



UNIVERSITY OF  
**OXFORD**

# Fast Quasi-Centroid Molecular Dynamics

**THEO FLETCHER**

Balliol College  
University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*

**August 2023**



# Fast Quasi-Centroid Molecular Dynamics

THEO FLETCHER

University of Oxford

A thesis submitted for the degree of  
*Doctor of Philosophy*

Imaginary time path integrals are a proven way of accounting for quantum effects in theoretical chemistry. Despite considerable interest, and for a variety of reasons, no path integral method has been suggested that is able to robustly simulate vibrational dynamics. The best attempt so far is quasi-centroid molecular dynamics (QCMD), but this still has some major problems: it is quite slow, and the only gas-phase molecules it has been shown to work for so far are  $\text{H}_2\text{O}$  and  $\text{NH}_3$ . This thesis presents fast quasi-centroid molecular dynamics (f-QCMD), a new implementation that is simple, efficient and demonstrably applicable to more complex molecules. Furthermore, this thesis develops and tests two additional methods that compliment f-QCMD. The first is a simple way of correcting the intensities of overtone and combination bands in infrared spectroscopy, which trajectory-based methods like f-QCMD tend to get wrong. The second is an extremely versatile force-based technique for calculating any type of distribution function quickly and with arbitrary precision.

*August 2023*



## ACKNOWLEDGEMENTS

First and foremost, the utmost thanks goes to my supervisor, David Manolopoulos. I will always be grateful for him sharing his wealth of experience, knowledge, and opinions on Brexit.

Thank you also to the Manolopoulos group: Lachlan Lindoy, Joe Lawrence, Tom Fay and Ben Sutherland. Special thanks goes to Andrew Zhu for being the best part II student anyone could ask for. This thesis would not have been possible without you.

Thank you to EPSRC and all the staff at the TMCS, for supporting me financially and academically. It was a privilege to learn from such an esteemed group of individuals. To the students of cohort V, my friends, thank you for being such great company.

Thank you to my family for their unwavering support over the years.

Finally, words are not enough to thank my partner Alexandra, for giving me somewhere to stay through lockdown, and pretending not to notice that I was still there once it had ended. Thank you for your patience and for putting up with me all these years; you always said the things I needed to hear, and I wouldn't be anywhere without you.



# CONTENTS

<b>Spectroscopic Quantities</b>	<b>ix</b>
<b>Notation</b>	<b>x</b>
<b>Introduction</b>	<b>xi</b>
<b>1 Background and Theory</b>	<b>1</b>
1.1 Infrared Spectroscopy . . . . .	2
1.2 Classical Dynamics . . . . .	5
1.3 Path Integral Methods . . . . .	6
1.4 Other Methods . . . . .	13
1.5 Overtone and Combination Intensity . . . . .	16
1.6 Summary of Methods . . . . .	20
<b>2 Quasi-Centroid Molecular Dynamics</b>	<b>23</b>
2.1 The Centroid Formalism . . . . .	24
2.2 The Curvature Problem . . . . .	28
2.3 Polar Constraints . . . . .	33
2.4 The Quasi-Centroid Formalism . . . . .	37
2.5 Computational Details . . . . .	41
<b>3 Fast Quasi-Centroid Molecular Dynamics</b>	<b>49</b>
3.1 Choosing Coordinates . . . . .	50
3.2 Implementing Centroid Molecular Dynamics . . . . .	53
3.3 Coarse-Graining . . . . .	61
<b>4 Application to Gas-Phase Molecules</b>	<b>69</b>
4.1 Computational Details . . . . .	71
4.2 H <sub>2</sub> O . . . . .	74
4.3 HDO . . . . .	78
4.4 NH <sub>3</sub> . . . . .	80
4.5 CH <sub>4</sub> . . . . .	88
4.6 H <sub>2</sub> O <sub>2</sub> . . . . .	91

<b>5</b>	<b>Estimators for Distribution Functions</b>	<b>99</b>
5.1	Reduced Variance Estimators . . . . .	100
5.2	Classical Estimators . . . . .	105
5.3	Path Integral Estimators . . . . .	106
5.4	Distribution Functions for Molecules . . . . .	110
5.5	Summary of Equations . . . . .	120
	<b>Conclusions</b>	<b>123</b>
	<b>References</b>	<b>131</b>

## APPENDICES

<b>A</b>	<b>Intensity Correction Factor</b>	<b>141</b>
A.1	Dipole Moment Matrix Elements . . . . .	142
A.2	Allowed Transitions . . . . .	146
A.3	Integrated Intensities . . . . .	148
<b>B</b>	<b>Iterative Boltzmann Inversion Convergence</b>	<b>157</b>
B.1	H <sub>2</sub> O . . . . .	159
B.2	HDO . . . . .	160
B.3	NH <sub>3</sub> . . . . .	162
B.4	CH <sub>4</sub> . . . . .	164
B.5	H <sub>2</sub> O <sub>2</sub> . . . . .	166
<b>C</b>	<b>Reduced Variance Estimators</b>	<b>171</b>
C.1	Jacobian Derivation . . . . .	171
C.2	Implementation . . . . .	174

## SPECTROSCOPIC QUANTITIES

Quantities commonly used to describe spectroscopic transitions, and their relationship with each other.

Quantity	Symbol	$E_k - E_j =$
Energy	$E$	$\Delta E_{kj}$
Frequency	$\nu$	$h\nu_{kj}$
Angular Frequency	$\omega$	$\hbar\omega_{kj}$
Wavenumber	$\tilde{\nu}$	$hc\tilde{\nu}_{kj}$

Unless specified otherwise, all spectra are given in arbitrary intensity units, such that the integrated area of the spectrum is unity.

# NOTATION

## Ring polymer coordinates

Applies to momenta,  $\mathbf{p}$ , and positions,  $\mathbf{q}$ .

---

$\mathbf{q}$      An  $N$ -dimensional position vector.

$q_i$      The  $i^{\text{th}}$  element of  $\mathbf{q}$ .

---

$\mathbf{q}^{(j)}$      Position vector of the  $j^{\text{th}}$  ring polymer bead.

$\underline{\mathbf{q}}$      Shorthand for  $\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(n)}\}$ .

---

$\tilde{\mathbf{q}}^{(k)}$      Position vector of the  $k^{\text{th}}$  ring polymer normal mode.

$\underline{\tilde{\mathbf{q}}}$      Shorthand for  $\{\tilde{\mathbf{q}}^{(0)}, \tilde{\mathbf{q}}^{(1)}, \dots, \tilde{\mathbf{q}}^{(n-1)}\}$ .

---

For a general position dependent function,  $f(\mathbf{q})$ :

$$f^{(j)} = f(\mathbf{q}^{(j)})$$

$$\bar{f}(\underline{\mathbf{q}}) = \frac{1}{n} \sum_{j=1}^n f^{(j)}$$

## Coordinate transformation

---

$\boldsymbol{\xi}(\mathbf{q})$      A generic coordinate transformation.

