{-\beta \hat{H}} | \mathbf{q} \rangle \,, \tag{4.4}$$

and the thermal expectation value of an operator $\hat{O}$ is

$$\langle \hat{O} \rangle_{\beta \hat{H}} = \frac{1}{Z} \operatorname{Tr} e^{-\beta \hat{H}} \hat{O} = \frac{\int \mathrm{d}\mathbf{q} \, \langle \mathbf{q} | e^{-\beta \hat{H}} \hat{O} | \mathbf{q} \rangle}{\int \mathrm{d}\mathbf{q} \, \langle \mathbf{q} | e^{-\beta \hat{H}} | \mathbf{q} \rangle}. \tag{4.5}$$

As a means of evaluating $\langle \hat{O} \rangle_{\beta \hat{H}}$, we may construct a discretized imaginary time path integral for the partition function. To that end, we first rename $\mathbf{q}$ to $\mathbf{Q}^{(1)}$ and then insert $P - 1$ resolutions of the identity

$$\hat{\mathbb{1}} = \int \mathrm{d}\mathbf{Q}^{(j)} \, |\mathbf{Q}^{(j)}\rangle\langle\mathbf{Q}^{(j)}| \,, \tag{4.6}$$

which introduce the additional Cartesian coordinates $\mathbf{Q}^{(2)}$, …, $\mathbf{Q}^{(P)}$ along the imaginary time path; we combine them all into the vector $\mathbf{Q}$ and refer to them as "beads", picturing the path as a necklace. This results in

$$Z = \int \mathrm{d}\mathbf{Q} \prod_{j=1}^{P} \langle \mathbf{Q}^{(j)} | e^{-\frac{\beta}{P} \hat{H}} | \mathbf{Q}^{(j+1)} \rangle \,, \tag{4.7}$$

where it should be understood that the path is cyclic in imaginary time (that is, $\mathbf{Q}^{(P+1)}$ is an alias for $\mathbf{Q}^{(1)}$).

To evaluate each high-temperature propagator, since $[\hat{K}, \hat{V}] \neq 0$, we rely on the Trotter factorization

$$e^{-\beta \hat{H}} = \lim_{P \to \infty} \left(e^{-\frac{\beta}{P} \hat{K}} e^{-\frac{\beta}{P} \hat{V}} \right)^P, \tag{4.8}$$

which allows us to start with the approximation

$$\langle \mathbf{q}' | e^{-\frac{\beta}{P} \hat{H}} | \mathbf{q} \rangle \approx \sqrt{\frac{|\mathbf{M}| P^f}{(2\pi\hbar^2\beta)^f}} \, e^{-\frac{P}{2\hbar^2\beta} (\mathbf{q}' - \mathbf{q}) \cdot \mathbf{M} \cdot (\mathbf{q}' - \mathbf{q}) - \frac{\beta}{P} V(\mathbf{q})} \tag{4.9}$$

and systematically improve the error in the product of these approximate factors by increasing P. For any finite P, we may construct the approximate path density

$$\pi(\mathbf{Q}) = \left(\frac{|\mathbf{M}|P^f}{(2\pi\hbar^2\beta)^f}\right)^{\frac{P}{2}} e^{-\beta V_{\mathrm{cl}}(\mathbf{Q})} \tag{4.10}$$

with the classical potential

$$V_{\mathrm{cl}}(\mathbf{Q}) = \sum_{i=1}^{f} \frac{m_i P}{2\hbar^2\beta^2} \sum_{j=1}^{P} \left(Q_i^{(j)} - Q_i^{(j+1)}\right)^2 + \frac{1}{P}\sum_{j=1}^{P} V(\mathbf{Q}^{(j)}), \tag{4.11}$$

so that

$$Z = \lim_{P\to\infty} \int \mathrm{d}\mathbf{Q}\ \pi(\mathbf{Q}). \tag{4.12}$$

For the remainder of this article, we drop the $P \to \infty$ limit for the sake of brevity.

In order to use PIMC sampling to calculate $\langle \hat{O} \rangle_{\beta\hat{H}}$, it is necessary to procure an estimator function $\mathcal{E}_{\hat{O}}(\mathbf{Q})$, the details of which depend on the nature of the operator. The operator expression in Eq. (4.5) is then replaced by a ratio of integrals containing only regular functions:

$$\langle \hat{O} \rangle_{\beta\hat{H}} = \langle \mathcal{E}_{\hat{O}} \rangle_{\pi} = \frac{\int \mathrm{d}\mathbf{Q}\ \pi(\mathbf{Q})\mathcal{E}_{\hat{O}}(\mathbf{Q})}{\int \mathrm{d}\mathbf{Q}\ \pi(\mathbf{Q})}. \tag{4.13}$$

This ratio is commonly evaluated as

$$\langle \mathcal{E}_{\hat{O}} \rangle_{\pi} \approx \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} \mathcal{E}_{\hat{O}}(\mathbf{Q}_{[i]}) \tag{4.14}$$

by drawing the samples $\{\mathbf{Q}_{[i]}\}_{i=1}^{N_{\mathrm{MC}}}$ from $\pi(\mathbf{Q})$ using Markov chain Monte Carlo.

4.3 Estimators

It is generally more convenient to work with path integrals in Cartesian coordinates $\mathbf{q}$, but the PMF $A(\xi^\star)$ is expressed in terms of an arbitrary curvilinear coordinate ξ at some value $\xi^\star$. To connect the two, we introduce an invertible coordinate transformation to the generalized coordinates X_1, X_2, …, X_{f-1}, ξ, where the first $f-1$ of these are grouped into the vector $\mathbf{X}$. This transformation has non-zero Jacobian determinant $J(\mathbf{q}) = J(\mathbf{X}, \xi)$. The special coordinate ξ is referred to as the *reaction coordinate*; for example, it may be the distance between two specific centers of mass in a cluster.

Using the diagonal reduced density

$$\varrho(\xi^\star) = \frac{1}{Z}\,\langle\xi^\star|\mathrm{Tr}_{\mathbf{X}}\, e^{-\beta\hat{H}}|\xi^\star\rangle \tag{4.15a}$$

$$= \frac{1}{Z}\int \mathrm{d}\mathbf{X}\ \langle \mathbf{X}\,\xi^\star|e^{-\beta\hat{H}}|\mathbf{X}\,\xi^\star\rangle \tag{4.15b}$$

at reciprocal temperature β, we may construct the overall object of interest: the potential of mean force

$$A(\xi^\star) = -\frac{1}{\beta}\log\frac{\varrho(\xi^\star)}{\varrho_0}, \tag{4.16}$$

where ϱ_0 is an arbitrary constant with the same physical dimension as ξ^{-1}. Choosing a value for ϱ_0 sets the zero of energy for the PMF. Note that our definitions imply that $\varrho_0^{-1} = \int \mathrm{d}\xi^\star\, e^{-\beta A(\xi^\star)}$, which does not contain any explicit volume element factors; instead, we encounter a geometric term in the estimators. Even though the momentum operator $\hat{p}_\xi$ conjugate to the reaction coordinate operator $\hat{\xi}$ satisfies[128]

$$\langle\xi|\,\hat{p}_\xi = -i\hbar\frac{\partial}{\partial\xi}\,\langle\xi|\,, \tag{4.17}$$

in general we find that

$$\langle\mathbf{q}|\,\hat{p}_\xi \neq -i\hbar\frac{\partial}{\partial\xi}\,\langle\mathbf{q}|\,, \tag{4.18}$$

and the missing portion is directly responsible for the geometric term.

We wish to compute $A(\xi^\star)$ via its derivative

$$A'(\xi^\star) = -\frac{1}{\beta}\frac{\mathrm{d}}{\mathrm{d}\xi^\star}\log\frac{\varrho(\xi^\star)}{\varrho_0} = -\frac{1}{\beta}\frac{\varrho'(\xi^\star)}{\varrho(\xi^\star)}. \tag{4.19}$$

As shown in Appendix 4.A, we may write the diagonal reduced density in Cartesian coordinates as

$$\varrho(\xi^\star) = \frac{1}{Z}\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star)\,\langle\mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle\,, \tag{4.20}$$

so

$$-\beta A'(\xi^\star) = \frac{\frac{\mathrm{d}}{\mathrm{d}\xi^\star}\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star)\,\langle\mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle}{\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star)\,\langle\mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle}. \tag{4.21}$$

Because the denominator resembles a constrained version of the partition function Z in Eq. (4.4), we use this as the starting point to derive two path integral estimators $\mathcal{E}_1(\mathbf{Q})$ and $\mathcal{E}_2(\mathbf{Q})$ which

satisfy

$$-\beta A'(\xi^\star) = \langle \mathcal{E}_i \rangle_{\pi,\xi^\star} = \frac{\int \mathrm{d}\mathbf{Q}\ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big)\pi(\mathbf{Q})\mathcal{E}_i(\mathbf{Q})}{\int \mathrm{d}\mathbf{Q}\ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big)\pi(\mathbf{Q})}. \tag{4.22}$$

For $\mathcal{E}_1(\mathbf{Q})$, we first differentiate and then discretize the path integral in the numerator, and for $\mathcal{E}_2(\mathbf{Q})$ we do the reverse.

4.3.1 Estimator 1: Differentiate then discretize

The main quantity in question is

$$Z\varrho'(\xi^\star) = \frac{\mathrm{d}}{\mathrm{d}\xi^\star} \int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star)\, \langle \mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle\,, \tag{4.23}$$

which we write using Appendix 4.B as

$$Z\varrho'(\xi^\star) = \int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star) \left[J_\xi(\mathbf{q}) + \frac{\partial}{\partial \xi}\right] \langle \mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle\,. \tag{4.24}$$

The simpler of the two terms is the geometric one, which stems from the coordinate transformation:

$$\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star)\, \langle \mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle\, J_\xi(\mathbf{q}). \tag{4.25}$$

The remaining term

$$\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star) \sum_{i=1}^{f} \frac{\partial q_i}{\partial \xi} G_i(\mathbf{q}) \tag{4.26}$$

is more involved, requiring the derivatives of the imaginary time propagator:

$$G_i(\mathbf{q}) = \frac{\partial}{\partial q_i} \langle \mathbf{q}|e^{-\beta\hat{H}}|\mathbf{q}\rangle = \frac{1}{i\hbar} \langle \mathbf{q}| \Big[e^{-\beta\hat{H}}, \hat{p}_i\Big] |\mathbf{q}\rangle\,, \tag{4.27}$$

where the derivative–commutator identity is derived in Appendix 4.C.

Using the Kubo formula for the commutator with the exponential of an operator,[129] we find that

$$G_i(\mathbf{q}) = -\frac{1}{i\hbar} \int_0^\beta \mathrm{d}\lambda\ \langle \mathbf{q}|e^{-(\beta-\lambda)\hat{H}} [\hat{H}, \hat{p}_i]\, e^{-\lambda\hat{H}}|\mathbf{q}\rangle\,. \tag{4.28}$$

Since the kinetic energy operator in Eq. (4.1) commutes with $\hat{p}_i$, only the commutator with the

potential energy remains: $[\hat{V}, \hat{p}_i]$. It follows from $[\hat{q}_i, \hat{p}_j] = i\hbar\delta_{ij}$ that

$$[V(\hat{\mathbf{q}}), \hat{p}_i] = -i\hbar F_i(\hat{\mathbf{q}}), \tag{4.29}$$

where the component of the force vector $\mathbf{F}(\mathbf{q})$ on the coordinate q_i is given by

$$F_i(\mathbf{q}) = -\frac{\partial}{\partial q_i} V(\mathbf{q}). \tag{4.30}$$

Hence,

$$G_i(\mathbf{q}) = \int_0^\beta \mathrm{d}\lambda \ \langle \mathbf{q}| e^{-(\beta-\lambda)\hat{H}} F_i(\hat{\mathbf{q}}) e^{-\lambda\hat{H}} |\mathbf{q}\rangle \,. \tag{4.31}$$

Having obtained the necessary expressions, we may discretize the path integral in the usual fashion. The geometric term in Eq. (4.25) poses no difficulty, and we get

$$\int \mathrm{d}\mathbf{Q} \ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big) \pi(\mathbf{Q}) J_\xi(\mathbf{Q}^{(1)}). \tag{4.32}$$

The integral from the Kubo formula is discretized into an average over the path, and because $F_i(\hat{\mathbf{q}})$ is diagonal in the additional path coordinates, we only need to perform the substitution

$$G_i(\mathbf{q}) \to \pi(\mathbf{Q}) \frac{\beta}{P} \sum_{j=1}^{P} F_i(\mathbf{Q}^{(j)}), \tag{4.33}$$

turning Eq. (4.26) into

$$\int \mathrm{d}\mathbf{Q} \ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big) \pi(\mathbf{Q}) \frac{\beta}{P} \sum_{j=1}^{P} \mathbf{F}(\mathbf{Q}^{(j)}) \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial \xi}. \tag{4.34}$$

Thus, the PMF derivative may be written as

$$-\beta A'(\xi^\star) = \langle \mathcal{E}_1 \rangle_{\pi,\xi^\star} = \frac{\int \mathrm{d}\mathbf{Q} \ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big) \pi(\mathbf{Q}) \mathcal{E}_1(\mathbf{Q})}{\int \mathrm{d}\mathbf{Q} \ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big) \pi(\mathbf{Q})}, \tag{4.35}$$

where

$$\mathcal{E}_1(\mathbf{Q}) = \frac{\partial}{\partial \xi} \log |J(\mathbf{Q}^{(1)})| \ + \frac{\beta}{P} \sum_{j=1}^{P} \mathbf{F}(\mathbf{Q}^{(j)}) \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial \xi} \tag{4.36}$$

is the first estimator. In the $P = 1$ case, it reduces to a form recognizable from classical mechanics:[115]

$$-\frac{1}{\beta}\mathcal{E}_1(\mathbf{q}) = -\frac{1}{\beta}\frac{\partial}{\partial\xi}\log|J(\mathbf{q})| - \mathbf{F}(\mathbf{q})\cdot\frac{\partial\mathbf{q}}{\partial\xi} \tag{4.37a}$$

$$= \frac{\partial}{\partial\xi}\left[V(\mathbf{q}) - \frac{1}{\beta}\log|J(\mathbf{q})|\right]. \tag{4.37b}$$

4.3.2 Estimator 2: Discretize then differentiate

To derive another estimator, we start from Eq. (4.24), but first discretize the path into P imaginary time steps to find

$$Z\varrho'(\xi^\star) = \int \mathrm{d}\mathbf{Q}\ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big)\left[J_\xi(\mathbf{Q}^{(1)}) + \frac{\partial\mathbf{Q}^{(1)}}{\partial\xi}\cdot\frac{\partial}{\partial\mathbf{Q}^{(1)}}\right]\pi(\mathbf{Q}). \tag{4.38}$$

The geometric term will again be as in Eq. (4.32), but the other term is now straightforward to compute via ordinary calculus, requiring only

$$\frac{1}{\pi(\mathbf{Q})}\frac{\partial\pi(\mathbf{Q})}{\partial\mathbf{Q}^{(j)}} = -\beta\frac{\partial V_{\mathrm{cl}}(\mathbf{Q})}{\partial\mathbf{Q}^{(j)}} = \beta\mathbf{F}_{\mathrm{cl}}^{(j)}(\mathbf{Q}) = \frac{\beta}{P}\mathbf{F}(\mathbf{Q}^{(j)}) - \frac{P}{\hbar^2\beta}\mathbf{M}\cdot\left[2\mathbf{Q}^{(j)} - \mathbf{Q}^{(j+1)} - \mathbf{Q}^{(j-1)}\right], \tag{4.39}$$

in which the classical force $\mathbf{F}_{\mathrm{cl}}^{(j)}(\mathbf{Q})$ on bead j is obtained from the classical potential. The PMF derivative may therefore also be written as

$$-\beta A'(\xi^\star) = \langle\mathcal{E}_2\rangle_{\pi,\xi^\star} = \frac{\int\mathrm{d}\mathbf{Q}\ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big)\pi(\mathbf{Q})\mathcal{E}_2(\mathbf{Q})}{\int\mathrm{d}\mathbf{Q}\ \delta\Big(\xi(\mathbf{Q}^{(1)}) - \xi^\star\Big)\pi(\mathbf{Q})}, \tag{4.40}$$

where

$$\mathcal{E}_2(\mathbf{Q}) = \frac{\partial}{\partial\xi}\log|J(\mathbf{Q}^{(1)})| + \beta\mathbf{F}_{\mathrm{cl}}^{(1)}(\mathbf{Q})\cdot\frac{\partial\mathbf{Q}^{(1)}}{\partial\xi} \tag{4.41}$$

is the second estimator.