---

$\Xi = \boldsymbol{\xi}(\mathbf{Q}),$

i.e.: The transformation of the vector  $\mathbf{Q}$ .

---

$J(\boldsymbol{\xi}) = \left| \left| \frac{\partial(q_1, \dots, q_N)}{\partial(\xi_1, \dots, \xi_N)} \right| \right|,$

i.e.: The Jacobian for the  $\mathbf{q} \rightarrow \boldsymbol{\xi}$  transformation.

---

## INTRODUCTION

Accurately describing the dynamics of molecules is of fundamental importance to computational chemistry. It is also incredibly difficult. Quantum mechanics has significant consequences for molecular vibrations, but full-scale quantum calculations are far too expensive for anything but the very smallest systems.

The most promising idea for solving this problem is to use (imaginary time) path integrals. Specific details vary, but the general concept here is to create several interacting replicas of the classical system, which can be seen as representing quantum dispersion.<sup>13</sup> This approach has already been used to accurately and efficiently include quantum effects in studies of diffusion coefficients,<sup>16,43</sup> reaction rates<sup>17–20,42</sup> and more.<sup>15,41</sup> Naturally, recent years have seen a lot of interest in developing path integral methods to model the dynamics of molecules, specifically vibrational spectra. However, everything so far has had some sort of critical limitation and the problem remains unsolved.<sup>2,21,22</sup>

For example, ring polymer molecular dynamics<sup>15</sup> generally gives good predictions of low-frequency band positions and shapes. However, vibrations of the normal modes of the ring polymer can contaminate the spectrum at higher frequencies.<sup>2,21</sup> The attempt at fixing this was to attach a thermostat to the ring polymer normal modes, to damp out the vibrations. While this did remove the spurious frequencies from the spectrum, it also resulted in a fairly significant broadening of the spectral band shape.<sup>22,23</sup> To make matters worse, as with all path integral methods, the predicted intensities of overtone and combination bands are usually completely wrong.<sup>31,32</sup>

Another noteworthy example is centroid molecular dynamics, or CMD, which

focuses on the dynamics of the Cartesian centroid of the path integral.<sup>25,39</sup> The main issue here is that the Cartesian centroid of a path integral tends to have shorter bonds than the actual replicas.<sup>21,45</sup> This so called ‘curvature problem’ causes CMD to severely underestimate the frequency of bond stretching at low temperatures.<sup>9,21,23,26,44,45</sup>

The response to this was quasi-centroid molecular dynamics, or QCMD: a modification to CMD that uses a curvilinear ‘quasi-centroid’ instead of the usual Cartesian average.<sup>26</sup> This avoids the curvature problem, but unfortunately, using curvilinear coordinates also introduces a new set of issues. For one, the forces are much more complicated. This means that the various tricks that make CMD a viable tool<sup>62,63,67–69,72</sup> do not always translate well to QCMD, and a like-for-like calculation can be up to 8 times slower.<sup>26,65</sup> More fundamentally, the method in its current state is already pushed to its limits. In the gas-phase, QCMD has only ever been applied to H<sub>2</sub>O and NH<sub>3</sub> and it is still unclear whether it can even be used, for example, on CH<sub>4</sub>.<sup>8</sup>

Although these are only a few simple examples, QCMD has been consistently more accurate than every other path integral approach.<sup>9,24,26</sup> The method definitely has potential, but its issues and limited scope mean it still has a way to go before it can be used practically. This thesis gives QCMD the attention it deserves, and presents a fast, simple implementation that can be applied to any molecule.

Fast quasi-centroid molecular dynamics (f-QCMD) avoids the complicated task of calculating forces from a path integral simulation. Instead, it produces the actual effective potential energy surface that the dynamics is run on. This only requires a single short path integral calculation, and the effective quasi-centroid potential is then known once-and-for-all.

In short, f-QCMD works by constructing a classical potential that reproduces distribution functions of path integrals at equilibrium. These can be difficult to obtain precisely with the standard histogram approach, because the statistical error

is linked to the resolution of the calculation. Based on recent studies,<sup>102–104</sup> this thesis presents a new way of generating distribution functions rapidly and to an arbitrary level of precision. This completely general and highly versatile method can be applied to a wide range of problems, including classical and path integral systems, as well as any type of distribution function (e.g.: bond length, bond angle, torsion angle, etc).

The final piece of the puzzle is how to correct the intensities of overtone and combination bands, which are still predicted incorrectly by f-QCMD. Fortunately, a possible solution has recently been published: a simple correction factor that can be applied as a final step to the calculated spectrum.<sup>31,32</sup> As with QCMD, this has worked well in a few cases, but its performance for more complex systems is still yet to be demonstrated.

The following is a brief outline of the rest of the thesis:

Chapter 1 gives a complete background on the essential theory behind predicting vibrational spectra. This includes exact quantum mechanics, how path integrals approximate this, and why these methods often go wrong. It also introduces the intensity correction factor, the derivation of which is given in appendix A. A wide range of approximate methods are then tested and compared, giving a comprehensive picture of the current state of vibrational dynamics simulations.

Chapter 2 takes a closer look at CMD and the so-called curvature problem that makes it unsuitable for vibrational dynamics. This leads onto QCMD, and explores how it is more than just an *ad hoc* fix; it turns out that it is actually a more general version of CMD, and has some interesting theoretical links to quantum mechanics.

Chapter 3 delves further into CMD, this time looking at how it is implemented, and why the same ideas can not always be translated to QCMD. It also examines the more fundamental problem of actually finding the quasi-centroid. These problems motivate the new implementation, f-QCMD, which is presented here.

Chapter 4 uses f-QCMD to generate spectra for a range of gas-phase molecules,

going well beyond the limits of previous studies. This provides a comprehensive test of the methods that have been developed, and allows for a critical assessment on whether we finally have a robust and accurate way to predict the dynamics of molecules.

Chapter 5 demonstrates how distribution functions can be calculated quickly and precisely. The actual framework here is completely general, but particular focus is given to the specific cases most relevant to f-QCMD.

## CHAPTER 1

# BACKGROUND AND THEORY

Accurately describing the dynamics of molecules is immensely difficult. The reason for this is that a completely faithful description requires quantum mechanics, which becomes exponentially more complex as the size of the system grows. As a result, for most applications of any interest to chemistry, solving pure quantum mechanics is completely impossible, and alternative approximations are necessary to make any progress.

As chemists, vibrational (infrared) spectroscopy is one of the best tools at our disposal to investigate the dynamics of chemical systems. In fact, accurately simulating infrared spectra is the key focus of this thesis, and this chapter presents the underlying theory required for this.