It is perhaps a little surprising that the sum $\frac{\beta}{P}\sum_{j=2}^{P}\mathbf{F}(\mathbf{Q}^{(j)})$ from $\mathcal{E}_1$, which involves all coordinates except the constrained one, appears to have been replaced by $-\frac{P}{\hbar^2\beta}\mathbf{M}\cdot\Big(2\mathbf{Q}^{(1)} - \mathbf{Q}^{(2)} - \mathbf{Q}^{(P)}\Big)$, which depends on only three coordinates. This results in the peculiar identity

$$\left\langle\mathbf{F}_{\mathrm{cl}}^{(1)}(\mathbf{Q})\cdot\frac{\partial\mathbf{Q}^{(1)}}{\partial\xi}\right\rangle_{\pi,\xi^\star} \overset{P\to\infty}{=} \left\langle\frac{1}{P}\sum_{j=1}^{P}\mathbf{F}(\mathbf{Q}^{(j)})\cdot\frac{\partial\mathbf{Q}^{(1)}}{\partial\xi}\right\rangle_{\pi,\xi^\star}, \tag{4.42}$$

which relates the classical force on the constrained coordinates to the average force over the path.

4.3.3 Removal of geometric term

It is occasionally more convenient to work with

$$\tilde{\varrho}(\xi^\star) = \frac{\varrho(\xi^\star)}{f(\xi^\star)}, \tag{4.43}$$

for some function f, than with $\varrho(\xi^\star)$ directly. Consider, for example, the spherical coordinates (ξ, $\cos\theta$, φ), which have Jacobian determinant

$$J(\xi, \cos\theta, \varphi) = -\xi^2. \tag{4.44}$$

The normalization

$$1 = \int \mathrm{d}\xi^\star \, (\xi^\star)^2 \tilde{\varrho}(\xi^\star) \tag{4.45}$$

is often more natural than

$$1 = \int \mathrm{d}\xi^\star \, \varrho(\xi^\star). \tag{4.46}$$

Using $\tilde{\varrho}$ leads to the modified PMF

$$\tilde{A}(\xi^\star) = -\frac{1}{\beta} \log \frac{\tilde{\varrho}(\xi^\star)}{\tilde{\varrho}_0} = A(\xi^\star) + \frac{1}{\beta} \log \frac{\tilde{\varrho}_0 f(\xi^\star)}{\varrho_0}, \tag{4.47}$$

where the arbitrary constant $\tilde{\varrho}_0$ has the same physical dimension as $(\xi f(\xi))^{-1}$. The corresponding derivative is

$$-\beta \tilde{A}'(\xi^\star) = -\beta A'(\xi^\star) - \frac{\partial}{\partial \xi} \log f(\xi^\star), \tag{4.48}$$

and it follows immediately that

$$\tilde{\mathcal{E}}_1(\mathbf{Q}) = \mathcal{E}_1(\mathbf{Q}) - \frac{\partial}{\partial \xi} \log f(\xi^\star) \tag{4.49a}$$

and

$$\tilde{\mathcal{E}}_2(\mathbf{Q}) = \mathcal{E}_2(\mathbf{Q}) - \frac{\partial}{\partial \xi} \log f(\xi^\star) \tag{4.49b}$$

may be used to estimate $-\beta \tilde{A}'(\xi^\star)$.

Whenever a transformation from $\mathbf{q}$ to $\mathbf{X}, \xi$ exists with Jacobian determinant $J(\mathbf{X}, \xi)$ that is a function of only ξ (as in the above spherical coordinates example) the geometric term may be exactly cancelled from $\mathcal{E}_1$ and $\mathcal{E}_2$ by setting $f(\xi) = |J(\xi)|$. In such a situation,

$$\frac{\partial}{\partial \xi} \log f(\xi^\star) = \frac{\partial}{\partial \xi} \log |J(\mathbf{Q}^{(1)})| \,, \tag{4.50}$$

so we are left with just

$$\tilde{\mathcal{E}}_1(\mathbf{Q}) = \frac{\beta}{P} \sum_{j=1}^{P} \mathbf{F}(\mathbf{Q}^{(j)}) \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial \xi} \tag{4.51a}$$

and

$$\tilde{\mathcal{E}}_2(\mathbf{Q}) = \beta \mathbf{F}_{\mathrm{cl}}^{(1)}(\mathbf{Q}) \cdot \frac{\partial \mathbf{Q}^{(1)}}{\partial \xi}. \tag{4.51b}$$

This modification provides no substantial computational benefits, as the omitted expression will be a constant with respect to the integration (for example, $2/\xi^\star$ for spherical coordinates). However, it may make sense to exclude the term from the calculation entirely if it is destined to be excised after the calculation is completed.

4.3.4 Kubo formula in generalized coordinates

Starting from Eq. (4.15), which expresses the diagonal reduced density in terms of the generalized coordinates, application of Appendix 4.C immediately yields

$$Z \varrho'(\xi^\star) = \frac{1}{i\hbar} \int \mathrm{d}\mathbf{X} \, \langle \mathbf{X}\, \xi^\star | \left[e^{-\beta \hat{H}}, \hat{p}_\xi \right] | \mathbf{X}\, \xi^\star \rangle \,. \tag{4.52}$$

This bypasses many of the convoluted steps found above and leaves us with a succinct expression, which takes on the form

$$-\frac{1}{i\hbar} \int \mathrm{d}\mathbf{X} \int_0^\beta \mathrm{d}\lambda \, \langle \mathbf{X}\, \xi^\star | e^{-(\beta-\lambda)\hat{H}} [\hat{H}, \hat{p}_\xi] \, e^{-\lambda \hat{H}} | \mathbf{X}\, \xi^\star \rangle \tag{4.53}$$

after treatment with the Kubo formula.[129] Since

$$[V(\hat{\mathbf{q}}), \hat{p}_\xi] = -i\hbar F_\xi(\hat{\mathbf{q}}) \tag{4.54}$$

involves only the force along the reaction coordinate, proceeding in this direction seems like the obvious choice. However, $\hat{p}_\xi$ is not guaranteed to commute with the Cartesian momenta, and the commutator $[\hat{K}, \hat{p}_\xi]$ is not always diagonal in the position representation. Consequently, Eq. (4.53)

does not lend itself well to discretization and we do not pursue this approach to the derivation further.

4.4 Results

As a proof of concept, we use the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ to compute derivatives of the PMF of two small systems, for which reference results (either exact or numerical) may be calculated: a one-dimensional harmonic oscillator and a Lennard-Jones model of the Ar_2 dimer. To perform the path integral Monte Carlo sampling, we have implemented a basic Markov chain integrator[130] using the Metropolis–Hastings acceptance criterion. The constraint is exactly enforced by sampling in the generalized coordinates for the first bead: updates are proposed for $\mathbf{X}^{(1)}$, but $\xi^{(1)}$ is held fixed at $\xi^\star$.

4.4.1 Harmonic oscillator

The simplest non-trivial problem we can consider is the dependable harmonic oscillator, with the Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{1}{2}m\omega^2\hat{q}^2 \tag{4.55}$$

and the reaction coordinate $\xi = q$. Since the eigenstates of this Hamiltonian are known analytically, we may write down the normalized diagonal density

$$\varrho(\xi^\star) = \frac{1}{Z}e^{-\frac{\beta\hbar\omega}{2}}\frac{\alpha}{\sqrt{\pi}}e^{-(\alpha\xi^\star)^2}\sum_{n=0}^{\infty}\frac{e^{-\beta\hbar\omega n}}{2^n n!}H_n^2(\alpha\xi^\star), \tag{4.56}$$

where $H_n(x)$ is the order-n Hermite polynomial at x, $\alpha = \sqrt{m\omega/\hbar}$, and the partition function is

$$Z = \frac{1}{2}\operatorname{csch}(\beta\hbar\omega/2). \tag{4.57}$$

Using the identity[131]

$$\sum_{n=0}^{\infty}\frac{k^n}{2^n n!}H_n(x)H_n(y) = \frac{e^{\frac{k^2(x^2+y^2)-2kxy}{k^2-1}}}{\sqrt{1-k^2}} \tag{4.58}$$

for $|k| < 1$, which in our case simplifies to

$$\sum_{n=0}^{\infty}\frac{k^n}{2^n n!}H_n^2(x) = \frac{e^{\frac{2kx^2}{k+1}}}{\sqrt{1-k^2}}, \tag{4.59}$$

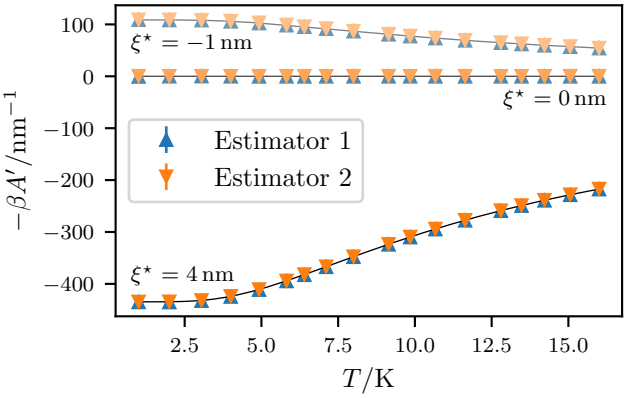

Figure 4.1: Comparison of the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ in Eqs. (4.36) and (4.41) for the computation of the PMF derivative of a harmonic oscillator at $\xi^\star = -1\,\mathrm{nm}$ (top curve, least saturated), $0\,\mathrm{nm}$ (middle curve), and $4\,\mathrm{nm}$ (bottom curve, most saturated). Error bars are not visible, because they are smaller than the symbols. The solid curves show the exact result from Eq. (4.61).

we find that

$$\varrho(\xi^\star) = \sqrt{\frac{\alpha^2 \tanh(\beta\hbar\omega/2)}{\pi}} e^{-\alpha^2 \tanh(\beta\hbar\omega/2)(\xi^\star)^2}. \tag{4.60}$$

Thus,

$$-\beta A'(\xi^\star) = -2\alpha^2 \tanh(\beta\hbar\omega/2)\xi^\star, \tag{4.61}$$

which is proportional to $\xi^\star$.

The necessary quantities for the PIMC estimators are

$$\frac{\partial}{\partial\xi} \log|J(q)| = 0, \tag{4.62a}$$

$$\frac{\partial q}{\partial\xi} = 1, \tag{4.62b}$$

and

$$F(q) = -m\omega^2 q. \tag{4.62c}$$

For this example, we have arbitrarily chosen $m = 1.5\,\mathrm{g\,mol^{-1}}$ and $\omega = 2.3\,\mathrm{ps^{-1}}$. The derivative of the PMF as computed using the Monte Carlo estimators agrees very well with the exact result over a range of temperatures and constraint positions, as shown in Fig. 4.1.

4.4.2 Lennard-Jones dimer

To demonstrate that these estimators are applicable to a curvilinear reaction coordinate, we study a diatomic molecule with reduced mass μ and Lennard-Jones interactions. Without the term for translation of the center of mass, its Hamiltonian is

$$\hat{H} = \frac{\hat{p}_{\mathbf{q}}^2}{2\mu} + V_{\mathrm{LJ}}(\hat{\xi}), \tag{4.63}$$

where $\mathbf{q}$ is the radial separation vector between the atoms, whose magnitude $\xi = |\mathbf{q}|$ we use as the reaction coordinate, and

$$V_{\mathrm{LJ}}(\xi) = 4\varepsilon \left[\left(\frac{\sigma}{\xi}\right)^{12} - \left(\frac{\sigma}{\xi}\right)^{6} \right] \tag{4.64}$$

is the Lennard-Jones potential. Unlike the harmonic oscillator example, this system has a potential that vanishes at large separation, allowing the dimer to dissociate.

Expressing $\mathbf{q}$ in spherical coordinates $(\xi, \cos\theta, \varphi)$, we have that the magnitude of the Jacobian determinant is

$$|J(\mathbf{X}, \xi)| = \xi^2. \tag{4.65}$$

In order to evaluate the PIMC estimators, we therefore require the following quantities:

$$\frac{\partial}{\partial \xi} \log |J(\mathbf{q})| = \frac{2}{\xi}, \tag{4.66a}$$

$$\frac{\partial \mathbf{q}}{\partial \xi} = \frac{\mathbf{q}}{\xi}, \tag{4.66b}$$

and

$$\mathbf{F}(\mathbf{q}) = 24\varepsilon \frac{\mathbf{q}}{\xi^2} \left[2\left(\frac{\sigma}{\xi}\right)^{12} - \left(\frac{\sigma}{\xi}\right)^{6} \right]. \tag{4.66c}$$

Note that we retain the geometric term during the simulation and explicitly remove it in the subsequent numerical integration. To perform a reference calculation, we use numerical matrix multiplication (NMM), as described in Appendix 4.D.

For the Lennard-Jones parameters provided in Ref. [132] for Ar_2 ($\varepsilon = 119.8\,\mathrm{K}$ and $\sigma = 3.405\,\text{Å}$), the results in Fig. 4.2 confirm that the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ function correctly with radial distance as a reaction coordinate. In particular, the rapid change in the slope of the PMF is captured at the lower temperatures.

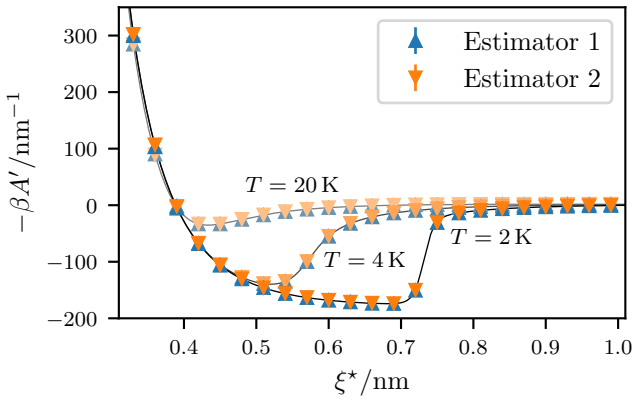

Figure 4.2: Comparison of the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ in Eqs. (4.36) and (4.41) for the computation of the PMF derivative of a Lennard-Jones dimer at $T = 20\,\mathrm{K}$ (top curve, least saturated), $4\,\mathrm{K}$ (middle curve), and $2\,\mathrm{K}$ (bottom curve, most saturated). Error bars are not visible, because they are smaller than the symbols. The solid curves show the NMM results.

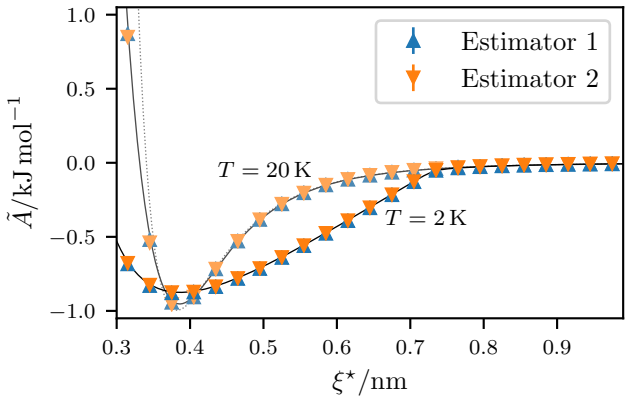

Figure 4.3: Comparison of the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ in Eqs. (4.36) and (4.41) for the computation of the PMF of a Lennard-Jones dimer at $T = 20\,\mathrm{K}$ (narrow curve, least saturated) and $2\,\mathrm{K}$ (wide curve, most saturated). Error bars are not visible, because they are smaller than the symbols. Additional points extending to $\xi_0^\star = 1.5\,\mathrm{nm}$ are not displayed. The solid curves show the NMM results, while the dotted curve is the Lennard-Jones potential in Eq. (4.64).

It is possible to numerically integrate the derivative $A'(\xi^\star)$ to recover the PMF $A(\xi^\star)$. We do so using the midpoint rule on a grid of points $\xi_i^\star$ with spacing $\Delta\xi^\star$ (as in Fig. 4.2), and with $\xi_1^\star$ placed at the largest value of $\xi^\star$. We also define the virtual point $\xi_0^\star = \xi_1^\star + \Delta\xi^\star$ and a shifted grid of points

$$\bar{\xi}_i^\star = \xi_i^\star - \frac{\Delta\xi^\star}{2}, \tag{4.67}$$

with $\bar{\xi}_0^\star$ acting as a "point at infinity" (the dimer is considered to have dissociated when the atoms are at least $\bar{\xi}_0^\star$ apart). Correspondingly, we set $\tilde{A}(\bar{\xi}_0^\star) = 0 = \tilde{A}'(\xi_0^\star)$, using the normalization in Eq. (4.45).

In Fig. 4.3, we show

$$\tilde{A}(\bar{\xi}_j^\star) = \frac{\Delta\xi^\star}{\beta} \sum_{i=1}^{j} \left[-\beta A'(\xi_i^\star) - \frac{2}{\xi_i^\star} \right], \tag{4.68}$$

which is the renormalized PMF with the desired energy offset. The matching NMM curves are calculated from $\varrho(\xi^\star)$ as

$$\tilde{A}(\xi^\star) = -\frac{1}{\beta} \log \frac{\varrho(\xi^\star)(\bar{\xi}_0^\star)^2}{\varrho(\bar{\xi}_0^\star)(\xi^\star)^2} \tag{4.69}$$

to ensure a compatible energy offset. Even though the integration grid is rather sparse, especially

where the slope of the PMF changes suddenly for $T = 2\,\mathrm{K}$, the obtained PMFs are consistent with the reference results.