Starting from the ideal quantum description, a number of approximate methods are introduced and critiqued, with direct comparisons of example spectra. Of particular note are a class of imaginary time path integral methods, which are a tried-and-tested solution to quantum dynamics in many other applications. When it comes to spectroscopy, however, they nearly all have severe problems. As this thesis goes some way to rectifying this, particular focus is given to these approaches.

It also turns out that while a decent approximation to vibrational frequencies might be possible, the intensity of the spectrum is another problem entirely, and a naive approach can get this badly wrong. Fortunately, a simple correction factor can be applied as a finishing touch. This is described briefly, with a full derivation given in appendix A.

### 1.1 INFRARED SPECTROSCOPY

Consider a quantum mechanical system with a Hamiltonian

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

where  $\hat{\mathbf{p}}$  and  $\hat{\mathbf{q}}$  are the momentum and position operators respectively, and  $V(\mathbf{q})$  is the potential energy surface. There are  $N$  degrees of freedom, each of which is Cartesian and has mass  $m_i$ .

Suppose the system has a volume  $v$ , and is at thermal equilibrium at an inverse temperature  $\beta = 1/k_B T$ . Using Fermi's Golden Rule, the infrared (IR) absorption spectrum of this system is<sup>1,2</sup>

$$n(\omega)\alpha(\omega) = \frac{\pi\beta}{3c\epsilon_0 v} \frac{1}{2\pi} \int_{-\infty}^{\infty} dt e^{-i\omega t} \tilde{C}(t), \quad (1.2)$$

where  $n(\omega)$  is the refractive index and  $\alpha(\omega)$  is the Beer-Lambert absorption coefficient. Fundamentally, the spectrum is the Fourier transform of the Kubo-transformed dipole-derivative autocorrelation function,

$$\tilde{C}(t) = \frac{1}{Z\beta} \int_0^\beta d\lambda \operatorname{tr} \left[ \left( e^{-(\beta-\lambda)\hat{H}} \hat{\boldsymbol{\mu}} e^{-\lambda\hat{H}} \right) \cdot \left( e^{i\hat{H}t/\hbar} \hat{\boldsymbol{\mu}} e^{-i\hat{H}t/\hbar} \right) \right], \quad (1.3)$$

where  $Z$  is the partition function,

$$Z = \operatorname{tr} \left[ e^{-\beta\hat{H}} \right], \quad (1.4)$$

$\hat{\boldsymbol{\mu}}$  is the dipole moment operator and its time derivative is

$$\dot{\hat{\boldsymbol{\mu}}} = \frac{i}{\hbar} [\hat{H}, \hat{\boldsymbol{\mu}}]. \quad (1.5)$$

Evaluating Eq. (1.2) quantum mechanically requires computation that scales exponentially with the number of system degrees of freedom. As a result, generating a purely quantum mechanical spectrum is impossible for anything beyond the smallest gas-phase molecules. In these cases however,  $\tilde{C}(t)$  is rarely calculated directly. Instead, expanding  $\tilde{C}(t)$  in the basis of eigenstates of  $\hat{H}$  transforms Eq. (1.2) into

$$n(\omega)\alpha(\omega) = \frac{\pi}{3c\epsilon_0 v \hbar Z} \sum_{jk} \omega_{kj} e^{-\beta E_j} (1 - e^{-\beta \hbar \omega_{kj}}) |\langle j | \hat{\boldsymbol{\mu}} | k \rangle|^2 \delta(\omega - \omega_{kj}). \quad (1.6)$$

Finally, partially summing over sets of degenerate states,

$$\begin{aligned} n(\omega)\alpha(\omega) &= \frac{\pi}{3c\epsilon_0 v \hbar Z} \sum_{if} \sum_{j \in i} \sum_{k \in f} \omega_{kj} e^{-\beta E_j} (1 - e^{-\beta \hbar \omega_{kj}}) |\langle j | \hat{\boldsymbol{\mu}} | k \rangle|^2 \delta(\omega - \omega_{kj}), \\ &= \frac{\pi}{3c\epsilon_0 v \hbar Z} \sum_{if} \omega_{fi} e^{-\beta E_i} (1 - e^{-\beta \hbar \omega_{fi}}) \delta(\omega - \omega_{fi}) \sum_{j \in i} \sum_{k \in f} |\langle j | \hat{\boldsymbol{\mu}} | k \rangle|^2, \end{aligned}$$

gives<sup>3</sup>

$$n(\omega)\alpha(\omega) = \frac{1}{v} \sum_{if} S_{fi} \delta(\omega - \omega_{fi}), \quad (1.7)$$

where  $S_{fi}$  is the absorption line strength for a transition between the energy levels  $E_i$  and  $E_f$ ,

$$S_{fi} = \frac{\pi}{3c\epsilon_0 \hbar} \frac{e^{-\beta E_i}}{Z} (1 - e^{-\beta \hbar \omega_{fi}}) \omega_{fi} \sum_{j \in i} \sum_{k \in f} |\langle j | \hat{\boldsymbol{\mu}} | k \rangle|^2. \quad (1.8)$$

For a given system, the set of energy levels and line strengths can be calculated and tabulated once-and-for-all, allowing for the quick recovery of the complete spectrum at any temperature.<sup>4-6</sup>

In practice, and due to a variety of effects,<sup>7</sup> spectral lines are not infinitely sharp. This is modelled by the convolution of a normalised line shape function,  $F(\omega)$ , onto the spectrum<sup>6</sup>

$$n(\omega)\alpha(\omega) = \frac{1}{v} \sum_{if} S_{fi} F(\omega - \omega_{fi}). \quad (1.9)$$

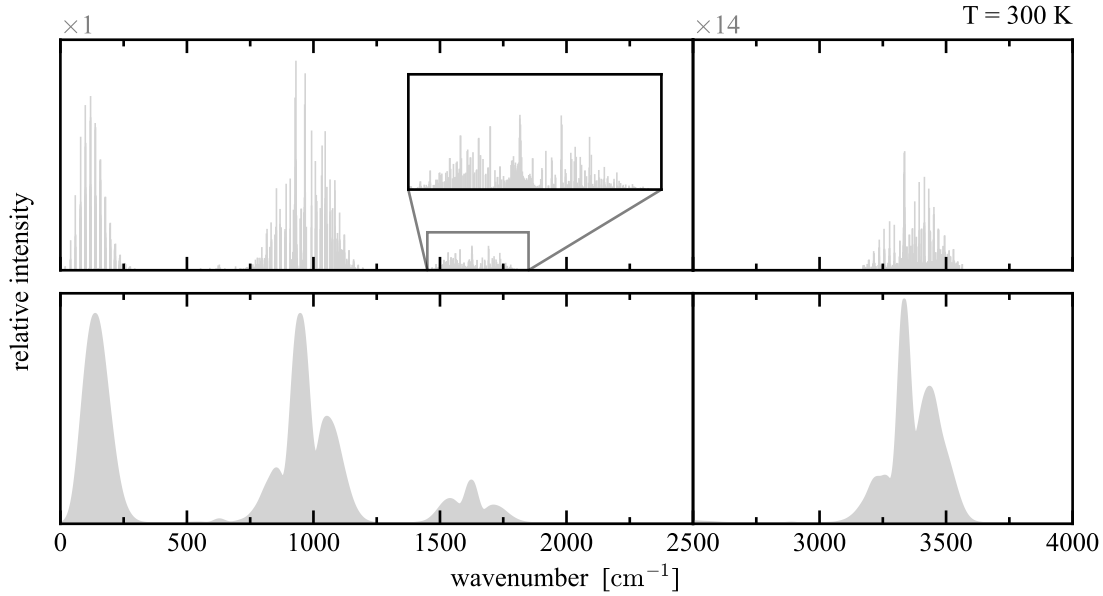
Assuming that  $F(\omega)$  is identical for all lines, this has the equivalent effect of damping the autocorrelation function,

$$n(\omega)\alpha(\omega) = \frac{\pi\beta}{3c\epsilon_0 v} \frac{1}{2\pi} \int_{-\infty}^{\infty} dt e^{-i\omega t} \tilde{C}(t) f(t), \quad (1.10)$$

where the damping function  $f(t)$  is

$$f(t) = \int_{-\infty}^{\infty} d\omega e^{i\omega t} F(\omega). \quad (1.11)$$

The effect of a line profile on a spectrum is shown in fig. 1.1.



**Figure 1.1:** The IR spectrum of gas-phase NH<sub>3</sub> at 300 K, calculated with Eq. (1.6) (top), and with a line shape function (bottom). Panels are scaled in the ratio 1:14.

See section 1.6 for more details.

## 1.2 CLASSICAL DYNAMICS

Exponential scaling limits the scope of purely quantum mechanical calculations. This issue can be avoided by evaluating Eq. (1.3) using classical dynamics instead,

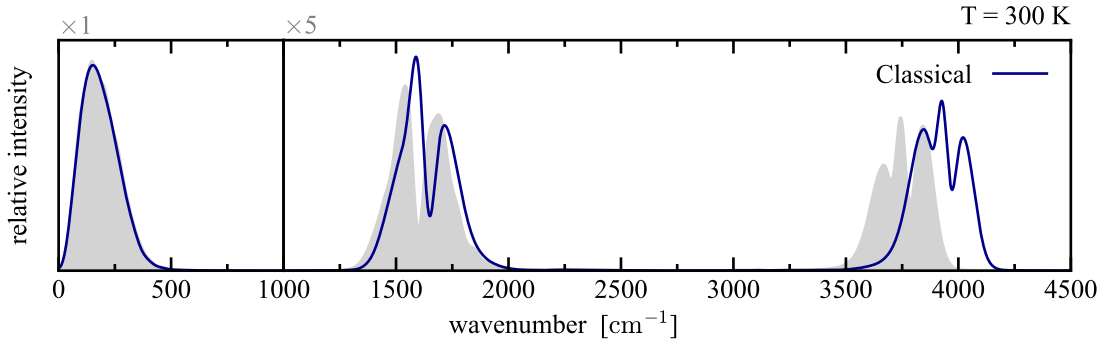
$$\tilde{C}(t) \approx \frac{1}{Z_1} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p}, \mathbf{q})} \dot{\boldsymbol{\mu}}(\mathbf{p}, \mathbf{q}) \cdot \dot{\boldsymbol{\mu}}(\mathbf{p}(t), \mathbf{q}(t)), \quad (1.12)$$

where the classical partition function is

$$Z_1 = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p}, \mathbf{q})}. \quad (1.13)$$

Classical dynamics achieves linear scaling with system size, but at the cost of neglecting quantum effects. In vibrational spectroscopy, this typically manifests as blue-shifted bands in the classical spectrum relative to exact results,<sup>8–10</sup> as demonstrated in fig. 1.2.

It should also be noted that there are many ways to formulate a quantum correlation function, which would change the overall form of Eq. (1.2).<sup>11</sup> However, in many ways, the Kubo-transformed quantum correlation function can be considered as the most classical choice, making Eq. (1.2) as written the correct form for classical spectra. For one, the Kubo-transformed correlation function is necessarily real and a symmetric function of time, and is identical to its classical counterpart in the



**Figure 1.2:** The quantum (shaded) and classical spectrum of gas-phase  $\text{H}_2\text{O}$  at 300 K. Panels are scaled in the ratio 1:5. See section 1.6 for more details.

case of a harmonic potential and linear dipole moment.<sup>2,12</sup> Furthermore, Eq. (1.2) as written is independent of  $\hbar$ , which naturally lends itself to the classical ( $\hbar \rightarrow 0$ ) limit.

### 1.3 PATH INTEGRAL METHODS

Imaginary time path integral methods are an intermediate between quantum and classical calculations, recovering much of the lost (nuclear) quantum effects while maintaining efficient linear computational complexity. The justification behind these approaches comes from a Trotter discretisation of  $e^{-\beta\hat{H}}$ , the quantum Boltzmann operator.<sup>13</sup> This results in  $n$  replicas of the classical system, producing an extended phase space  $\underline{\mathbf{p}} = \{\mathbf{p}^{(1)}, \dots, \mathbf{p}^{(n)}\}$  and  $\underline{\mathbf{q}} = \{\mathbf{q}^{(1)}, \dots, \mathbf{q}^{(n)}\}$ , at inverse temperature  $\beta_n = \beta/n$ . These replicas are connected into a ring, by temperature dependent harmonic springs,

$$H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}}; \beta) = \sum_{j=1}^n H(\mathbf{p}^{(j)}, \mathbf{q}^{(j)}) + \sum_{j=1}^n \sum_{i=1}^N \frac{1}{2} m_i \omega_n^2 (q_i^{(j)} - q_i^{(j+1)})^2, \quad (1.14)$$

where  $\omega_n = 1/\beta_n \hbar$ . Because the replicas are connected in a ring,  $\mathbf{p}^{(j+n)} = \mathbf{p}^{(j)}$  and  $\mathbf{q}^{(j+n)} = \mathbf{q}^{(j)}$ , and the complete set of replicas is referred to as a ring polymer.