4.5 Conclusions

We have described two path integral Monte Carlo estimators for the calculation of the derivative of the quantum mechanical potential of mean force. Notably, the curves obtained from these estimators are in terms of the true quantum reaction coordinate, unlike other methods that utilize the path centroid.

The first estimator, Eq. (4.36), was obtained by initially differentiating the exact path integral and then discretizing the resulting path integral into imaginary time steps. Alternatively, the second estimator, Eq. (4.41), was obtained by discretizing the exact path integral first and then performing the differentiation after. In principle, these should be equivalent operations in the $P \to \infty$ limit, and we have demonstrated that both estimators reproduce the correct derivative of the PMF for the one-dimensional harmonic oscillator and Lennard-Jones dimer. In contrast to existing histogram-based methods for the evaluation of free energies, these novel estimators can be used to ascertain information about the free energy profile at just a single point along the reaction coordinate.

Furthermore, it is possible to numerically integrate the computed derivatives evaluated from these estimators to obtain the PMF itself. As shown in the argon dimer example, even when the integration grid is not very dense, this method successfully reproduces the known PMF obtained from numerical matrix multiplication.

In Paper II of this series, we show how these estimators may be used with path integral molecular dynamics. This is achieved by applying techniques from constrained Langevin dynamics to the PILE integrator in order to constrain one of the beads. The extension of these estimators to path integral molecular dynamics simulations will allow for their application to more general systems and potentials, such as small water clusters.

Acknowledgments

We thank Raymond Kapral for providing the initial direction for the derivation of the discretized constrained path integral. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2016-04403), the Ontario Ministry of Research and Innovation (MRI), the Canada Research Chair program (950-231024), and the Canada Foundation for Innovation (CFI) (project No. 35232).

4.A Kets in curvilinear coordinates

A wavefunction $\psi(\mathbf{q})$ may be thought of as the concrete manifestation of an abstract ket $|\psi\rangle$ in a continuous representation:

$$\psi(\mathbf{q}) = \langle \mathbf{q}|\psi\rangle \,. \tag{4.70}$$

Although the object $|\mathbf{q}\rangle$ (which represents a state with definite Cartesian position $\mathbf{q}$) is not an element of Hilbert space, it is common to formally treat it as if it were. Given a change of variables from $\mathbf{q}$ to $\mathbf{X}$, ξ with Jacobian determinant $J(\mathbf{q}) = J(\mathbf{X}, \xi)$, it is useful to define $|\mathbf{X}\,\xi\rangle$ in a way that fulfills

$$\int \mathrm{d}\mathbf{q}\; |\langle \mathbf{q}|\psi\rangle|^2 = \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\; |\langle \mathbf{X}\,\xi|\psi\rangle|^2, \tag{4.71}$$

which is analogous to the statement that the resolution of the identity

$$\hat{\mathbb{1}} = \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\; |\mathbf{X}\,\xi\rangle\langle \mathbf{X}\,\xi| \tag{4.72}$$

should have the usual form, even in curvilinear coordinates.

Since

$$\int \mathrm{d}\mathbf{q}\; |\langle \mathbf{q}|\psi\rangle|^2 = \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\; |J(\mathbf{X}, \xi)|\, |\langle \mathbf{q}(\mathbf{X}, \xi)|\psi\rangle|^2, \tag{4.73}$$

it follows that the definition

$$|\mathbf{X}\,\xi\rangle = \sqrt{|J(\mathbf{X}, \xi)|}\, |\mathbf{q}(\mathbf{X}, \xi)\rangle = \sqrt{|J(\mathbf{q})|}\, |\mathbf{q}\rangle \tag{4.74}$$

is sufficient. This is the approach described in Ref. [128], and the one we use in the present work. Using this definition, we see that the diagonal matrix elements of the partial trace of an operator $\hat{O}$ with respect to $\mathbf{X}$ may be expressed as

$$\begin{aligned}
\langle \xi^\star|\mathrm{Tr}_{\mathbf{X}}\, \hat{O}|\xi^\star\rangle &= \int \mathrm{d}\mathbf{X}\; \langle \mathbf{X}\,\xi^\star|\hat{O}|\mathbf{X}\,\xi^\star\rangle && \text{(4.75a)}\\
&= \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\; \delta(\xi - \xi^\star)\, \langle \mathbf{X}\,\xi|\hat{O}|\mathbf{X}\,\xi\rangle && \text{(4.75b)}\\
&= \int \mathrm{d}\mathbf{q}\; \delta(\xi(\mathbf{q}) - \xi^\star) \frac{\langle \mathbf{X}\,\xi|\hat{O}|\mathbf{X}\,\xi\rangle}{|J(\mathbf{q})|} && \text{(4.75c)}\\
&= \int \mathrm{d}\mathbf{q}\; \delta(\xi(\mathbf{q}) - \xi^\star)\, \langle \mathbf{q}|\hat{O}|\mathbf{q}\rangle && \text{(4.75d)}
\end{aligned}$$

in Cartesian coordinates.

4.B Derivative of a Dirac delta function integral

We wish to take the derivative

$$D(\xi^\star) = \frac{\mathrm{d}}{\mathrm{d}\xi^\star} \int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star) f(\mathbf{q}). \tag{4.76}$$

We first obtain the one-dimensional result

$$\frac{\mathrm{d}}{\mathrm{d}\xi^\star} \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) f(\xi) = \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) \frac{\mathrm{d}}{\mathrm{d}\xi} f(\xi) \tag{4.77}$$

by noting that

$$\frac{\mathrm{d}}{\mathrm{d}\xi^\star} f(\xi^\star) = \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) \frac{\mathrm{d}}{\mathrm{d}\xi} f(\xi). \tag{4.78}$$

For the general case, we change coordinates to those in which ξ appears explicitly:

$$D(\xi^\star) = \int \mathrm{d}\mathbf{X} \frac{\mathrm{d}}{\mathrm{d}\xi^\star} \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) |J(\mathbf{X}, \xi)| f(\mathbf{X}, \xi) \tag{4.79a}$$

$$= \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) \frac{\partial}{\partial \xi} |J(\mathbf{X}, \xi)| f(\mathbf{X}, \xi) \tag{4.79b}$$

$$= \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) \left[\frac{\partial}{\partial \xi} |J(\mathbf{X}, \xi)| \right] f(\mathbf{X}, \xi) + \int \mathrm{d}\mathbf{X} \int \mathrm{d}\xi\ \delta(\xi - \xi^\star) |J(\mathbf{X}, \xi)| \frac{\partial}{\partial \xi} f(\mathbf{X}, \xi) \tag{4.79c}$$

$$= \int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}) - \xi^\star) \left[J_\xi(\mathbf{q}) + \frac{\partial}{\partial \xi} \right] f(\mathbf{q}), \tag{4.79d}$$

where

$$J_\xi(\mathbf{q}) = \frac{\partial}{\partial \xi} \log |J(\mathbf{q})| = \frac{\frac{\partial}{\partial \xi} |J(\mathbf{X}, \xi)|}{|J(\mathbf{X}, \xi)|}, \tag{4.80}$$

and we formally apply the logarithmic derivative notation even when the function is not dimensionless.

4.C Derivative–commutator identity for diagonal matrix elements

It is well-known that momentum operators lead to differentiation in the position representation. For example,

$$\langle \mathbf{q} | \hat{p}_i \hat{A} | \mathbf{q}' \rangle = -i\hbar \frac{\partial}{\partial q_i} \langle \mathbf{q} | \hat{A} | \mathbf{q}' \rangle \tag{4.81}$$

for an arbitrary operator $\hat{A}$, where $\hat{p}_i$ is the momentum operator conjugate to $\hat{q}_i$. However, this relationship does not generally hold when $\mathbf{q}$ and $\mathbf{q}'$ are the same variable:

$$\langle \mathbf{q}|\hat{p}_i\hat{A}|\mathbf{q}\rangle \neq -i\hbar\frac{\partial}{\partial q_i}\langle \mathbf{q}|\hat{A}|\mathbf{q}\rangle\,. \tag{4.82}$$

Instead, for a Hermitian operator $\hat{A}$ with the eigenvalue equation $\hat{A}\,|a\rangle = a\,|a\rangle$, we have that

$$\langle \mathbf{q}|\hat{p}_i\hat{A}|\mathbf{q}\rangle = \sum_a \langle \mathbf{q}|\hat{p}_i|a\rangle\ \langle a|\hat{A}|\mathbf{q}\rangle \tag{4.83a}$$

$$= -i\hbar\sum_a a\left[\frac{\partial}{\partial q_i}\langle \mathbf{q}|a\rangle\right]\langle a|\mathbf{q}\rangle \tag{4.83b}$$

and

$$\langle \mathbf{q}|\hat{A}\hat{p}_i|\mathbf{q}\rangle = i\hbar\sum_a a\,\langle \mathbf{q}|a\rangle\left[\frac{\partial}{\partial q_i}\langle a|\mathbf{q}\rangle\right]. \tag{4.84}$$

Thus, we conclude that

$$\frac{\partial}{\partial q_i}\langle \mathbf{q}|\hat{A}|\mathbf{q}\rangle = \frac{\partial}{\partial q_i}\sum_a \langle \mathbf{q}|\hat{A}|a\rangle\,\langle a|\mathbf{q}\rangle \tag{4.85a}$$

$$= \sum_a a\,\langle \mathbf{q}|a\rangle\left[\frac{\partial}{\partial q_i}\langle a|\mathbf{q}\rangle\right] + \sum_a a\left[\frac{\partial}{\partial q_i}\langle \mathbf{q}|a\rangle\right]\langle a|\mathbf{q}\rangle \tag{4.85b}$$

$$= \frac{1}{i\hbar}\langle \mathbf{q}|\hat{A}\hat{p}_i|\mathbf{q}\rangle - \frac{1}{i\hbar}\langle \mathbf{q}|\hat{p}_i\hat{A}|\mathbf{q}\rangle \tag{4.85c}$$

$$= \frac{1}{i\hbar}\langle \mathbf{q}|\left[\hat{A},\hat{p}_i\right]|\mathbf{q}\rangle\,. \tag{4.85d}$$

4.D Numerical matrix multiplication for a radial coordinate

In Ref. [132], expressions for numerical matrix multiplication of the path integral of a system described by a three-dimensional relative coordinate are given, but not derived. In this section, we briefly explain why the radial propagator has such a curious form.

The operator in the kinetic energy of Eq. (4.63) may be expressed as

$$\hat{p}_{\mathbf{q}}^2 = \hat{p}_\xi^2 + \frac{\hat{\ell}^2}{\hat{\xi}^2}, \tag{4.86}$$

where $\hat{p}_\xi$ is the radial momentum operator, and $\hat{\ell}^2$ is the squared angular momentum operator, whose eigenstates are the spherical harmonics $|\ell\, m\rangle$ with eigenvalues $\hbar^2\ell(\ell+1)$. The radial momentum

operator is not self-adjoint and does not have a spectrum of eigenstates,[133,134] so the spectral theorem does not apply to it and the appropriate resolution of identity is not given by

$$\int \mathrm{d}p_\xi \, |p_\xi\rangle\langle p_\xi| \,, \tag{4.87}$$

despite the wavefunctions

$$\langle \xi | p_\xi \rangle = \frac{e^{\frac{i\xi p_\xi}{\hbar}}}{\sqrt{2\pi\hbar}} \tag{4.88}$$

satisfying $\hat{p}_\xi \, |p_\xi\rangle = p_\xi \, |p_\xi\rangle$. Thus, we must be careful when rederiving Eq. (16c) of Ref. [132].

We turn to the operator $\hat{p}_\xi^2$, which is well-behaved and has the eigenstates

$$\langle \xi | p_\xi^{(2)} \rangle = \frac{e^{\frac{i\xi p_\xi}{\hbar}} - e^{\frac{-i\xi p_\xi}{\hbar}}}{2i\sqrt{\pi\hbar}} = \frac{1}{\sqrt{\pi\hbar}} \sin\frac{\xi p_\xi}{\hbar} \tag{4.89}$$

with eigenvalues p_ξ^2. We may use these states to construct the correct resolution of the identity,

$$\hat{\mathbb{1}} = \int \mathrm{d}p_\xi \; |p_\xi^{(2)}\rangle\langle p_\xi^{(2)}| = \int \mathrm{d}p_\xi \, \Big(\, |p_\xi\rangle\langle p_\xi| - |p_\xi\rangle\langle -p_\xi| \, \Big), \tag{4.90}$$

which, as expected, results in

$$\begin{aligned}
\langle \xi' | e^{-\frac{\tau \hat{p}_\xi^2}{2\mu}} | \xi \rangle &= \int \mathrm{d}p_\xi \; \langle \xi' | e^{-\frac{\tau p_\xi^2}{2\mu}} | p_\xi^{(2)} \rangle \; \langle p_\xi^{(2)} | \xi \rangle && \text{(4.91a)} \\
&= \frac{1}{4\pi\hbar} \int \mathrm{d}p_\xi \, e^{-\frac{\tau p_\xi^2}{2\mu} + \frac{ip_\xi}{\hbar}(\xi'-\xi)} + \frac{1}{4\pi\hbar} \int \mathrm{d}p_\xi \, e^{-\frac{\tau p_\xi^2}{2\mu} - \frac{ip_\xi}{\hbar}(\xi'-\xi)} \\
&\quad - \frac{1}{4\pi\hbar} \int \mathrm{d}p_\xi \, e^{-\frac{\tau p_\xi^2}{2\mu} + \frac{ip_\xi}{\hbar}(\xi'+\xi)} - \frac{1}{4\pi\hbar} \int \mathrm{d}p_\xi \, e^{-\frac{\tau p_\xi^2}{2\mu} - \frac{ip_\xi}{\hbar}(\xi'+\xi)} && \text{(4.91b)} \\
&= \sqrt{\frac{\mu}{2\pi\hbar^2\tau}} \left[e^{-\frac{\mu}{2\hbar^2\tau}(\xi'-\xi)^2} - e^{-\frac{\mu}{2\hbar^2\tau}(\xi'+\xi)^2} \right]. && \text{(4.91c)}
\end{aligned}$$

Chapter 5

On the quantum mechanical potential of mean force. II. Constrained path integral molecular dynamics integrators

This chapter is reproduced from Dmitri Iouchtchenko, Kevin P. Bishop, and Pierre-Nicholas Roy, "On the quantum mechanical potential of mean force. II. Constrained path integral molecular dynamics integrators", arXiv:2101.00762v1.[4]

5.1 Introduction

In Paper I of this series, we demonstrated that the quantum potential of mean force (PMF) may be obtained from constrained path integral Monte Carlo (PIMC) simulations using novel estimators for the derivative of the PMF. However, devising efficient configuration update schemes for Monte Carlo methods is a nontrivial endeavor, as the volume of configuration space grows exponentially both with the number of particles and the number of beads per particle. For most systems, it is not feasible to simply perturb the position of each bead by a random amount, since the magnitude of the perturbations must be decreased substantially when increasing the number of beads, in order to avoid rejecting all proposed configurations. Thus, it is often beneficial to turn to path integral molecular dynamics (PIMD), in which configurations are updated based on the force arising from an effective Hamiltonian. In the unconstrained case, the path integral Langevin equation (PILE) integrator provides a straightforward implementation of thermostatted ring polymer time evolution for PIMD.[16]

In the present work, we incorporate bead-local holonomic constraints into the PILE in order to calculate the quantum PMF of molecular systems. We also show that when the reaction coordinate

for the PMF is the radial separation between two centers of mass, the additional computational effort necessary to enforce the constraint is negligible, as the Lagrange multiplier may be computed directly without resorting to an iterative scheme.

A recent study has found that an approach which combines PIMD with umbrella sampling and histogram unbiasing is sufficient to calculate the quantum PMF of a water dimer.[24] While it may be advantageous to stitch together multiple histograms, as each simulation will contribute data over a range of the reaction coordinate, it can be quite challenging to obtain and converge a collection of smooth histograms. On the other hand, independent values obtained from an estimator at single points are simple to examine and improve in a systematic fashion. Indeed, we observe that when the PMF has a very rapidly changing slope, as in the water dimer at low temperature, it can be favorable to compute its derivative using constrained PIMD and our estimators rather than by numerical differentiation of an unbiased PMF.

The remainder of this article is organized as follows: in Sec. 5.2, we describe our notation; in Sec. 5.3, we develop two integrators for constrained path integral Langevin dynamics; in Sec. 5.4, we derive exact expressions for a special case of the reaction coordinate; in Sec. 5.5, we apply the integrators to a water dimer; and in Sec. 5.6, we summarize our findings.