The above Hamiltonian reflects the picture of a ring polymer as a set of connected classical replicas, or ‘beads’. However, it is often convenient to instead use the normal mode coordinates,  $\tilde{\mathbf{p}}$  and  $\tilde{\mathbf{q}}$ , such that

$$H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}}; \beta) = \sum_{k=0}^{n-1} \sum_{i=1}^N \left[ \frac{(\tilde{p}_i^{(k)})^2}{2m_i} + \frac{1}{2} m_i (\omega^{(k)})^2 (\tilde{q}_i^{(k)})^2 \right] + \sum_{j=1}^n V(\mathbf{q}^{(j)}) \quad (1.15)$$

where  $\omega^{(k)}$  is the frequency of the  $k^{\text{th}}$  normal mode,

$$\omega^{(k)} = 2\omega_n \sin\left(\frac{k\pi}{n}\right). \quad (1.16)$$

The normal mode coordinates of a ring polymer can be obtained by an orthogonal transformation of the original bead representation

$$\tilde{\mathbf{p}}^{(k)} = \sum_{j=1}^n C_{jk} \mathbf{p}^{(j)}, \quad \tilde{\mathbf{q}}^{(k)} = \sum_{j=1}^n C_{jk} \mathbf{q}^{(j)}, \quad (1.17a)$$

$$\mathbf{p}^{(j)} = \sum_{k=0}^{n-1} C_{jk} \tilde{\mathbf{p}}^{(k)}, \quad \mathbf{q}^{(j)} = \sum_{k=0}^{n-1} C_{jk} \tilde{\mathbf{q}}^{(k)}, \quad (1.17b)$$

where, for even  $n$ , the elements  $C_{jk}$  are given by<sup>14</sup>

$$C_{jk} = \begin{cases} \sqrt{1/n}, & k = 0, \\ \sqrt{2/n} \cos(2\pi jk/n), & 0 < k < n/2, \\ \sqrt{1/n} (-1)^j, & k = n/2, \\ \sqrt{2/n} \sin(2\pi jk/n), & n/2 < k < n. \end{cases} \quad (1.18)$$

A key observation here is that the 0<sup>th</sup> normal mode

$$\tilde{\mathbf{p}}^{(0)} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{p}^{(j)}, \quad \tilde{\mathbf{q}}^{(0)} = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{q}^{(j)}, \quad (1.19)$$

is proportional to the centroid coordinates of the ring polymer,

$$\bar{\mathbf{p}} = \frac{1}{n} \sum_{j=1}^n \mathbf{p}^{(j)}, \quad \bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}^{(j)}. \quad (1.20)$$

A qualitative explanation of how ring polymers capture quantum effects follows from examining  $\omega_n$ : the strength of the springs between the replicas. At low temperatures, the springs become weaker, and the resulting expansion of the polymer represents quantum dispersion. Increasing the temperature causes the springs to stiffen, and the polymer collapses onto a single point, recovering the appropriate classical description. In fact, the  $n = 1$  limit of Eq. (1.14) exactly recovers the

purely classical Hamiltonian.

Path integrals provide an equivalent description of quantum mechanics for static equilibrium properties, where quantum averages are mapped onto averages in an entirely classical, albeit extended, phase space.<sup>13</sup> For example,

$$\frac{1}{Z} \text{tr} \left[ e^{-\beta \hat{H}} A(\hat{\mathbf{q}}) \right] = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \bar{A}(\mathbf{q}) + \mathcal{O}(\beta_n^2), \quad (1.21)$$

where the ring polymer partition function is

$$Z_n = \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \quad (1.22)$$

and  $\bar{A}(\mathbf{q})$  is the path integral estimator for  $A$ ,

$$\bar{A}(\mathbf{q}) = \frac{1}{n} \sum_{j=1}^n A(\mathbf{q}^{(j)}). \quad (1.23)$$

It is straightforward to evaluate these averages by using classical dynamics to sample the canonical ring polymer ensemble. This so-called path integral molecular dynamics (PIMD) approach has been used successfully for many years in a wide variety of contexts. A particularly convenient approach is Langevin dynamics of the normal modes,<sup>14</sup>

$$\frac{d}{dt} \tilde{q}_i^{(k)} = \frac{\tilde{p}_i^{(k)}}{m_i}, \quad (1.24a)$$

$$\frac{d}{dt} \tilde{p}_i^{(k)} = - \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial q_i^{(j)}} C_{jk} - m_i (\omega^{(k)})^2 \tilde{q}_i^{(k)} - \gamma^{(k)} \tilde{p}_i^{(k)} + \sqrt{\frac{2m_i \gamma^{(k)}}{\beta_n}} \xi_i^{(k)}(t), \quad (1.24b)$$

where  $\xi_i^{(k)}(t)$  is an uncorrelated Gaussian noise for each of the degrees of freedom. As the normal modes are harmonic, a friction coefficient of  $\gamma^{(k)} = 2\omega^{(k)}$  provides optimal sampling of the ring polymer distribution.<sup>14</sup>

The relationship between quantum mechanics and imaginary time path inte-

grals is exact for static equilibrium properties. However, the extension to dynamic properties is not so straightforward, and we must rely on approximate methods.

### 1.3.1 Ring Polymer Molecular Dynamics

PIMD uses the classical dynamics of a ring polymer purely as a means to sample phase space. Ring polymer molecular dynamics (RPMD)<sup>15</sup> takes the dynamics literally, giving

$$\tilde{C}(t) \approx \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \dot{\boldsymbol{\mu}}(\mathbf{p}, \mathbf{q}) \cdot \dot{\boldsymbol{\mu}}(\mathbf{p}(t), \mathbf{q}(t)). \quad (1.25)$$

The evolution of  $\mathbf{p}(t)$  and  $\mathbf{q}(t)$  is the Hamiltonian dynamics of  $H_n$ , which is the same thing as Eq. (1.24) with all  $\gamma^{(k)} = 0$ .

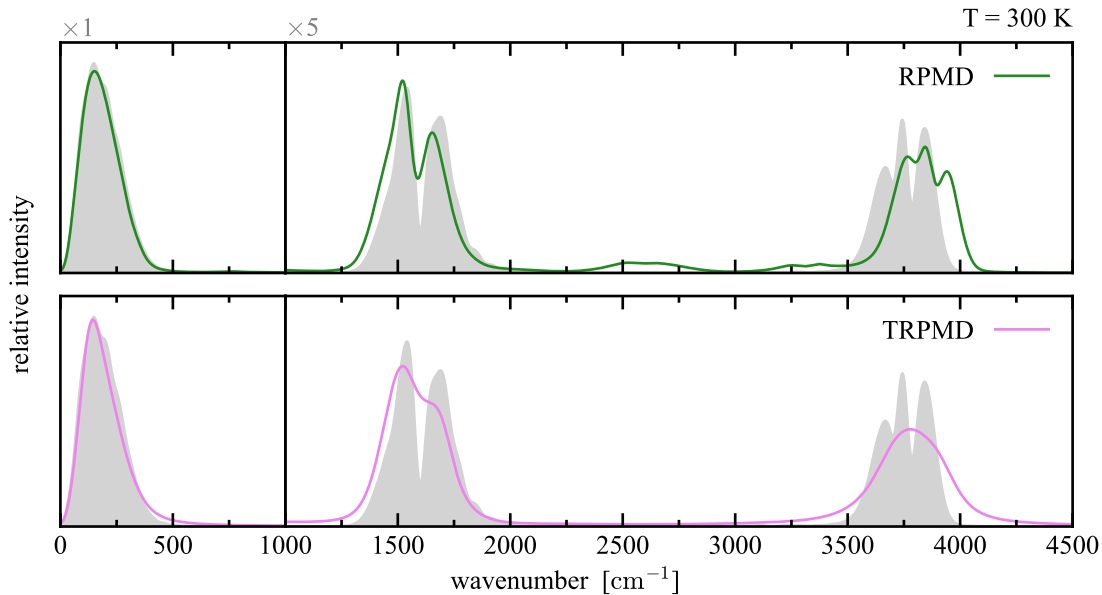
RPMD is entirely *ad hoc*, but Eq. (1.25) is able to reflect the Kubo-transformed correlation function in a number of ways:<sup>15</sup>

- RPMD is consistent with the quantum Boltzmann distribution.
- It reduces to classical dynamics in the  $\beta \rightarrow 0$  or  $n = 1$  limit.
- It is exact in the  $t \rightarrow 0$  limit.
- If  $V(\mathbf{q})$  is harmonic and  $\boldsymbol{\mu}(\mathbf{q})$  is a linear function of  $\mathbf{q}$ , it is exact at all times.

RPMD has been successful in a variety of ‘zero-frequency’<sup>2</sup> applications, including calculating diffusion coefficients<sup>16</sup> and reaction rates.<sup>17–20</sup> However RPMD fails to correctly describe spectroscopy, as vibrations of the  $k \neq 0$  normal modes of the ring polymer can contribute to the dipole moment correlation function. These vibrations are at unphysical frequencies,  $\omega^{(k)}$ , and thus contaminate the spectrum.<sup>2,21,22</sup> For context, at 300 K in the limit that  $n \rightarrow \infty$ , the first normal mode frequency occurs at  $\omega^{(1)}/(2\pi c) \approx 1310 \text{ cm}^{-1}$ , comfortably within the typical frequency range of interest for IR spectroscopy.

Thermostatted RPMD (TRPMD)<sup>22</sup> alleviates these issues by attaching a Langevin thermostat to the internal modes of the ring polymer, damping out the spurious vibrations. In practice, this just means using Eq. (1.24) to describe the dynamics, with  $\gamma^{(k)} = \omega^{(k)}$  as the standard choice for the friction coefficients. This comes at the cost of a fairly significant line broadening effect, and TRPMD fails to correctly predict spectral line shapes.<sup>22,23</sup> Nevertheless, TRPMD maintains all of the desirable features of RPMD<sup>22</sup> and gets band frequencies mostly right, while still being one of the most efficient path integral methods.<sup>9,23,24</sup>

Example spectra produced by both RPMD and TRPMD are shown in fig. 1.3.



**Figure 1.3:** The quantum (shaded), RPMD and TRPMD spectrum of gas-phase  $\text{H}_2\text{O}$  at 300 K. Panels are scaled in the ratio 1:5. See section 1.6 for more details.

### 1.3.2 Centroid Molecular Dynamics

Generally speaking, path integral approaches map quantum problems onto the classical, but extended, phase space of a ring polymer. Centroid molecular dynamics (CMD) takes this even further, to a single classical particle: the ring polymer cen-

troid,

$$\bar{\mathbf{p}} = \frac{1}{n} \sum_{j=1}^n \mathbf{p}^{(j)}, \quad \bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}^{(j)}.$$

The distribution of the centroid is governed by an effective Hamiltonian<sup>25</sup>

$$\bar{H}_n(\mathbf{P}, \mathbf{Q}; \beta) = \sum_{i=1}^N \frac{P_i^2}{2m_i} + \bar{V}_n(\mathbf{Q}; \beta), \quad (1.26)$$

where the effective centroid potential  $\bar{V}_n(\mathbf{Q}; \beta)$  is defined according to

$$e^{-\beta \bar{V}_n(\mathbf{Q}; \beta)} = \left[ \prod_{i=1}^N \Lambda_i \right] \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\mathbf{q}} - \mathbf{Q}), \quad (1.27)$$

and  $\Lambda_i$  is the de Broglie thermal wavelength for the  $i^{\text{th}}$  degree of freedom,

$$\Lambda_i = \sqrt{\frac{2\pi\beta\hbar^2}{m_i}}. \quad (1.28)$$

The CMD approximation to Eq. (1.3) is given by<sup>25</sup>

$$\tilde{C}(t) \approx \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \mathbf{Q})} \dot{\boldsymbol{\mu}}(\mathbf{P}, \mathbf{Q}) \cdot \dot{\boldsymbol{\mu}}(\mathbf{P}(t), \mathbf{Q}(t)), \quad (1.29)$$

where the evolution of  $\mathbf{P}(t)$  and  $\mathbf{Q}(t)$  is the classical dynamics under  $\bar{H}_n$ , and  $Z_n$  is

$$Z_n = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \mathbf{Q})}. \quad (1.30)$$

The CMD approximation of  $\tilde{C}(t)$  shares many of the same desirable properties as RPMD<sup>25</sup>

- CMD is consistent with the quantum Boltzmann distribution.
- It reduces to classical dynamics ( $\bar{V}_n \rightarrow V$ ) in the  $\beta \rightarrow 0$  or  $n = 1$  limit.
- Provided  $\boldsymbol{\mu}(\mathbf{q})$  is a linear function of  $\mathbf{q}$ , it is exact in the  $t \rightarrow 0$  limit.
  - If  $V(\mathbf{q})$  is also harmonic, then it is exact at all times.

When applied to vibrational spectroscopy, CMD performs well at high temperatures but fails drastically at low temperatures, where the spectrum suffers from severe artificial red-shifts and line broadening effects. This is demonstrated in fig. 1.4, and a more detailed discussion of this ‘curvature problem’ is given in chapter 2.