5.2 Background

As in Paper I, we consider Hamiltonians of the form

$$\hat{H} = \frac{1}{2}\hat{\tilde{\mathbf{p}}} \cdot \tilde{\mathbf{M}}^{-1} \cdot \hat{\tilde{\mathbf{p}}} + V(\hat{\tilde{\mathbf{q}}}) \tag{5.1}$$

(note the addition of tildes to distinguish these quantities from the fictitious ones in the molecular dynamics simulations), and work with the classical potential

$$V_{\mathrm{cl}}(\mathbf{q}) = \sum_{i=1}^{f} \frac{\tilde{m}_i P}{2\hbar^2\tilde{\beta}^2} \sum_{j=1}^{P} \left(q_i^{(j)} - q_i^{(j+1)}\right)^2 + \frac{1}{P}\sum_{j=1}^{P} V(\mathbf{q}^{(j)}), \tag{5.2}$$

which is extracted from the discrete imaginary time path integral of the quantum partition function $Z = \mathrm{Tr}\, e^{-\tilde{\beta}\hat{H}}$ at reciprocal temperature $\tilde{\beta}$ with P beads.

To sample from the unconstrained path density $e^{-\tilde{\beta}V_{\mathrm{cl}}(\mathbf{q})}$ using molecular dynamics, one may associate a fictitious mass $m_i^{(j)}$ and momentum $p_i^{(j)}$ with each Cartesian bead coordinate $q_i^{(j)}$.[15] The masses form the $Pf \times Pf$ diagonal mass matrix $\mathbf{M}$, while the momenta are collected into the vector $\mathbf{p}$ of length Pf. We require that all the fictitious masses for a single degree of freedom be equal ($m_i^{(j)} = m_i$), so that each corresponding block of $\mathbf{M}$ is guaranteed to be invariant under every

similarity transformation. We consider only the typical case when the ratio $\tilde{m}_i/m_i$ is the same for all degrees of freedom; we refer to said ratio as $\tilde{m}/m$ and define the single-bead fictitious mass matrix as

$$\mathbf{M}_1 = \frac{m}{\tilde{m}}\tilde{\mathbf{M}}. \tag{5.3}$$

For a simulation at a fictitious reciprocal temperature β, the momentum partition function

$$\int \mathrm{d}\mathbf{p}\; e^{-\frac{\beta}{2}\mathbf{p}\cdot\mathbf{M}^{-1}\cdot\mathbf{p}} = \left(\frac{2\pi}{\beta}\right)^{\frac{Pf}{2}} |\mathbf{M}|^{\frac{1}{2}} \tag{5.4}$$

may be inserted into configurational averages to express them as phase space averages:

$$\frac{\int \mathrm{d}\mathbf{q}\; e^{-\tilde{\beta}V_{\mathrm{cl}}(\mathbf{q})} f(\mathbf{q})}{\int \mathrm{d}\mathbf{q}\; e^{-\tilde{\beta}V_{\mathrm{cl}}(\mathbf{q})}} = \frac{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q}\; e^{-\beta H_{\mathrm{cl}}(\mathbf{p},\mathbf{q})} f(\mathbf{q})}{\int \mathrm{d}\mathbf{p} \int \mathrm{d}\mathbf{q}\; e^{-\beta H_{\mathrm{cl}}(\mathbf{p},\mathbf{q})}}, \tag{5.5}$$

with the classical Hamiltonian given by

$$H_{\mathrm{cl}}(\mathbf{p},\mathbf{q}) = \frac{1}{2}\mathbf{p}\cdot\mathbf{M}^{-1}\cdot\mathbf{p} + \frac{\tilde{\beta}}{\beta}V_{\mathrm{cl}}(\mathbf{q}). \tag{5.6}$$

Although it appears that we've increased the complexity of the integrals, the phase space formulation of the expectation value is readily evaluated using molecular dynamics techniques.

5.2.1 Path integral Langevin equation (PILE) integrator

The PILE integrator is a combination of a white noise Langevin thermostat with a generalized velocity Verlet scheme that operates on path normal modes.[16] The normal mode transformation is done via an orthogonal matrix C with elements

$$C_{jk} = \begin{cases} \sqrt{\dfrac{1}{P}} & \text{if } k = 0 \\ \sqrt{\dfrac{2}{P}}\cos\dfrac{2\pi jk}{P} & \text{if } 1 \le k < \frac{P}{2} \\ \sqrt{\dfrac{1}{P}}(-1)^j & \text{if } k = \frac{P}{2} \\ \sqrt{\dfrac{2}{P}}\sin\dfrac{2\pi jk}{P} & \text{if } \frac{P}{2} < k \le P-1, \end{cases} \tag{5.7}$$

where j labels the beads (1 to P) and k labels the modes (0 to $P-1$). We use the vectors $\mathbf{P}$ and $\mathbf{Q}$ for the transformed coordinates.

This integrator is composed of three exact sub-integrators. The propagation of harmonic oscillators in normal mode coordinates is

$$\mathrm{A}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk}\,;\ \mathbf{Q}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{q}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow S_k^{\mathbf{PP}}(\Delta t)\mathbf{P}^{(k)} + S_k^{\mathbf{PQ}}(\Delta t)\mathbf{M}_1\mathbf{Q}^{(k)} \\ \bar{\mathbf{Q}}^{(k)} \leftarrow S_k^{\mathbf{QP}}(\Delta t)\mathbf{M}_1^{-1}\mathbf{P}^{(k)} + S_k^{\mathbf{QQ}}(\Delta t)\mathbf{Q}^{(k)} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)}\,;\ \bar{\mathbf{q}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{Q}}^{(k)},\end{cases} \tag{5.8}$$

where we have made explicit the transformations to and from normal modes. The propagation coefficients

$$S_k^{\mathbf{PP}}(\Delta t) = S_k^{\mathbf{QQ}}(\Delta t) = \cos(\omega_k \Delta t), \tag{5.9a}$$

$$S_k^{\mathbf{PQ}}(\Delta t) = -\omega_k \sin(\omega_k \Delta t), \tag{5.9b}$$

and

$$S_k^{\mathbf{QP}}(\Delta t) = \Delta t \,\mathrm{sinc}\,(\omega_k \Delta t) \tag{5.9c}$$

arise from the exact solution of Hamilton's equations of motion for a harmonic oscillator with angular frequency

$$\omega_k = 2\sqrt{\frac{\tilde{m}P}{\hbar^2 m\tilde{\beta}\beta}}\sin\frac{\pi k}{P}. \tag{5.10}$$

Application of the remaining force may be done in Cartesian coordinates:

$$\mathrm{B}\left\{\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \frac{\Delta t\tilde{\beta}}{P\beta}\mathbf{F}(\mathbf{q}^{(j)}),\right. \tag{5.11}$$

where

$$\mathbf{F}(\mathbf{q}^{(j)}) = -\boldsymbol{\nabla} V(\mathbf{q}^{(j)}). \tag{5.12}$$

The normal mode degrees of freedom are independently thermostatted using a Langevin thermostat:

$$\mathrm{O}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow T_k(\Delta t)\mathbf{P}^{(k)} + U_k(\Delta t)\mathbf{M}_1^{\frac{1}{2}}\boldsymbol{\eta} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)},\end{cases} \tag{5.13}$$

where the coefficients are

$$T_k(\Delta t) = e^{-\Delta t \gamma_k} \tag{5.14a}$$

and

$$U_k(\Delta t) = \sqrt{\frac{1}{\beta}(1 - e^{-2\Delta t \gamma_k})}, \tag{5.14b}$$

γ_k is a friction coefficient (the same for all degrees of freedom), and $\boldsymbol{\eta}$ is a vector of f pseudorandom numbers sampled from a standard normal distribution. When $k \geq 1$, the value of γ_k that minimizes the energy autocorrelation time in the free ($V(\tilde{\mathbf{q}}) = 0$) case is known analytically to be $2\omega_k$, and this value is typically used even when interactions are present; the centroid friction γ_0 is a tunable simulation parameter.

If these sub-integrators are referred to as A, B, and O, respectively, then they may be combined in the order OBABO to form the PILE integrator. It is implied by this notation that all the steps other than the central (that is, both repetitions of B and O) have a halved duration of $\Delta t/2$, as required by the symmetric splitting of the Fokker–Planck operator.

5.2.2 Constrained Hamiltonian integrators

The addition of holonomic constraints to a symplectic integrator for Hamilton's equations of motion may be accomplished by a straightforward scheme, in which non-dynamical momentum perturbations ensure that both the position and velocity constraints are satisfied at the end of each step.[135] For example, this may be used to obtain the well-known RATTLE algorithm[136] from velocity Verlet.

In this scheme, a generic integration step of the form

$$\begin{cases} \bar{\mathbf{p}} \leftarrow f_{\mathbf{p}}(\mathbf{p}, \mathbf{q}) \\ \bar{\mathbf{q}} \leftarrow f_{\mathbf{q}}(\mathbf{p}, \mathbf{q}) \end{cases} \tag{5.15}$$

becomes the two-step sequence

$$\begin{cases} \mathbf{p}' \leftarrow \mathbf{p} + \boldsymbol{\nabla\xi}(\mathbf{q}) \cdot \boldsymbol{\Lambda} \\ \bar{\mathbf{p}} \leftarrow f_{\mathbf{p}}(\mathbf{p}', \mathbf{q}) \\ \bar{\mathbf{q}} \leftarrow f_{\mathbf{q}}(\mathbf{p}', \mathbf{q}) \\ \boldsymbol{\xi}(\bar{\mathbf{q}}) = \mathbf{z} \end{cases} \tag{5.16a}$$

$$\begin{cases} \bar{\mathbf{p}} \leftarrow \mathbf{p} + \boldsymbol{\nabla\xi}(\mathbf{q}) \cdot \boldsymbol{\Lambda} \\ \dot{\boldsymbol{\xi}}(\mathbf{q}) = 0, \end{cases} \tag{5.16b}$$

where $\boldsymbol{\xi}(\mathbf{q}) = \mathbf{z}$ is the holonomic constraint to be maintained, $\dot{\boldsymbol{\xi}}(\mathbf{q}) = 0$ is the implicit velocity constraint, and each $\boldsymbol{\Lambda}$ is a vector of Lagrange multipliers that results in the final line of the corresponding step being valid. Note how the first step begins by perturbing the momentum away from the constraint manifold in order to ensure that the position constraint is satisfied, while the second step projects the momentum back onto the constraint manifold.

In principle, the Lagrange multipliers may be found at each step of the simulation by integrating their equations of motion explicitly. However, this will result in a growing discrepancy between the calculated values and the values that are needed to correctly enforce the constraints.[137] Instead, one should solve a system of equations for the Lagrange multipliers at each step; since these equations are generally nonlinear, they are most often solved by iteration.

5.3 Integrators

Our aim is to compute the derivative of the PMF,

$$-\tilde{\beta}A'(\xi^\star) = \frac{\int \mathrm{d}\mathbf{q}\ \delta\big(\xi(\mathbf{q}^{(1)}) - \xi^\star\big) e^{-\tilde{\beta}V_{\mathrm{cl}}(\mathbf{q})} \mathcal{E}_i(\mathbf{q})}{\int \mathrm{d}\mathbf{q}\ \delta(\xi(\mathbf{q}^{(1)}) - \xi^\star) e^{-\tilde{\beta}V_{\mathrm{cl}}(\mathbf{q})}}, \tag{5.17}$$

in a PIMD setting using the two path integral estimators

$$\mathcal{E}_1(\mathbf{q}) = \frac{\partial}{\partial\xi} \log|J(\mathbf{q}^{(1)})| + \frac{\tilde{\beta}}{P} \sum_{j=1}^{P} \mathbf{F}(\mathbf{q}^{(j)}) \cdot \frac{\partial \mathbf{q}^{(1)}}{\partial\xi} \tag{5.18a}$$

and

$$\mathcal{E}_2(\mathbf{q}) = \frac{\partial}{\partial\xi} \log|J(\mathbf{q}^{(1)})| + \tilde{\beta}\mathbf{F}_{\mathrm{cl}}^{(1)}(\mathbf{q}) \cdot \frac{\partial \mathbf{q}^{(1)}}{\partial\xi} \tag{5.18b}$$

from Paper I. Although the molecular dynamics simulations will take place in Cartesian coordinates, it is substantially more convenient to develop the theory using the generalized coordinates ($\mathbf{X}$, ξ, $\mathbf{u}$) which include the reaction coordinate. The transformation to these coordinates has non-zero Jacobian determinant $J(\mathbf{q}) = J(\mathbf{X}, \xi, \mathbf{u})$; since the unconstrained beads are not transformed ($\mathbf{u} = \mathbf{q}^{(2)}, \ldots, \mathbf{q}^{(P)}$), J has no dependence on their values and we write simply $J(\mathbf{X}, \xi)$, noting that it has the same value as the Jacobian determinant of the transformation on just the first bead.

By taking the time derivative of the explicit constraint equation $\xi(\mathbf{q}^{(1)}) = \xi^\star$, we find the implicit velocity constraint $\dot{\xi}(\mathbf{q}^{(1)}) = 0$, which prevents us from using Eq. (5.4) directly to obtain the

necessary phase space integral. Instead, we find that the appropriate expression is

$$\int \mathrm{d}\mathbf{p}_\mathbf{U}\; e^{-\frac{\beta}{2}\mathbf{p}_\mathbf{U}\cdot\mathbf{A}^{-1}\cdot\mathbf{p}_\mathbf{U}} = \left(\frac{2\pi}{\beta}\right)^{\frac{Pf-1}{2}} |\mathbf{A}|^{\frac{1}{2}} = \left(\frac{2\pi}{\beta}\right)^{\frac{Pf-1}{2}} |\mathbf{\Gamma}|^{\frac{1}{2}}\, Z_\xi^{\frac{1}{2}} \tag{5.19a}$$

$$= \left(\frac{2\pi}{\beta}\right)^{\frac{Pf-1}{2}} |\mathbf{M}|^{\frac{1}{2}}\, |J(\mathbf{X},\xi^\star)|\, Z_\xi(\mathbf{X},\xi^\star)^{\frac{1}{2}}, \tag{5.19b}$$

where $\mathbf{p}_\mathbf{U}$ are the momenta conjugate to all the unconstrained bead coordinates $\mathbf{U} = \mathbf{X}, \mathbf{u}$,

$$\mathbf{\Gamma} = \mathbf{J}^\mathrm{T} \cdot \mathbf{M} \cdot \mathbf{J} \tag{5.20}$$

is the generalized mass matrix, $\mathbf{J}$ is the Jacobian matrix (whose determinant is J), $\mathbf{A}$ is $\mathbf{\Gamma}$ without the row and column corresponding to ξ, and

$$Z_\xi = \boldsymbol{\nabla}\xi \cdot \mathbf{M}^{-1} \cdot \boldsymbol{\nabla}\xi = \boldsymbol{\nabla}_1\xi \cdot \mathbf{M}_1^{-1} \cdot \boldsymbol{\nabla}_1\xi \tag{5.21}$$

is the element of $\mathbf{\Gamma}^{-1}$ at the row and column corresponding to ξ. The requisite determinant identity is proved in Appendix 5.A.

Therefore, we have

$$-\tilde{\beta}A'(\xi^\star) = \frac{\int \mathrm{d}\mathbf{p}_\mathbf{U} \int \mathrm{d}\mathbf{U}\; e^{-\beta H_\mathrm{cl}^\mathrm{c}(\mathbf{p}_\mathbf{U},\mathbf{U};\xi^\star)} Z_\xi^{-\frac{1}{2}} \mathcal{E}_i(\mathbf{U},\xi^\star)}{\int \mathrm{d}\mathbf{p}_\mathbf{U} \int \mathrm{d}\mathbf{U}\; e^{-\beta H_\mathrm{cl}^\mathrm{c}(\mathbf{p}_\mathbf{U},\mathbf{U};\xi^\star)} Z_\xi^{-\frac{1}{2}}}, \tag{5.22}$$

where the constrained classical Hamiltonian

$$H_\mathrm{cl}^\mathrm{c}(\mathbf{p}_\mathbf{U},\mathbf{U};\xi^\star) = \frac{1}{2}\mathbf{p}_\mathbf{U}\cdot\mathbf{A}^{-1}\cdot\mathbf{p}_\mathbf{U} + \frac{\tilde{\beta}}{\beta}V_\mathrm{cl}(\mathbf{U},\xi^\star) \tag{5.23}$$

can be obtained from Eq. (5.6) by setting $\dot{\xi} = 0$. We may write the above expression in terms of averages over constrained molecular dynamics simulations as

$$-\tilde{\beta}A'(\xi^\star) = \frac{\langle Z_\xi^{-\frac{1}{2}} \mathcal{E}_i \rangle_{\xi^\star}}{\langle Z_\xi^{-\frac{1}{2}} \rangle_{\xi^\star}}, \tag{5.24}$$

where the dependence of the constrained momentum partition function on the coordinates has given rise to the Fixman correction.[138,139] These constrained molecular dynamics simulations are in practice carried out in Cartesian coordinates with the constraint enforced via the standard method of Lagrange multipliers.

5.3.1 Constrained OBABO (c-OBABO) integrator

The method described in Sec. 5.2.2 is applicable to constrained Hamiltonian systems, but unfortunately, the simulations of interest are to be run at constant temperature using a Langevin thermostat. However, as Lelièvre *et al.* have shown, the thermostat step may be adjusted in exactly the same manner for Langevin dynamics.[140,141] This allows us to transform the PILE (OBABO) integrator into the constrained version, c-OBABO. Although for our present purposes (computing the quantum PMF), we only require a single constraint on one bead, we derive here a more general integrator that supports multiple independent holonomic constraints on all beads.

Consider P functions $\boldsymbol{\xi}_j(\mathbf{q}^{(j)})$ and constant vectors $\mathbf{z}_j$ (not necessarily of the same length for each bead), making up the constraint equations $\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) = \mathbf{z}_j$. The implicit velocity constraints, when expressed in momentum form, are $\boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \mathbf{p}^{(j)} = \mathbf{0}$. Upon adding these constraints to the OBABO integrator and removing redundant constraint steps, we obtain the c-OBABO integrator for PIMD:

$$\mathrm{O}\begin{cases} \mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow T_k(\Delta t/2)\mathbf{P}^{(k)} + U_k(\Delta t/2)\mathbf{M}_1^{\frac{1}{2}}\boldsymbol{\eta} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)} \end{cases} \tag{5.25a}$$

$$\mathrm{B}\begin{cases} \bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \dfrac{\Delta t\tilde{\beta}}{2P\beta}\mathbf{F}(\mathbf{q}^{(j)}) \end{cases} \tag{5.25b}$$

$$\mathrm{C}\begin{cases} \bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0} \end{cases} \tag{5.25c}$$

$$\mathrm{A}\begin{cases} \mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk}\,;\ \mathbf{Q}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{q}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow S_k^{\mathbf{PP}}(\Delta t)\mathbf{P}^{(k)} + S_k^{\mathbf{PQ}}(\Delta t)\mathbf{M}_1\mathbf{Q}^{(k)} \\ \bar{\mathbf{Q}}^{(k)} \leftarrow S_k^{\mathbf{QP}}(\Delta t)\mathbf{M}_1^{-1}\mathbf{P}^{(k)} + S_k^{\mathbf{QQ}}(\Delta t)\mathbf{Q}^{(k)} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)} \\ \qquad\qquad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{PP}}(\Delta t)\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \bar{\mathbf{q}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{Q}}^{(k)} \\ \qquad\qquad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{QP}}(\Delta t)\mathbf{M}_1^{-1}\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \boldsymbol{\xi}_j(\bar{\mathbf{q}}^{(j)}) = \mathbf{z}_j \end{cases} \tag{5.25d}$$

$$\mathrm{B}\begin{cases} \bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \dfrac{\Delta t\tilde{\beta}}{2P\beta}\mathbf{F}(\mathbf{q}^{(j)}) \end{cases} \tag{5.25e}$$

$$\mathcal{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0}\end{cases} \tag{5.25f}$$

$$\mathcal{O}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow T_k(\Delta t/2)\mathbf{P}^{(k)} + U_k(\Delta t/2)\mathbf{M}_1^{\frac{1}{2}}\boldsymbol{\eta} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)}\end{cases} \tag{5.25g}$$

$$\mathcal{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0},\end{cases} \tag{5.25h}$$

where

$$\tilde{S}_{j\ell}^{\mathbf{xx}}(\Delta t) = \sum_{k=0}^{P-1} C_{jk} S_k^{\mathbf{xx}}(\Delta t) C_{\ell k} \tag{5.26}$$

are the normal mode propagation coefficients transformed to real space. For the velocity constraints, the Lagrange multipliers may be computed directly as

$$\boldsymbol{\Lambda}_j = -\left(\boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})\right)^{-1} \cdot \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \mathbf{p}^{(j)}, \tag{5.27}$$

without an iterative scheme.

Because both the normal mode transformations and the harmonic oscillator equations of motion are linear, in going from A in Eq. (5.8) to $\tilde{\mathrm{A}}$ in Eq. (5.25d), the constraint force was threaded through the sub-integrator, and is applied only at the very end of $\tilde{\mathrm{A}}$, after the inverse normal mode transformation. Although all the beads are coupled by the constraint and the Lagrange multipliers cannot be obtained directly in the general case, in Sec. 5.4 we show that this form can enable direct evaluation of the Lagrange multiplier when only one constraint is needed. Additionally, because the constraints couple the normal modes even in the absence of interactions, the standard derivation for the optimal friction of a thermostatted harmonic oscillator is not applicable. Despite this, we find that using the unmodified friction values from the PILE ($\gamma_k = 2\omega_k$ for $k \geq 1$) is a valid strategy in practice.

5.3.2 Constrained BAOAB (c-BAOAB) integrator

It has been demonstrated that the alternate integrator step order BAOAB may result in a smaller time step error for PIMD.[142] While Lelièvre *et al.* only consider the "side" scheme (which places the thermostat on the outer sides of the integrator), the feasibility and benefits of the "middle" scheme (which has the thermostat centered in the integrator) have also been recently established

for molecular dynamics with holonomic constraints.[143] The c-BAOAB integrator for PIMD may therefore be written as the following arrangement of sub-integrators and constraint steps:

$$\mathrm{B}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \dfrac{\Delta t\tilde{\beta}}{2P\beta}\mathbf{F}(\mathbf{q}^{(j)})\end{cases} \tag{5.28a}$$

$$\mathrm{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0}\end{cases} \tag{5.28b}$$

$$\tilde{\mathrm{A}}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk}\,;\ \mathbf{Q}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{q}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow S_k^{\mathbf{PP}}(\Delta t/2)\mathbf{P}^{(k)} + S_k^{\mathbf{PQ}}(\Delta t/2)\mathbf{M}_1\mathbf{Q}^{(k)} \\ \bar{\mathbf{Q}}^{(k)} \leftarrow S_k^{\mathbf{QP}}(\Delta t/2)\mathbf{M}_1^{-1}\mathbf{P}^{(k)} + S_k^{\mathbf{QQ}}(\Delta t/2)\mathbf{Q}^{(k)} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)} \\ \qquad\quad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{PP}}(\Delta t/2)\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \bar{\mathbf{q}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{Q}}^{(k)} \\ \qquad\quad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{QP}}(\Delta t/2)\mathbf{M}_1^{-1}\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \boldsymbol{\xi}_j(\bar{\mathbf{q}}^{(j)}) = \mathbf{z}_j\end{cases} \tag{5.28c}$$

$$\mathrm{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0}\end{cases} \tag{5.28d}$$

$$\mathrm{O}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow T_k(\Delta t)\mathbf{P}^{(k)} + U_k(\Delta t)\mathbf{M}_1^{\frac{1}{2}}\boldsymbol{\eta} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)}\end{cases} \tag{5.28e}$$

$$\mathrm{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j \\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0}\end{cases} \tag{5.28f}$$

$$\tilde{\mathrm{A}}\begin{cases}\mathbf{P}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{p}^{(j)} C_{jk}\,;\ \mathbf{Q}^{(k)} \leftarrow \sum_{j=1}^{P} \mathbf{q}^{(j)} C_{jk} \\ \bar{\mathbf{P}}^{(k)} \leftarrow S_k^{\mathbf{PP}}(\Delta t/2)\mathbf{P}^{(k)} + S_k^{\mathbf{PQ}}(\Delta t/2)\mathbf{M}_1\mathbf{Q}^{(k)} \\ \bar{\mathbf{Q}}^{(k)} \leftarrow S_k^{\mathbf{QP}}(\Delta t/2)\mathbf{M}_1^{-1}\mathbf{P}^{(k)} + S_k^{\mathbf{QQ}}(\Delta t/2)\mathbf{Q}^{(k)} \\ \bar{\mathbf{p}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{P}}^{(k)} \\ \qquad\quad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{PP}}(\Delta t/2)\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \bar{\mathbf{q}}^{(j)} \leftarrow \sum_{k=0}^{P-1} C_{jk}\bar{\mathbf{Q}}^{(k)} \\ \qquad\quad + \sum_{\ell=1}^{P} \tilde{S}_{j\ell}^{\mathbf{QP}}(\Delta t/2)\mathbf{M}_1^{-1}\boldsymbol{\nabla}_\ell\boldsymbol{\xi}_\ell(\mathbf{q}^{(\ell)}) \cdot \boldsymbol{\Lambda}_\ell \\ \boldsymbol{\xi}_j(\bar{\mathbf{q}}^{(j)}) = \mathbf{z}_j\end{cases} \tag{5.28g}$$

$$\mathrm{B}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \dfrac{\Delta t\tilde{\beta}}{2P\beta}\mathbf{F}(\mathbf{q}^{(j)}) & \text{(5.28h)}\\ \mathrm{C}\begin{cases}\bar{\mathbf{p}}^{(j)} \leftarrow \mathbf{p}^{(j)} + \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)}) \cdot \boldsymbol{\Lambda}_j\\ \boldsymbol{\nabla}_j\boldsymbol{\xi}_j(\mathbf{q}^{(j)})^{\mathrm{T}} \cdot \mathbf{M}_1^{-1} \cdot \bar{\mathbf{p}}^{(j)} = \mathbf{0}.\end{cases} & \text{(5.28i)}\end{cases}$$

Our rather unsophisticated implementation[144] of c-BAOAB suffers an increase in run time on the order of 10 % compared to c-OBABO due to the additional work required. Despite this, as we show in Sec. 5.5, the former may still outperform the latter, since it can allow a much larger time step to be used.

5.4 Exact relations for a center of mass distance constraint

When the reaction coordinate ξ is the radial separation between two centers of mass, several useful expressions may be derived. From this point onward, we work explicitly with N particles in 3-dimensional space, with coordinates $\mathbf{x}_i$ (where $\mathbf{x}$ describes all the degrees of freedom at the first bead, and can be thought of as a more structured reinterpretation of $\mathbf{q}^{(1)}$). The convex sums

$$\mathbf{x}_\alpha = \sum_{i=1}^{N_\alpha} \frac{m_{\alpha i}}{m_\alpha}\mathbf{x}_{\alpha i} \tag{5.29a}$$

and

$$\mathbf{x}_\beta = \sum_{i=1}^{N_\beta} \frac{m_{\beta i}}{m_\beta}\mathbf{x}_{\beta i} \tag{5.29b}$$

are the centers of mass of α and β (which we require to be non-empty and disjoint), where $m_{\alpha i}$ is the mass corresponding to $\mathbf{x}_{\alpha i}$ and

$$m_\alpha = \sum_{i=1}^{N_\alpha} m_{\alpha i}, \tag{5.30}$$

and likewise for β. We also use $\mathbf{x}_{\gamma i}$ to refer to the remaining particles, if any, which do not participate in either center of mass, with $N_\alpha + N_\beta + N_\gamma = N$.

We constrain the distance between the centers of mass as

$$\xi(\mathbf{x}) = |\mathbf{r}| = |\mathbf{x}_\alpha - \mathbf{x}_\beta| = z, \tag{5.31}$$

for an arbitrary positive z. The derivatives of this function are

$$\boldsymbol{\nabla}_{\alpha i}\xi(\mathbf{x}) = \frac{m_{\alpha i}\mathbf{r}}{m_\alpha \xi(\mathbf{x})}, \tag{5.32a}$$

$$\boldsymbol{\nabla}_{\beta i}\xi(\mathbf{x}) = -\frac{m_{\beta i}\mathbf{r}}{m_\beta \xi(\mathbf{x})}, \tag{5.32b}$$

and

$$\boldsymbol{\nabla}_{\gamma i}\xi(\mathbf{x}) = 0. \tag{5.32c}$$

Immediately, it follows that Eq. (5.21) evaluates to

$$Z_\xi(\mathbf{x}) = \sum_{i=1}^{N_\alpha} \frac{1}{m_{\alpha i}} |\boldsymbol{\nabla}_{\alpha i}\xi(\mathbf{x})|^2 + \sum_{i=1}^{N_\beta} \frac{1}{m_{\beta i}} |\boldsymbol{\nabla}_{\beta i}\xi(\mathbf{x})|^2 \tag{5.33a}$$

$$= \frac{1}{m_\alpha} + \frac{1}{m_\beta} = \frac{1}{\mu_{\alpha\beta}}, \tag{5.33b}$$

which is a constant (with $\mu_{\alpha\beta}$ being the reduced mass), so the Fixman correction cancels from Eq. (5.24) leaving us with just

$$-\tilde{\beta} A'(\xi^\star) = \langle \mathcal{E}_i \rangle_{\xi^\star}. \tag{5.34}$$

5.4.1 Lagrange multiplier

It is also straightforward in this circumstance to obtain a closed-form solution for the Lagrange multiplier λ for the constrained harmonic oscillator propagation step $\tilde{\mathrm{A}}$. Because the momentum perturbation was propagated through the linear equations of motion, we know that the equation to be satisfied is

$$\xi\Big(\bar{\mathbf{x}} + \lambda \tilde{S}_{11}^{\mathbf{QP}} \mathbf{M}_1^{-1} \boldsymbol{\nabla}_1 \xi(\mathbf{x})\Big) = z, \tag{5.35}$$

where $\mathbf{x}$ is the value of $\mathbf{q}^{(1)}$ before the propagation, and $\bar{\mathbf{x}}$ is its value after propagation by the unconstrained step A. After some algebraic manipulations, we get

$$\left| \left(1 + \frac{\lambda \tilde{S}_{11}^{\mathbf{QP}}}{\mu_{\alpha\beta}\xi(\mathbf{x})}\right) \mathbf{r} + \Delta\mathbf{r} \right| = z, \tag{5.36}$$

where $\Delta\mathbf{r} = \bar{\mathbf{r}} - \mathbf{r}$ is the change in relative position of the centers of mass due to the unconstrained step. We assume that $\tilde{S}_{11}^{\mathbf{QP}}$ is not zero.

From geometric considerations, we see that solutions exist only when the component of $\Delta\mathbf{r}$ orthogonal to $\mathbf{r}$ does not surpass z:

$$|\Delta\mathbf{r}|^2 - (\Delta\mathbf{r}\cdot\hat{\mathbf{r}})^2 \leq z^2, \tag{5.37}$$

where $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$ is the unit vector in the direction of $\mathbf{r}$. When this condition is met, the Lagrange multiplier is given by

$$\lambda = -\frac{\mu_{\alpha\beta}}{\tilde{S}_{11}^{\mathbf{QP}}}\left[|\mathbf{r}| + \Delta\mathbf{r}\cdot\hat{\mathbf{r}} - \sqrt{z^2 - \left(|\Delta\mathbf{r}|^2 - (\Delta\mathbf{r}\cdot\hat{\mathbf{r}})^2\right)}\right], \tag{5.38}$$

which may be computed directly for very little cost. When the square root is not zero, there is another solution for λ, which has the root being added rather than subtracted, but this corresponds to the interchange of the two centers of mass; we assume that the time step is sufficiently small that this will never be the desired outcome.

The interpretation of the expression for λ is simple. The resulting shifts of the particles in α and β are, respectively,

$$-\frac{m_\beta\hat{\mathbf{r}}}{m_\alpha + m_\beta}\left[|\mathbf{r}| + \Delta\mathbf{r}\cdot\hat{\mathbf{r}} - \sqrt{z^2 - \left(|\Delta\mathbf{r}|^2 - (\Delta\mathbf{r}\cdot\hat{\mathbf{r}})^2\right)}\right] \tag{5.39a}$$

and

$$\frac{m_\alpha\hat{\mathbf{r}}}{m_\alpha + m_\beta}\left[|\mathbf{r}| + \Delta\mathbf{r}\cdot\hat{\mathbf{r}} - \sqrt{z^2 - \left(|\Delta\mathbf{r}|^2 - (\Delta\mathbf{r}\cdot\hat{\mathbf{r}})^2\right)}\right]. \tag{5.39b}$$

The first two terms of each are responsible for completely removing any separation between the centers of mass along $\mathbf{r}$: the first term takes care of the original vector, while the second term handles the component of $\Delta\mathbf{r}$ that is parallel to $\mathbf{r}$. The prefactors ensure that the groups move toward one another, and that the lighter group moves farther. The square root term then restores some of the separation along $\mathbf{r}$, with the exact amount being determined by the excess allowance from Eq. (5.37).

It is important to note that Eq. (5.38) is not an equation of motion for the constraint. Rather, this is an exact solution to the Lagrange multiplier optimization, equivalent to one that would be obtained by an iterative scheme. As such, it is not susceptible to numerical drift of the constrained coordinate.

5.4.2 Derivatives for estimators

Evaluation of the estimators $\mathcal{E}_1$ and $\mathcal{E}_2$ requires knowledge of $\frac{\partial}{\partial\xi}\log|J(\mathbf{x})|$ and $\frac{\partial\mathbf{x}}{\partial\xi}$, which we find by performing several coordinate transformations. We begin by transforming $\mathbf{x}$ into the Jacobi coordinates

$$\mathbf{y}_{\alpha\ell} = \sum_{i=1}^{\ell} \frac{m_{\alpha i}}{m_\alpha^{1\to\ell}}\mathbf{x}_{\alpha i} - \begin{cases} \mathbf{x}_{\alpha,\ell+1} & \text{if } 1 \le \ell \le N_\alpha - 1 \\ 0 & \text{if } \ell = N_\alpha, \end{cases} \tag{5.40}$$

with

$$m_\alpha^{1\to\ell} = \sum_{i=1}^{\ell} m_{\alpha i}, \tag{5.41}$$

and similarly for β. The remaining coordinates $\mathbf{x}_{\gamma i}$ are left unmodified. As shown in Appendix 5.B, this transformation has unit Jacobian determinant.