### 1.3.3 Quasi-Centroid Molecular Dynamics

Suppose that we define a set of alternative coordinates  $\boldsymbol{\xi}(\mathbf{q})$ , where the Jacobian associated with this transform is

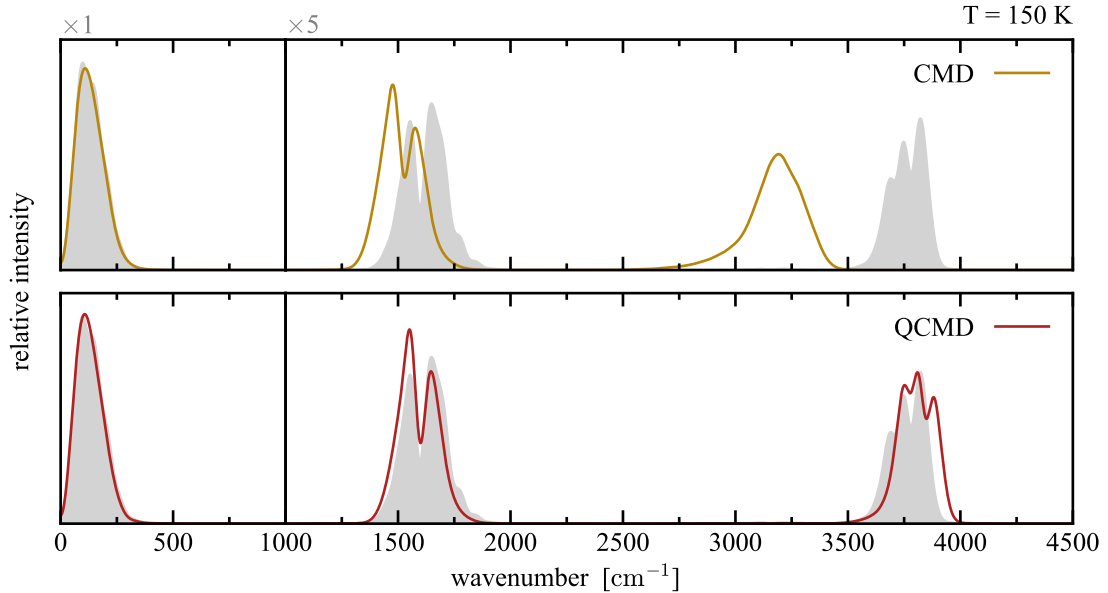
$$J(\boldsymbol{\xi}) = \left\| \frac{\partial(q_1, \dots, q_N)}{\partial(\xi_1, \dots, \xi_N)} \right\|. \quad (1.31)$$

As is shown more fully in chapter 2, it is then possible to create a more general framework that is basically identical to the centroid formalism, except that the effective classical potential is

$$e^{-\beta \bar{V}_n(\mathbf{Q}; \beta)} = \left[ \prod_{i=1}^N \Lambda_i \right] \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \frac{\delta(\bar{\boldsymbol{\xi}} - \boldsymbol{\xi}(\mathbf{Q}))}{J(\boldsymbol{\xi}(\mathbf{Q}))}, \quad (1.32)$$

where

$$\bar{\boldsymbol{\xi}} = \frac{1}{n} \sum_{j=1}^n \boldsymbol{\xi}^{(j)}. \quad (1.33)$$



**Figure 1.4:** The quantum (shaded), CMD and QCMD spectrum of gas-phase  $\text{H}_2\text{O}$  at 150 K. Panels are scaled in the ratio 1:5. See section 1.6 for more details.

Because the constraint on the ring polymer acts via the new coordinates  $\xi$ , the classical position  $\mathbf{Q}$  is no longer equal to the centroid. Instead, it is referred to as the *quasi*-centroid.

The true importance of this quasi-centroid molecular dynamics (QCMD) is that, with a sensible choice of curvilinear coordinates, the ‘curvature problem’ of CMD can be avoided entirely.<sup>8,26,27</sup> This is demonstrated in fig. 1.4, and discussed further in chapter 2.

## 1.4 OTHER METHODS

The methodologies discussed so far feature heavily throughout this thesis, but a wide variety of different approaches to simulating vibrational spectra have also been put forward. These will not be focused on in the current work, but for completeness, a few of these methods are briefly presented here.

### 1.4.1 Matsubara Dynamics

In the limit that  $n \rightarrow \infty$ , a path integral becomes a continuous loop in imaginary time,

$$\begin{aligned} \{\mathbf{p}^{(j)}\} &\rightarrow \mathbf{p}(\tau), & \mathbf{p}(\tau) &= \mathbf{p}(\tau + \beta\hbar), & 0 \leq \tau < \beta\hbar. \\ \{\mathbf{q}^{(j)}\} &\rightarrow \mathbf{q}(\tau), & \mathbf{q}(\tau) &= \mathbf{q}(\tau + \beta\hbar), \end{aligned} \quad (1.34)$$

Expanding  $\mathbf{q}(\tau)$  in a Fourier series gives

$$\mathbf{q}(\tau) = \bar{\mathbf{q}} + \sqrt{2} \sum_{k=1}^{\infty} \tilde{\mathbf{Q}}^{(k)} \sin(\tilde{\omega}^{(k)} \tau) + \sqrt{2} \sum_{k=1}^{\infty} \tilde{\mathbf{Q}}^{(-k)} \cos(\tilde{\omega}^{(k)} \tau), \quad (1.35)$$

where  $\bar{\mathbf{q}} = \tilde{\mathbf{Q}}^{(0)}$  is the centroid and  $\tilde{\omega}^{(k)}$  are the Matsubara frequencies,

$$\tilde{\omega}^{(k)} = \frac{2\pi k}{\beta\hbar}. \quad (1.36)$$

Path integral estimators become integrals around the continuous path

$$\bar{A}(\underline{\mathbf{q}}) \rightarrow \tilde{A}(\tilde{\mathbf{Q}}) \equiv \frac{1}{\beta\hbar} \int_0^{\beta\hbar} d\tau A(\mathbf{q}(\tau)). \quad (1.37)$$

By truncating the Fourier series at terms where  $|k| \leq M \ll n$ , the path becomes a smooth and differentiable function of  $\tau$ . Matsubara dynamics<sup>28</sup> is the (classical) dynamics within this space, under the Hamiltonian

$$\tilde{H}_M(\tilde{\mathbf{P}}, \tilde{\mathbf{Q}}) \equiv \frac{1}{\beta\hbar} \int_0^{\beta\hbar} d\tau H(\mathbf{p}(\tau), \mathbf{q}(\tau)) \quad (1.38)$$

Within this approximation, the partition function becomes

$$\tilde{Z}_M = \left( \frac{M!^2}{2\pi\hbar^{2M+1}} \right)^N \int d^M \tilde{\mathbf{P}} \int d^M \tilde{\mathbf{Q}} e^{-\beta(\tilde{H}_M - i\tilde{\Theta}_M)}, \quad (1.39)$$

and Eq. (1.3) is approximated by

$$\tilde{C}(t) \approx \frac{1}{\tilde{Z}_M} \left( \frac{M!^2}{2\pi\hbar^{2M+1}} \right)^N \times \int d^M \tilde{\mathbf{P}} \int d^M \tilde{\mathbf{Q}} e^{-\beta(\tilde{H}_M - i\tilde{\Theta}_M)} \dot{\boldsymbol{\mu}}(\tilde{\mathbf{P}}, \tilde{\mathbf{Q}}) \cdot \dot{\boldsymbol{\mu}}(\tilde{\mathbf{P}}(t), \tilde{\mathbf{Q}}(t)). \quad (1.40)$$