The coordinates $\mathbf{y}_{\alpha N_\alpha}$ and $\mathbf{y}_{\beta N_\beta}$ are the centers of mass $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$, so we further transform them to

$$\mathbf{R} = \frac{m_\alpha \mathbf{y}_{\alpha N_\alpha} + m_\beta \mathbf{y}_{\beta N_\beta}}{m_\alpha + m_\beta} \tag{5.42a}$$

and

$$\mathbf{r} = \mathbf{y}_{\alpha N_\alpha} - \mathbf{y}_{\beta N_\beta}, \tag{5.42b}$$

and this change of variables also has unit Jacobian determinant.

Finally, the transformation of $\mathbf{r}$ to the spherical coordinates $(\xi, \cos\theta, \varphi)$ is known to have a Jacobian determinant whose absolute value is ξ^2. The overall transformation from $\mathbf{x}$ to $(\mathbf{y}_{\alpha i}, \mathbf{y}_{\beta i}, \mathbf{R}, \xi, \cos\theta, \varphi, \mathbf{x}_{\gamma i})$ therefore has $|J(\mathbf{x})| = \xi^2$, so

$$\frac{\partial}{\partial\xi}\log|J(\mathbf{x})| = \frac{2}{\xi}. \tag{5.43}$$

The original Cartesian coordinates may be written as

$$\mathbf{x}_{\alpha i} = \mathbf{R} + \frac{m_\beta \mathbf{r}}{m_\alpha + m_\beta} + \sum_{\ell=i}^{N_\alpha - 1} \frac{m_{\alpha,\ell+1}}{m_\alpha^{1\to\ell+1}}\mathbf{y}_{\alpha\ell} - \begin{cases} 0 & \text{if } i = 1 \\ \dfrac{m_\alpha^{1\to i-1}}{m_\alpha^{1\to i}}\mathbf{y}_{\alpha,i-1} & \text{if } 2 \le i \le N_\alpha \end{cases} \tag{5.44a}$$

and

$$\mathbf{x}_{\beta i} = \mathbf{R} - \frac{m_\alpha \mathbf{r}}{m_\alpha + m_\beta} + \sum_{\ell=i}^{N_\beta - 1} \frac{m_{\beta,\ell+1}}{m_\beta^{1\to\ell+1}} \mathbf{y}_{\beta\ell} - \begin{cases} 0 & \text{if } i = 1 \\ \dfrac{m_\beta^{1\to i-1}}{m_\beta^{1\to i}} \mathbf{y}_{\beta,i-1} & \text{if } 2 \leq i \leq N_\beta. \end{cases} \tag{5.44b}$$

Hence, the derivatives of $\mathbf{x}$ with respect to ξ are

$$\frac{\partial \mathbf{x}_{\alpha i}}{\partial \xi} = \frac{m_\beta}{m_\alpha + m_\beta} \hat{\mathbf{r}}, \tag{5.45a}$$

$$\frac{\partial \mathbf{x}_{\beta i}}{\partial \xi} = -\frac{m_\alpha}{m_\alpha + m_\beta} \hat{\mathbf{r}}, \tag{5.45b}$$

and

$$\frac{\partial \mathbf{x}_{\gamma i}}{\partial \xi} = 0. \tag{5.45c}$$

As observed in the shifts to enforce the constraints in Eq. (5.39), the prefactor for the lighter group is larger in magnitude than for heavier group: the former must move faster under a changing separation distance than the latter.

5.5 Results

To demonstrate the effectiveness of the constrained PIMD integrators, we apply them to a q-SPC/Fw[145] water dimer and obtain its quantum PMF as a function of the distance $\xi(\mathbf{x}) = |\mathbf{x}_\alpha - \mathbf{x}_\beta|$ between the molecules' centers of mass $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$. To generate reference results, we use the path integral umbrella sampling method (US/WHAM), which requires a histogram unbiasing step in order to stitch together the obtained histograms.[24]

Before attempting to generate a PMF, we first identify which integrator has better time step error characteristics. In Fig. 5.1, we show two combinations of temperature and constraint distance that have an appreciable difference in behavior between c-OBABO (solid) and c-BAOAB (dashed). In both cases, the time step error decreases faster for the latter, which is to be expected from past work with this ordering of integrator steps.[142,143] For the remainder of the calculations, we only use c-BAOAB, as it allows us to save some computational effort by using a larger time step to achieve the same level of error.

As shown in Fig. 5.2, direct estimation of the derivative of the PMF is successful using the c-BAOAB integrator and both estimators. The reference curves are obtained by numerically differentiating the US/WHAM PMF, which causes a slight wobble that is noticeable at 300 K. At 10 K, the PIMD

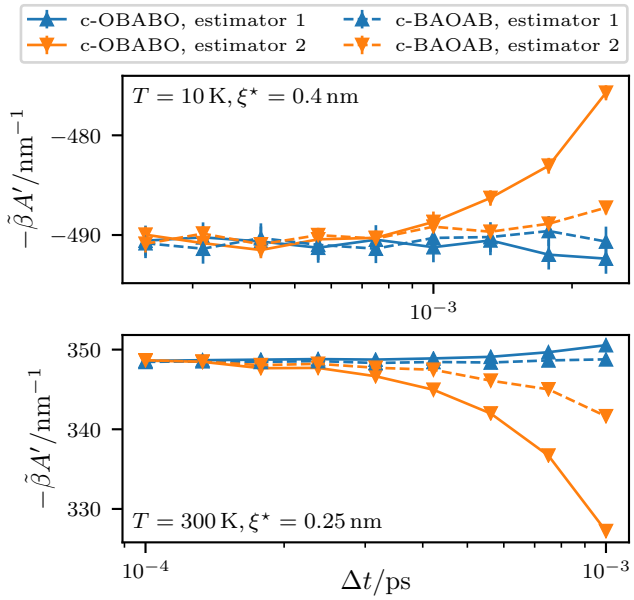

Figure 5.1: Comparison of the time step convergence of the PMF derivative of a water dimer using the integrators c-OBABO and c-BAOAB in Eqs. (5.25) and (5.28) at $T = 10\,\mathrm{K}$, $\xi^\star = 0.4\,\mathrm{nm}$ (top panel) and $T = 300\,\mathrm{K}$, $\xi^\star = 0.25\,\mathrm{nm}$ (bottom panel). Line segments added to guide the eye.

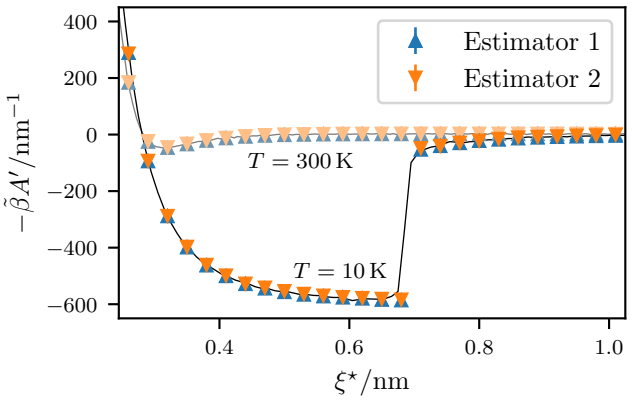

Figure 5.2: Derivative of the PMF of a water dimer at $T = 300\,\mathrm{K}$ (top curve, less saturated) and $10\,\mathrm{K}$ (bottom curve, more saturated). Error bars are not visible, because they are smaller than the symbols. The solid curves show the US/WHAM results.

simulations show that the derivative changes rather sharply near 0.7 nm, but the reference curve has rounded corners. This region of the reaction coordinate is difficult to simulate, as the system straddles the boundary between being two largely free monomers and a strongly bound dimer. This makes it challenging to obtain accurate histograms, and even more so for their derivatives, but we believe that the estimators for the direct calculation of the PMF derivative are not as sensitive to this sampling difficulty. This allows us to readily estimate the derivative of the PMF at any point along the reaction coordinate from a single PIMD simulation.

The error bars in Fig. 5.2 are small enough that they cannot be seen at that scale, so we present them separately in Fig. 5.3. At 300 K, the errors are relatively constant, without any interesting features. However, the picture is not quite as simple at 10 K, which has a crossing precisely in the problematic region near $\xi^\star = 0.7\,\mathrm{nm}$. This suggests that while both sides of the identity in Eq. (42) of Paper I (Eq. (4.42) in Chapter 4) converge to the same value, one may be preferable to the other depending on the circumstances.

Finally, we numerically integrate the obtained $-\tilde{\beta}A'$ values and plot the renormalized PMF $\tilde{A}$ in Fig. 5.4, using the same procedure as in Paper I. There is excellent agreement with the reference results at both temperatures, suggesting that this approach of integrating the derivative of the PMF is a viable alternative to umbrella sampling and histogram unbiasing.

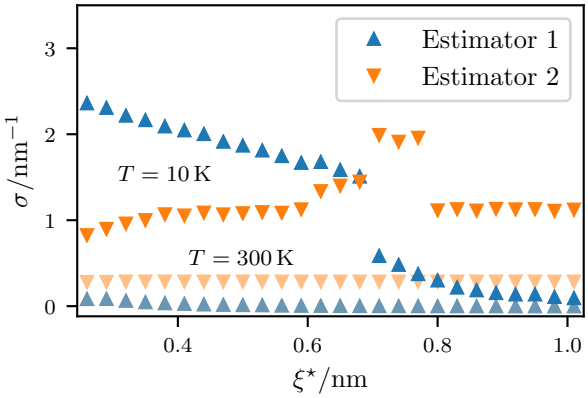

Figure 5.3: Comparison of the standard error of the mean of the PMF derivatives in Fig. 5.2 at $T = 300\,\mathrm{K}$ (lower points, less saturated) and 10 K (upper points, more saturated).

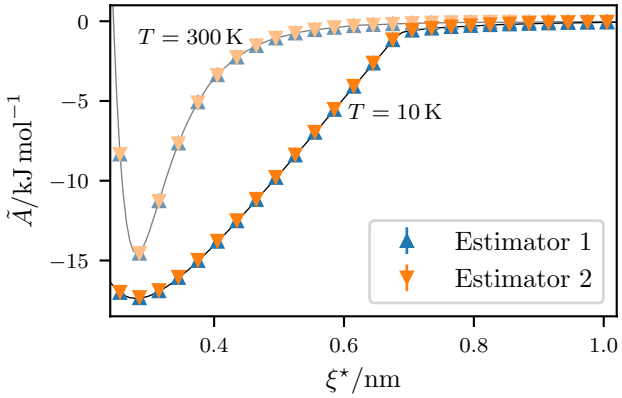

Figure 5.4: Potential of mean force of a water dimer at $T = 300\,\mathrm{K}$ (top curve, less saturated) and 10 K (bottom curve, more saturated). Error bars are not visible, because they are smaller than the symbols. Additional points extending to $\xi_0^\star = 1.5\,\mathrm{nm}$ are not displayed, and the integration grid is 3 times as dense as depicted here and in Fig. 5.2. The solid curves show the US/WHAM results.

5.6 Conclusions

We have augmented the PILE integrator with bead-local holonomic constraints to produce the c-OBABO and c-BAOAB integrators for PIMD. These integrators allow our estimators from Paper I to be used with molecular dynamics in addition to Monte Carlo, which greatly expands their scope of applicability. The constrained harmonic oscillator propagation step $\tilde{\mathrm{A}}$ of these integrators has the constraint force applied at the end rather than the beginning of the step, which allows us to find an exact expression for the Lagrange multiplier for some reaction coordinates, such as the distance between two centers of mass.

Using one of the constrained integrators, we have computed the derivative of the PMF of a water dimer at 10 K and 300 K. We observe that the constrained PIMD method captures the sharp step in the derivative better than umbrella sampling. We have also been able to successfully integrate the derivative to obtain the PMF itself.

Together, these novel estimators and integrators may be utilized to find free energy differences and potentials of mean force of molecular clusters using implementations based on existing PIMD simulation software. The conventional umbrella sampling and histogram unbiasing approach has several practical drawbacks: the user must choose a force constant for the restraint; the ends of the histogram have a tendency to be noisy, so histograms must be generated well past the region of interest; in addition to requiring a grid of points for the unbiasing step, the user must also select the locations of the restraint windows and make sure that the resulting histograms have sufficient

overlap; the iterative unbiasing procedure requires a stopping criterion. Our method, on the other hand, allows the result of each PIMD simulation to be converged independently of all the others, and there are no additional tunable parameters introduced into the simulations besides the choice of integration grid.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2016-04403), the Ontario Ministry of Research and Innovation (MRI), the Canada Research Chair program (950-231024), and the Canada Foundation for Innovation (CFI) (project No. 35232).

5.A Block matrix determinant identity

Consider an invertible block matrix and its inverse with identical block structure,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}^{-1}, \tag{5.46}$$

whose blocks $\mathbf{A}$ and $\mathbf{Z}$ are themselves invertible. Upon combining two manifestations of the Schur complement,[146]

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| \left|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right| \tag{5.47}$$

and

$$\mathbf{Z}^{-1} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}, \tag{5.48}$$

we conclude that

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}| \, |\mathbf{Z}|^{-1}. \tag{5.49}$$

5.B Jacobian determinant of transformation to Jacobi coordinates

The transformation to Jacobi coordinates described in Sec. 5.4.2 couples neither α and β nor the spatial degrees of freedom, and also does not involve γ, so its Jacobian matrix will be block diagonal with 7 independent blocks. The entire γ block is irrelevant, as its determinant is 1. The remaining 6 blocks are identical in form, so we may scrutinize only one of them, using x_i to label an arbitrary Cartesian component of either $\mathbf{x}_{\alpha i}$ or $\mathbf{x}_{\beta i}$, with corresponding masses m_i and transformed coordinates y_ℓ.

Trivially, the derivative of y_ℓ with respect to x_i is

$$\frac{\partial y_\ell}{\partial x_i} = \begin{cases} \dfrac{m_i}{m^{1\to\ell}} & \text{if } i \leq \ell \\ -1 & \text{if } i = \ell + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.50}$$

so the Jacobian matrix $\mathbf{J}$ with elements

$$J_{\ell i} = \frac{\partial y_\ell}{\partial x_i} \tag{5.51}$$

is a lower Hessenberg matrix (all elements above the superdiagonal $i = \ell + 1$ are zero). Thus, we may use a recurrence relation to find $|\mathbf{J}|$.[147]

In the following, $\mathbf{J}^{(n)}$ denotes the $n \times n$ leading principal submatrix of $\mathbf{J}$, and we see that $|\mathbf{J}^{(0)}| = 1$ and $|\mathbf{J}^{(1)}| = 1$. For subsequent n,

$$|\mathbf{J}^{(n)}| = J_{nn}|\mathbf{J}^{(n-1)}| + \sum_{i=1}^{n-1}(-1)^{n-i}J_{ni}\prod_{j=i}^{n-1}J_{j,j+1}|\mathbf{J}^{(j-1)}| \tag{5.52a}$$

$$= \frac{m_n}{m^{1\to n}}|\mathbf{J}^{(n-1)}| + \sum_{i=1}^{n-1}\frac{m_i}{m^{1\to n}}\prod_{j=i}^{n-1}|\mathbf{J}^{(j-1)}|\,. \tag{5.52b}$$

By induction, it stands to reason that $|\mathbf{J}^{(n)}| = 1$ for all n, and therefore $|\mathbf{J}| = 1$, so the entire transformation to Jacobi coordinates for both centers of mass has unit Jacobian determinant. This is consistent with the view of Jacobi coordinates as iterated pairwise transformations to center of mass and relative distance coordinates, which individually have unit Jacobian determinant.

Chapter 6

Conclusions and outlook

In the preceding chapters, we described new algorithms for quantum molecular dynamics, and applied them to various systems. In Chapter 2, we found a simple form for the ladder operators $\hat{\ell}^+$ and $\hat{\ell}^-$ for the ℓ quantum number of the spherical harmonics $|\ell\, m\rangle$. Because these operators have a well-defined flux, we were able to implement them efficiently using the ITensor package. The dipole–dipole interaction potential for linear rotors is straightforward to expand in terms of these operators, due to the compact form of the $\hat{x}$, $\hat{y}$, and $\hat{z}$ operators.

Using the DMRG implementation in ITensor, we were able to obtain ground states of the dipolar linear rotor chain Hamiltonian for up to 100 sites. The states near $R = 1$ showed a peak in the entanglement entropy S_{vN}, which is further evidence of a quantum phase transition in that region. We were also able to construct part of a dipole excitation spectrum using energy differences between the ground state of the lowest-lying symmetry block and ground states of other symmetry blocks.

This technique has since been used to benchmark the entanglement entropy in path integral ground state (PIGS) Monte Carlo calculations in Ref. [8]. Although the expectation is that the PIGS approach will one day be more efficient for three-dimensional structures of linear rotors, DMRG is currently able to compute ground state properties of long one-dimensional chains much faster. Once an MPS is found, it is possible to invoke a perfect sampling algorithm to draw samples distributed according to the squares of the state amplitudes.[148] This enabled us to provide training data for machine learning of a restricted Boltzmann machine for linear rotor ground states in Ref. [9].

In Chapter 3, we improved PIMC sampling for vibronic systems in three different ways. First, we replaced random component selection by a deterministic analogue, which helps on two fronts: it reduces the impact of severe outliers, and it provides a way to systematically recombine the RqMC results for the different components. It is a straightforward technique with little downside. Next, we substituted RqMC for MC when drawing samples from each component. This is a more involved

procedure, and there are potential issues when sampling from high-dimensional spaces, but we observed that it was able to reduce the statistical error in the partition function estimates for the model systems.

Finally, we introduced two flavors of parameter optimization. The first relies on already having a good sampling distribution, and it perturbs the parameters slightly to improve it further. This is achieved by minimizing a loss function built around the KL divergence between the desired distribution and the current distribution. The second starts from an empty Hamiltonian and gradually adds the desired terms back in. This ensures that the current distribution is never too different from the target distribution. Both of these variants were successful when applied to model systems, with the exception of a strongly-coupled Jahn–Teller system.

In Chapters 4 and 5, we derived novel path integral estimators ($\mathcal{E}_1$, $\mathcal{E}_2$) and integrators (c-OBABO, c-BAOAB) for free energy calculations. Unlike histogram-based approaches that combine histograms from several calculations using unbiasing techniques such as WHAM, our proposed method is able to calculate the derivative of the PMF directly at any point along the reaction coordinate. This simplifies the parameter convergence phase, as there are fewer user-serviceable parameters, and each estimate is a single value that may be improved independently of the others. The obtained derivative may be numerically integrated to find the PMF itself.

We computed the PMF of Ar_2 with a Lennard-Jones potential using our estimators and PIMC, and found good agreement with numerical matrix multiplication. We also used our PIMD integrators to compute the PMF of a water dimer with the q-SPC/Fw model, which agrees with the reference umbrella sampling result. These calculations were performed using custom proof-of-concept PIMC and PIMD implementations, which are unfortunately not easily extendable to more complex systems.

6.1 Future directions

Subsequent to the publication of Ref. [1], we have been working to compute ground state properties of the dipolar linear rotor chain Hamiltonian for even longer systems. To simplify the calculations, we also consider a modified Hamiltonian in which only interactions between nearest neighbors are present. Our preliminary results may be seen in Fig. 6.1, and appear promising.

For $\ell_{\max} = 1$ in the case of nearest neighbor interactions, we see that 1) the von Neumann entropy S_{vN} for an even bipartition of the system has a very peaked "λ"-like shape, 2) the difference between the largest Schmidt coefficients, $\lambda_1 - \lambda_2$, has a crisp corner, and 3) the energy gap $E_1 - E_0$ between the ground state and the first excited state is extremely pointy. However, as $\ell_{\max}$ grows, especially when all interactions are included in the Hamiltonian, it becomes progressively more time consuming to compute these properties near the critical point for large systems. Thus, we do not currently

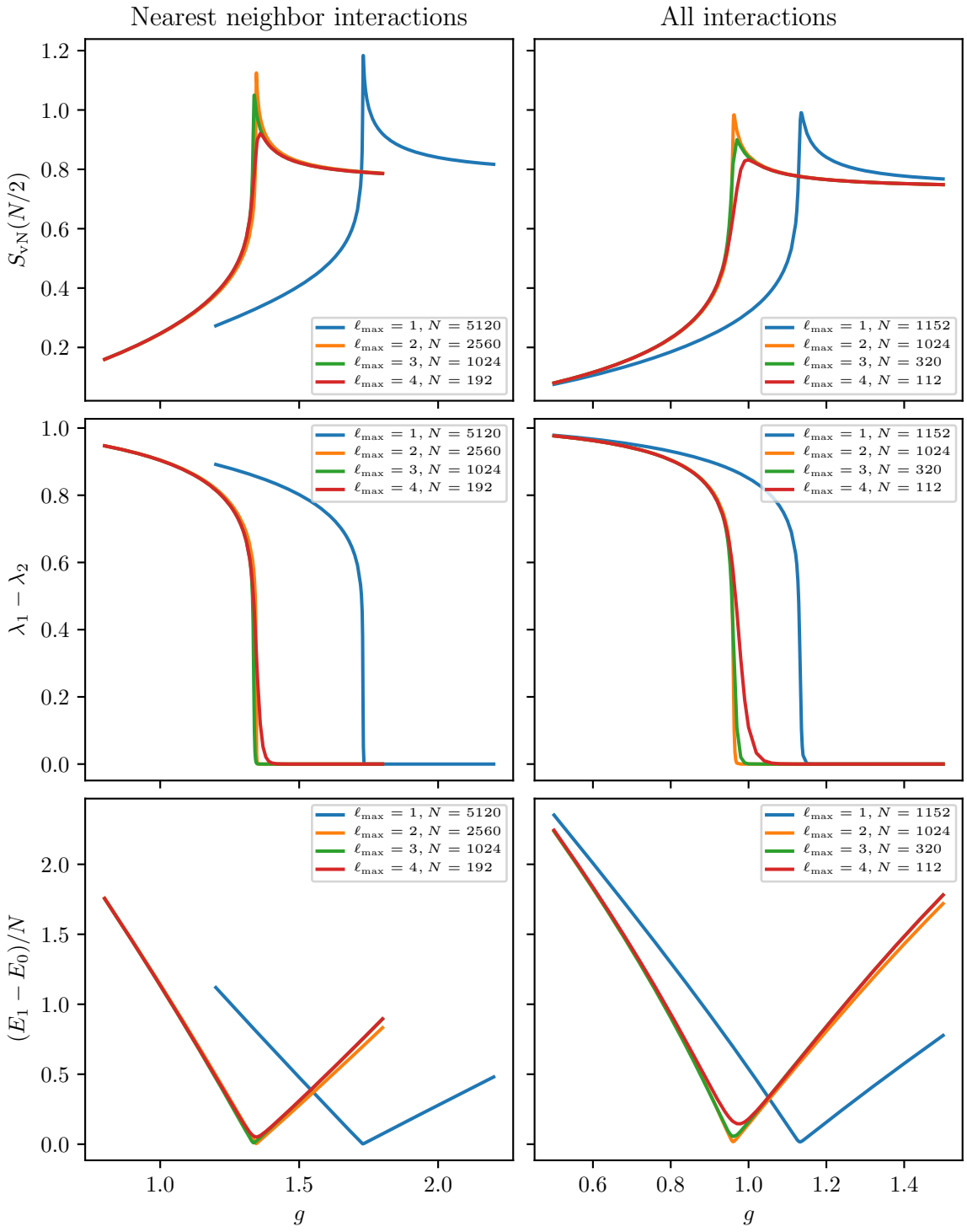

Figure 6.1: Convergence of several properties (von Neumann entanglement entropy, Schmidt gap, energy gap) of dipolar linear rotor chains with basis truncation parameter $\ell_{\max}$. Systems with only nearest neighbor interactions are in the left column; systems with all interactions are in the right column. The Hamiltonian parameter $g = 1/R^3$ serves the same purpose as R in Chapter 2.

have data for large enough system sizes N across the board, as evidenced by rounded-over features at larger $\ell_{\max}$. Calculations to remedy this are currently underway, and once they are completed, we hope to be able to extract the critical exponents and central charge associated with this quantum phase transition.[68,72]

In the meantime, our collaborators at the Université Paris-Saclay and Universität Heidelberg have been applying the multi-configuration time-dependent Hartree (MCTDH) method to the same system of dipolar rotors. This is not a typical use case for MCTDH, which is more often applied to the problem of molecular vibrations.[149] Our goal is to compare the efficiency of MCTDH (and the corresponding hierarchical Tucker decomposition) against DMRG (and matrix product states) for one-dimensional systems of quantum rotors, since both methods are based on the tensor network paradigm.[150] The calculations have been performed by all parties, and a manuscript is in progress.

Furthermore, we plan to extend the linear rotor DMRG method to asymmetric tops, such as water molecules, with arbitrary interaction potentials. The latter aspect of this has already seen some progress thanks to Adam Marr, who has successfully implemented a novel approach for the decomposition of two-body potentials into sums of products of one-body terms using the singular value decomposition (SVD). The extension of the method to more sophisticated rotors will involve the derivation of the appropriate ladder operators.

Now that the estimators and integrators for the calculation of the quantum PMF with PIMD have been validated, our aim is to apply them to more interesting systems. Work has been done by Kevin Bishop towards implementing the integrators in OpenMM, which opens the door to free energy calculations of systems with more degrees of freedom, as its highly optimized and GPU-enabled modules allow large-scale simulations to be completed in a reasonable amount of time.[151] OpenMM also provides straightforward access to the MB-pol water model,[152] which is preferable to the one that was used in this thesis for benchmarking purposes. Once this work is complete, we expect to be able to obtain an accurate PMF of a small water cluster using this method.

References

[1]D. Iouchtchenko and P.-N. Roy, J. Chem. Phys. **148**, 134115 (2018) 10.1063/1.5024403.

[2]D. Iouchtchenko, N. Raymond, P.-N. Roy, and M. Nooijen, arXiv preprint arXiv:1912.11594v1 (2019).

[3]D. Iouchtchenko, K. P. Bishop, and P.-N. Roy, arXiv preprint arXiv:2101.00761v1 (2020).

[4]D. Iouchtchenko, K. P. Bishop, and P.-N. Roy, arXiv preprint arXiv:2101.00762v1 (2020).

[5]R. Melko, C. Herdman, D. Iouchtchenko, P.-N. Roy, and A. Del Maestro, Phys. Rev. A **93**, 042336 (2016) 10.1103/PhysRevA.93.042336.

[6]T. Halverson, D. Iouchtchenko, and P.-N. Roy, J. Chem. Phys. **148**, 074112 (2018) 10.1063/1.5011769.

[7]N. Raymond, D. Iouchtchenko, P.-N. Roy, and M. Nooijen, J. Chem. Phys. **148**, 194110 (2018) 10.1063/1.5025058.

[8]T. Sahoo, D. Iouchtchenko, C. Herdman, and P.-N. Roy, J. Chem. Phys. **152**, 184113 (2020) 10.1063/5.0004602.

[9]I. J. De Vlugt, D. Iouchtchenko, E. Merali, P.-N. Roy, and R. G. Melko, Phys. Rev. B **102**, 035108 (2020) 10.1103/PhysRevB.102.035108.

[10]J. A. Pople, "Quantum Chemical Models", in *Nobel Lectures in Chemistry (1996-2000)*, edited by I. Grenthe (World Scientific, 2003), pp. 246–260, ISBN: 981-02-4958-6.

[11]E. Tiesinga, P. J. Mohr, D. B. Newell, and B. N. Taylor, *The 2018 CODATA Recommended Values of the Fundamental Physical Constants (Web Version 8.1)*, Database developed by J. Baker, M. Douma, and S. Kotochigova. Available at http://physics.nist.gov/constants, National Institute of Standards and Technology, Gaithersburg, MD 20899.

[12]H. Köppel, W. Domcke, and L. S. Cederbaum, Adv. Chem. Phys. **57**, 59–246 (1984) 10.1002/9780470142813.ch2.

[13]D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, 1976), ISBN: 0-06-044366-9.

[14]R. P. Feynman, A. R. Hibbs, and D. F. Styer, *Quantum Mechanics and Path Integrals* (Dover, 2010), ISBN: 978-0-486-47722-0.

[15]M. Parrinello and A. Rahman, J. Chem. Phys. **80**, 860–867 (1984) 10.1063/1.446740.

[16]M. Ceriotti, M. Parrinello, T. E. Markland, and D. E. Manolopoulos, J. Chem. Phys. **133**, 124104 (2010) 10.1063/1.3489925.

[17]J. Schmidt and J. C. Tully, J. Chem. Phys. **127**, 094103 (2007).

[18]J. Lu and Z. Zhou, J. Chem. Phys. **146**, 154110 (2017).

[19]J. C. Spall, IEEE Trans. Automat. Contr. **37**, 332–341 (1992) 10.1109/9.119632.

[20]J. C. Spall, IEEE Trans. Aerosp. Electron. Syst. **34**, 817–823 (1998) 10.1109/7.705889.

[21]H. Niederreiter, J. Number Theory **30**, 51–70 (1988) 10.1016/0022-314X(88)90025-X.

[22]D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, 2002), ISBN: 0-12-267351-4.

[23]S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comp. Chem. **13**, 1011–1021 (1992) 10.1002/jcc.540130812.

[24]K. P. Bishop and P.-N. Roy, J. Chem. Phys. **148**, 102303 (2018) 10.1063/1.4986915.

[25]U. Schollwöck, Ann. Phys. **326**, 96–192 (2011) 10.1016/j.aop.2010.09.012.

[26]J. Eisert, M. Cramer, and M. B. Plenio, Rev. Mod. Phys. **82**, 277 (2010).

[27]U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005) 10.1103/revmodphys.77.259.

[28]G. K.-L. Chan and S. Sharma, Annu. Rev. Phys. Chem. **62**, 465–481 (2011) 10.1146/annurev-physchem-032210-103338.

[29]A. Baiardi, C. J. Stein, V. Barone, and M. Reiher, J. Chem. Theory Comput. **13**, 3764–3777 (2017) 10.1021/acs.jctc.7b00329.

[30]B. P. Abolins, R. E. Zillich, and K. B. Whaley, J. Low Temp. Phys. **165**, 249–260 (2011) 10.1007/s10909-011-0398-1.

[31]J. Biskupek, S. T. Skowron, C. T. Stoppiello, G. A. Rance, S. Alom, K. L. Fung, R. J. Whitby, M. H. Levitt, Q. M. Ramasse, U. Kaiser, et al., ACS Nano **14**, 11178–11189 (2020) 10.1021/acsnano.0c02661.

[32]A. Krachmalnicoff, R. Bounds, S. Mamone, S. Alom, M. Concistrè, B. Meier, K. Kouřil, M. E. Light, M. R. Johnson, S. Rols, et al., Nat. Chem. **8**, 953–957 (2016) 10.1038/nchem.2563.

[33]A. A. Mason and A. H. Nielsen, J. Opt. Soc. Am. **57**, 1464–1470 (1967).

[34]K. Komatsu, M. Murata, and Y. Murata, Science **307**, 238–240 (2005) 10.1126/science.1106185.

[35]K. Kurotobi and Y. Murata, Science **333**, 613–616 (2011) 10.1126/science.1206376.

[36]T. Rõõm, L. Peedu, M. Ge, D. Hüvonen, U. Nagel, S. Ye, M. Xu, Z. Bačić, S. Mamone, M. H. Levitt, et al., Philos. Trans. R. Soc., A **371**, 20110631 (2013) 10.1098/rsta.2011.0631.

[37]M. Xu, S. Ye, and Z. Bačić, J. Phys. Chem. Lett. **6**, 3721–3725 (2015) 10.1021/acs.jpclett.5b01505.

[38]P. M. Felker and Z. Bačić, Chem. Phys. Lett. **683**, 172–178 (2017) 10.1016/j.cplett.2017.02.027.

[39]P. M. Felker and Z. Bačić, J. Chem. Phys. **146**, 084303 (2017) 10.1063/1.4976526.

[40]P. M. Felker, V. Vlček, I. Hietanen, S. FitzGerald, D. Neuhauser, and Z. Bačić, Phys. Chem. Chem. Phys. **19**, 31274–31283 (2017) 10.1039/c7cp06062a.

[41]Y. N. Kalugina and P.-N. Roy, J. Chem. Phys. **147**, 244303 (2017) 10.1063/1.5006589.

[42]B. W. Smith, M. Monthioux, and D. E. Luzzi, Nature **396**, 323 (1998) 10.1038/24521.

[43]B. W. Smith, M. Monthioux, and D. E. Luzzi, Chem. Phys. Lett. **315**, 31–36 (1999) 10.1016/s0009-2614(99)00896-9.

[44]B. Burteaux, A. Claye, B. W. Smith, M. Monthioux, D. E. Luzzi, and J. E. Fischer, Chem. Phys. Lett. **310**, 21–24 (1999) 10.1016/s0009-2614(99)00720-4.

[45]J. Sloan, R. E. Dunin-Borkowski, J. L. Hutchison, K. S. Coleman, V. C. Williams, J. B. Claridge, A. P. E. York, C. Xu, S. R. Bailey, G. Brown, et al., Chem. Phys. Lett. **316**, 191–198 (2000) 10.1016/s0009-2614(99)01250-6.

[46]M. Monthioux, Carbon **40**, 1809–1823 (2002) 10.1016/s0008-6223(02)00102-1.

[47]D. J. Hornbaker, S.-J. Kahng, S. Misra, B. W. Smith, A. T. Johnson, E. J. Mele, D. E. Luzzi, and A. Yazdani, Science **295**, 828–831 (2002) 10.1126/science.1068133.

[48]K. Hirahara, K. Suenaga, S. Bandow, H. Kato, T. Okazaki, H. Shinohara, and S. Iijima, Phys. Rev. Lett. **85**, 5384 (2000) 10.1103/physrevlett.85.5384.

[49]B.-Y. Sun, T. Inoue, T. Shimada, T. Okazaki, T. Sugai, K. Suenaga, and H. Shinohara, J. Phys. Chem. B **108**, 9011–9015 (2004) 10.1021/jp049130a.

[50]H. Shiozawa, H. Rauf, T. Pichler, M. Knupfer, M. Kalbac, S. Yang, L. Dunsch, B. Büchner, D. Batchelor, and H. Kataura, Phys. Rev. B **73**, 205411 (2006) 10.1103/physrevb.73.205411.

[51]R. Kitaura, H. Okimoto, H. Shinohara, T. Nakamura, and H. Osawa, Phys. Rev. B **76**, 172409 (2007) 10.1103/physrevb.76.172409.

[52]R. J. Nicholls, K. Sader, J. H. Warner, S. R. Plant, K. Porfyrakis, P. D. Nellist, G. A. D. Briggs, and D. J. H. Cockayne, ACS Nano **4**, 3943–3948 (2010) 10.1021/nn100823e.

[53]W. L. Yang, Z. Y. Xu, H. Wei, M. Feng, and D. Suter, Phys. Rev. A **81**, 032303 (2010) 10.1103/physreva.81.032303.

[54]F. Fritz, R. Westerström, A. Kostanyan, C. Schlesier, J. Dreiser, B. Watts, L. Houben, M. Luysberg, S. M. Avdoshenko, A. A. Popov, et al., Nanotechnology **28**, 435703 (2017) 10.1088/1361-6528/aa8b4c.

[55]A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner, Nature **545**, 462–466 (2017) 10.1038/nature22362.

[56]M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, T. Menke, D. Borgnia, P. M. Preiss, F. Grusdt, A. M. Kaufman, and M. Greiner, Nature **546**, 519–523 (2017) 10.1038/nature22811.

[57]A. V. Gorshkov, S. R. Manmana, G. Chen, J. Ye, E. Demler, M. D. Lukin, and A. M. Rey, Phys. Rev. Lett. **107**, 115301 (2011) 10.1103/physrevlett.107.115301.

[58]S. R. White, Phys. Rev. Lett. **69**, 2863 (1992) 10.1103/physrevlett.69.2863.

[59]S. R. White and R. L. Martin, J. Chem. Phys. **110**, 4127–4130 (1999) 10.1063/1.478295.

[60]C. Duperrouzel, P. Tecmer, K. Boguslawski, G. Barcza, Ö. Legeza, and P. W. Ayers, Chem. Phys. Lett. **621**, 160–164 (2015) 10.1016/j.cplett.2015.01.005.

[61]E. M. Stoudenmire and S. R. White, Annu. Rev. Condens. Matter Phys. **3**, 111–128 (2012) 10.1146/annurev-conmatphys-020911-125018.

[62]*ITensor library (version 2.1.1) http://itensor.org.*

[63]J. Ruhman, E. G. Dalla Torre, S. D. Huber, and E. Altman, Phys. Rev. B **85**, 125121 (2012) 10.1103/physrevb.85.125121.

[64]S. R. Manmana, E. M. Stoudenmire, K. R. A. Hazzard, A. M. Rey, and A. V. Gorshkov, Phys. Rev. B **87**, 081106 (2013) 10.1103/physrevb.87.081106.

[65]M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013), ISBN: 978-1-107-03425-9.

[66]M. B. Hastings, J. Stat. Mech.: Theory Exp. **2007**, P08024 (2007) 10.1088/1742-5468/2007/08/p08024.

[67]Z.-X. Gong, M. Foss-Feig, F. G. Brandão, and A. V. Gorshkov, Phys. Rev. Lett. **119**, 050501 (2017) 10.1103/physrevlett.119.050501.

[68]P. Calabrese and J. Cardy, J. Phys. A: Math. Theor. **42**, 504005 (2009) 10.1088/1751-8113/42/50/504005.

[69]L. E. Ballentine, *Quantum Mechanics* (Prentice-Hall, 1990), ISBN: 0-13-747932-8.

[70]Q. H. Liu, D. M. Xun, and L. Shan, Int. J. Theor. Phys. **49**, 2164–2171 (2010) 10.1007/s10773-010-0403-5.

[71]S. Sachdev, *Quantum Phase Transitions*, 2nd ed. (Cambridge University Press, 2011), ISBN: 0-521-58254-7.

[72]J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, 1996), ISBN: 0-521-49959-3.

[73]M. A. Robb, F. Bernardi, and M. Olivucci, Pure Appl. Chem. **67**, 783–789 (1995) 10.1351/pac199567050783.

[74]A. M. Wodtke, J. C. Tully, and D. J. Auerbach, Int. Rev. Phys. Chem. **23**, 513–539 (2004) 10.1080/01442350500037521.

[75]G. A. Worth and L. S. Cederbaum, Annu. Rev. Phys. Chem. **55**, 127–158 (2004) 10.1146/annurev.physchem.55.091602.094335.

[76]P. Goel and M. Nooijen, Phys. Chem. Chem. Phys. **18**, 11263–11277 (2016) 10.1039/C5CP07889J.

[77]S. Hahn and G. Stock, J. Phys. Chem. B **104**, 1146–1149 (2000) 10.1021/jp992939g.

[78]A. L. Sobolewski, W. Domcke, and C. Hättig, Proc. Natl. Acad. Sci. U.S.A. **102**, 17903–17906 (2005) 10.1073/pnas.0504087102.

[79]G. A. Worth, H. Köppel, E. Gindensperger, and L. S. Cederbaum, "Multidimensional Non-Adiabatic Dynamics", in *Multidimensional Quantum Dynamics: MCTDH Theory and Applications* (Wiley Online Library, 2009) Chap. 18, pp. 209–230, 10.1002/9783527627400.ch18.

[80]N. Ananth, J. Chem. Phys. **139**, 124102 (2013) 10.1063/1.4821590.

[81]A. R. Menzeleev, F. Bell, and T. F. Miller III, J. Chem. Phys. **140**, 064103 (2014) 10.1063/1.4863919.

[82]J. R. Duke and N. Ananth, Faraday Discuss. **195**, 253–268 (2017) 10.1039/C6FD00123H.

[83]X. Liu and J. Liu, J. Chem. Phys. **148**, 102319 (2018) 10.1063/1.5005059.

[84]C. Joy, P. P. Boyle, and K. S. Tan, Manag. Sci. **42**, 926–938 (1996) 10.1287/mnsc.42.6.926.

[85]R. E. Caflisch, Acta Numer. **7**, 1–49 (1998) 10.1017/S0962492900002804.

[86]H. Chi, P. Beerli, D. W. Evans, and M. Mascagni, in International Conference on Computational Science (Springer, 2005), pp. 775–782, 10.1007/11428862_105.

[87]S. E. Brown, I. Georgescu, and V. A. Mandelshtam, J. Chem. Phys. **138**, 044317 (2013) 10.1063/1.4788977.

[88]I. Georgescu, S. Jitomirskaya, and V. A. Mandelshtam, J. Chem. Phys. **139**, 204104 (2013) 10.1063/1.4829836.

[89]H. Niederreiter, Bull. Am. Math. Soc. **84**, 957–1041 (1978) 10.1090/S0002-9904-1978-14532-7.

[90]J. R. Birge, *Quasi-Monte Carlo approaches to option pricing*, tech. rep. 94-19 (Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, Michigan, 1995).

[91]I. M. Sobol', Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki (J. Comp. Math. Math. Phys.) **7**, 784–802 (1967) 10.1016/0041-5553(67)90144-9.

[92]W. J. Morokoff and R. E. Caflisch, J. Comp. Phys. **122**, 218–230 (1995) 10.1006/jcph.1995.1209.

[93]S. G. Johnson, *Sobol: The Sobol module for Julia*, https://github.com/stevengj/Sobol.jl, accessed 2019-05-08.

[94]P. A. Acworth, M. Broadie, and P. Glasserman, "A comparison of some Monte Carlo and quasi Monte Carlo techniques for option pricing", in *Monte Carlo and Quasi-Monte Carlo Methods 1996* (Springer, 1998), pp. 1–18, 10.1007/978-1-4612-1690-2_1.

[95]A. B. Owen, in Monte Carlo Ray Tracing: Siggraph 2003 Course 44 (2003), pp. 69–88.

[96]B. Tuffin, Monte Carlo Methods Appl. **10**, 617–628 (2004) 10.1515/mcma.2004.10.3-4.617.

[97]G. Ökten and W. Eastman, J. Econ. Dyn. Control **28**, 2399–2426 (2004) 10.1016/j.jedc.2003.11.003.

[98]P. Young, *Everything you wanted to know about data analysis and fitting but were afraid to ask* (Springer, 2015), 10.1007/978-3-319-19051-8.

[99]E. A. Youngs and E. M. Cramer, Technometrics **13**, 657–665 (1971) 10.1080/00401706.1971.10488826.

[100]T. F. Chan, G. H. Golub, and R. J. LeVeque, Am. Stat. **37**, 242–247 (1983) 10.1080/00031305.1983.10483115.

[101]S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79–86 (1951) 10.1214/aoms/1177729694.

[102]C. E. Shannon, Bell Syst. Tech. **27**, 379–423 (1948) 10.1002/j.1538-7305.1948.tb01338.x.

[103]D. Iouchtchenko, *VibronicToolkit*, https://github.com/0/VibronicToolkit.jl, accessed 2019-05-08.

[104]M. Matsumoto and T. Nishimura, ACM T. Model. Comput. S. **8**, 3–30 (1998) 10.1145/272991.272995.

[105]G. E. Box and M. E. Muller, Ann. Math. Stat. **29**, 610–611 (1958) 10.1214/aoms/1177706645.

[106]G. Marsaglia and W. W. Tsang, J. Stat. Soft. **5**, 1–7 (2000) 10.18637/jss.v005.i08.

[107]G. M. Torrie and J. P. Valleau, J. Comp. Phys. **23**, 187–199 (1977) 10.1016/0021-9991(77)90121-8.

[108]B. Roux, Comp. Phys. Comm. **91**, 275–282 (1995) 10.1016/0010-4655(95)00053-I.

[109]E. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472–477 (1989) 10.1016/S0009-2614(89)87314-2.

[110]M. Sprik and G. Ciccotti, J. Chem. Phys. **109**, 7737–7744 (1998) 10.1063/1.477419.

[111]G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, ChemPhysChem **6**, 1809–1814 (2005) 10.1002/cphc.200400669.

[112]A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562–12566 (2002) 10.1073/pnas.202427399.

[113]G. Bussi, A. Laio, and M. Parrinello, Phys. Rev. Lett. **96**, 090601 (2006) 10.1103/PhysRevLett.96.090601.

[114]T. Mülders, P. Krüger, W. Swegat, and J. Schlitter, J. Chem. Phys. **104**, 4869–4870 (1996) 10.1063/1.471182.

[115]W. K. den Otter and W. J. Briels, Mol. Phys. **98**, 773–781 (2000) 10.1080/00268970009483348.

[116]M. Ceriotti, W. Fang, P. G. Kusalik, R. H. McKenzie, A. Michaelides, M. A. Morales, and T. E. Markland, Chem. Rev. **116**, 7529–7550 (2016) 10.1021/acs.chemrev.5b00674.

[117]T. E. Markland and M. Ceriotti, Nat. Rev. Chem. **2**, 0109 (2018) 10.1038/s41570-017-0109.

[118]C. L. Vaillant, S. C. Althorpe, and D. J. Wales, J. Chem. Theory Comput. **15**, 33–42 (2018) 10.1021/acs.jctc.8b00675.

[119]E. Méndez and D. Laria, J. Chem. Phys. **153**, 054302 (2020) 10.1063/5.0016122.

[120]K. Hinsen and B. Roux, J. Chem. Phys. **106**, 3567–3577 (1997) 10.1063/1.473439.

[121]D. T. Major, M. Garcia-Viloca, and J. Gao, J. Chem. Theory Comput. **2**, 236–245 (2006) 10.1021/ct050257t.

[122]D. T. Major and J. Gao, J. Chem. Theory Comput. **3**, 949–960 (2007) 10.1021/ct600371k.

[123]B. Walker and A. Michaelides, J. Chem. Phys. **133**, 174306 (2010) 10.1063/1.3505038.

[124]A. Vardi-Kilshtain, D. T. Major, A. Kohen, H. Engel, and D. Doron, J. Chem. Theory Comput. **8**, 4786–4796 (2012) 10.1021/ct300628e.

[125]J. Cao and G. A. Voth, J. Chem. Phys. **100**, 5093–5105 (1994) 10.1063/1.467175.

[126]J. Cao and G. A. Voth, J. Chem. Phys. **101**, 6168–6183 (1994) 10.1063/1.468399.

[127]N. Blinov and P.-N. Roy, J. Chem. Phys. **120**, 3759–3764 (2004) 10.1063/1.1642600.

[128]B. Leaf, Found. Phys. **10**, 581–599 (1980) 10.1007/BF00715041.

[129]R. Wilcox, J. Math. Phys. **8**, 962–982 (1967) 10.1063/1.1705306.

[130]D. Iouchtchenko, *SnaggedNecklace*, https://github.com/0/SnaggedNecklace.jl, accessed 2021-01-03.

[131]Wolfram Research Inc., *Hermite polynomials: Summation (formula 05.01.23.0013)*, http://functions.wolfram.com/05.01.23.0013.01, accessed 2020-01-20.

[132]D. Thirumalai, E. J. Bruskin, and B. J. Berne, J. Chem. Phys. **79**, 5063–5069 (1983) 10.1063/1.445601.

[133]R. L. Liboff, I. Nebenzahl, and H. H. Fleischmann, Am. J. Phys. **41**, 976–980 (1973) 10.1119/1.1987445.

[134]G. Paz, J. Phys. A **35**, 3727 (2002) 10.1088/0305-4470/35/16/311.

[135]S. Reich, SIAM J. Numer. Anal. **33**, 475–491 (1996) 10.1137/0733025.

[136]H. C. Andersen, J. Comput. Phys. **52**, 24–34 (1983) 10.1016/0021-9991(83)90014-1.

[137]J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, J. Comp. Phys. **23**, 327–341 (1977) 10.1016/0021-9991(77)90098-5.

[138]M. Fixman, J. Chem. Phys. **69**, 1527–1537 (1978) 10.1063/1.436725.

[139]W. K. den Otter, J. Chem. Theory Comput. **9**, 3861–3865 (2013) 10.1021/ct400134d.

[140]T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations: A Mathematical Perspective* (Imperial College Press, 2010), ISBN: 978-1-84816-247-1.

[141]T. Lelièvre, M. Rousset, and G. Stoltz, Math. Comput. **81**, 2071–2125 (2012) 10.1090/S0025-5718-2012-02594-4.

[142]J. Liu, D. Li, and X. Liu, J. Chem. Phys. **145**, 024103 (2016) 10.1063/1.4954990.

[143]Z. Zhang, X. Liu, K. Yan, M. E. Tuckerman, and J. Liu, J. Phys. Chem. A **123**, 6056–6079 (2019) 10.1021/acs.jpca.9b02771.

[144]D. Iouchtchenko, *WaterMeanMeanForce*, https://github.com/0/WaterMeanMeanForce.jl, accessed 2021-01-03.

[145]F. Paesani, W. Zhang, D. A. Case, T. E. Cheatham III, and G. A. Voth, J. Chem. Phys. **125**, 184507 (2006) 10.1063/1.2386157.

[146]R. W. Cottle, Lin. Algebra Appl. **8**, 189–211 (1974) 10.1016/0024-3795(74)90066-4.

[147]N. D. Cahill, J. R. D'Errico, D. A. Narayan, and J. Y. Narayan, Coll. Math. J. **33**, 221–225 (2002) 10.1080/07468342.2002.11921945.

[148]A. J. Ferris and G. Vidal, Phys. Rev. B **85**, 165146 (2012) 10.1103/PhysRevB.85.165146.

[149]M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, Phys. Rep. **324**, 1–105 (2000) 10.1016/S0370-1573(99)00047-2.

[150]H. R. Larsson, J. Chem. Phys. **151**, 204102 (2019) 10.1063/1.5130390.

[151]P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, PLOS Comp. Biol. **13**, e1005659 (2017) 10.1371/journal.pcbi.1005659.

[152]V. Babin, C. Leforestier, and F. Paesani, J. Chem. Theory Comput. **9**, 5395–5403 (2013) 10.1021/ct400863t.