In any practical application, Matsubara dynamics is impossible to calculate at  $t > 0$  because of the ‘Matsubara phase’,<sup>28</sup>

$$\tilde{\Theta}_M(\tilde{\mathbf{P}}, \tilde{\mathbf{Q}}) = \sum_{k=1}^M \tilde{\omega}^{(k)} \left( \tilde{\mathbf{P}}^{(k)} \cdot \tilde{\mathbf{Q}}^{(-k)} + \tilde{\mathbf{P}}^{(-k)} \cdot \tilde{\mathbf{Q}}^{(k)} \right). \quad (1.41)$$

Instead, its primary use is pedagogical, as every other approximate method discussed so far can be derived as some approximation to Matsubara dynamics.<sup>12,27</sup>

### 1.4.2 Linearised Semiclassical Initial-Value Representation

In addition to path integrals, another map from quantum systems to classical phase space comes from the Wigner transform,

$$\mathcal{W}\{\hat{A}\}(\mathbf{p}, \mathbf{q}) = \int d\mathbf{x} \langle \mathbf{q} - \mathbf{x}/2 | \hat{A} | \mathbf{q} + \mathbf{x}/2 \rangle e^{i\mathbf{p} \cdot \mathbf{x}/\hbar}. \quad (1.42)$$

The partition function can now be expressed as

$$Z = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} \mathcal{W}\{e^{-\beta\hat{H}}\}(\mathbf{p}, \mathbf{q}). \quad (1.43)$$

Taking the Wigner transform of an initial quantum distribution and evolving the resulting phase space classically under  $H(\mathbf{p}, \mathbf{q})$  gives the linearised semiclassical initial-

value representation (LSC-IVR),<sup>28,29</sup>

$$\tilde{C}(t) \approx \frac{1}{\beta Z} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} \mathcal{W} \left\{ \int_0^\beta d\lambda e^{-(\beta-\lambda)\hat{H}} \hat{\boldsymbol{\mu}} e^{-\lambda\hat{H}} \right\} (\mathbf{p}, \mathbf{q}) \cdot \mathcal{W} \{ \hat{\boldsymbol{\mu}} \} (\mathbf{p}(t), \mathbf{q}(t)). \quad (1.44)$$

The LSC-IVR approximation can also be derived using path integrals. This reveals a somewhat surprising similarity to CMD, and LSC-IVR is similarly exact in the harmonic, high temperature and  $t \rightarrow 0$  limits.<sup>29</sup>

The main drawback of LSC-IVR is its failure to conserve the quantum Boltzmann distribution, resulting in ‘zero-point energy leakage’ between inter- and intramolecular modes.<sup>30</sup> In practical terms, this can result in an artificial heating of several hundred Kelvin, and spectral line broadening similar to that of TRPMD.<sup>9,30</sup>

For example spectra, as well as an excellent supplementary to the entire contents of this chapter, reference [9] applies LSC-IVR to exactly the same problem as figs. 1.6 and 1.7.

## 1.5 OVERTONE AND COMBINATION INTENSITY

Classical and path integral methods generally give good estimates for the intensities of fundamental bands, but can be drastically wrong for overtone and combination bands.<sup>9,31,32</sup> Recently, it has been shown that simple first-order perturbation theory gives good estimates for the quantum and classical intensities of these non-fundamental bands.<sup>31</sup> This naturally leads to a simple correction factor that can be applied as a final touch to the path integral spectrum.<sup>32</sup>

We start with a harmonic Hamiltonian with a weak cubic perturbation,

$$\hat{H} = \sum_{i=1}^N \left( \frac{\hat{p}_i^2}{2m_i} + \frac{1}{2} m_i \Omega_i^2 \hat{q}_i^2 \right) + \epsilon \frac{1}{6} \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^N \eta_{ijk} \hat{q}_i \hat{q}_j \hat{q}_k, \quad (1.45)$$

where the coefficients  $\eta_{ijk}$  are invariant under permutation of indices. We use a

linear dipole moment,

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0 + \sum_{i=1}^N \boldsymbol{\mu}_i \hat{q}_i. \quad (1.46)$$

The eigenstates of  $\hat{H}$  are

$$|\mathbf{n}\rangle = |n_1, n_2 \dots n_N\rangle = |\mathbf{n}^{(0)}\rangle + \epsilon |\mathbf{n}^{(1)}\rangle + \epsilon^2 |\mathbf{n}^{(2)}\rangle + \dots \quad (1.47)$$

where  $\mathbf{n}$  are the quantum numbers for the  $N$  vibrational states. It is straightforward to show that

$$\langle \mathbf{n} | \hat{H} | \mathbf{n} \rangle = \sum_{i=1}^N \hbar \Omega_i \left( n_i + \frac{1}{2} \right) + \mathcal{O}(\epsilon^2), \quad (1.48)$$

and thus

$$\omega_{\mathbf{n}\mathbf{n}'} = \sum_{i=1}^N \Omega_i (n_i - n'_i) + \mathcal{O}(\epsilon^2). \quad (1.49)$$

In other words, to first-order, the perturbation does not affect the position of the spectral bands.

The complete derivation of the spectrum associated with this system is given in appendix A. The end results can be summarised by considering the integrated absorption intensity

$$I(\Omega) = \lim_{\delta\omega \rightarrow 0} \int_{\Omega-\delta\omega}^{\Omega+\delta\omega} d\omega n(\omega) \alpha(\omega). \quad (1.50)$$

To first-order in perturbation theory, the intensities of the fundamental bands are unchanged,

$$I(\Omega_j) = \frac{\pi}{3c\epsilon_0 v} \frac{|\boldsymbol{\mu}_j|^2}{2m_j} + \mathcal{O}(\epsilon^2). \quad (1.51)$$

As the intensity is independent of  $\hbar$ , this also corresponds to the classical ( $\hbar \rightarrow 0$ ) limit. This corroborates the observation that classical and path integral methods tend to get the intensity of the fundamental bands right.

The more interesting result of the perturbation is that it gives rise to transitions that differ by two quanta (first overtones, binary combination bands and binary

difference bands)

$$I(\Omega_j \pm \Omega_k) = \epsilon^2 \frac{\pi}{3c\epsilon_0 v} \left| \frac{1}{4} \sum_i^N \frac{\boldsymbol{\mu}_i \eta_{ijk}/m_i}{(\Omega_j \pm \Omega_k)^2 - \Omega_i^2} \right|^2 \frac{1}{(2 - \delta_{jk})m_j m_k} \\ \times \hbar \frac{\Omega_j \pm \Omega_k}{\Omega_j \Omega_k} \left( \coth\left(\frac{\beta \hbar \Omega_k}{2}\right) \pm \coth\left(\frac{\beta \hbar \Omega_j}{2}\right) \right) + \mathcal{O}(\epsilon^3), \quad (1.52)$$

with intensities that do change in the classical limit

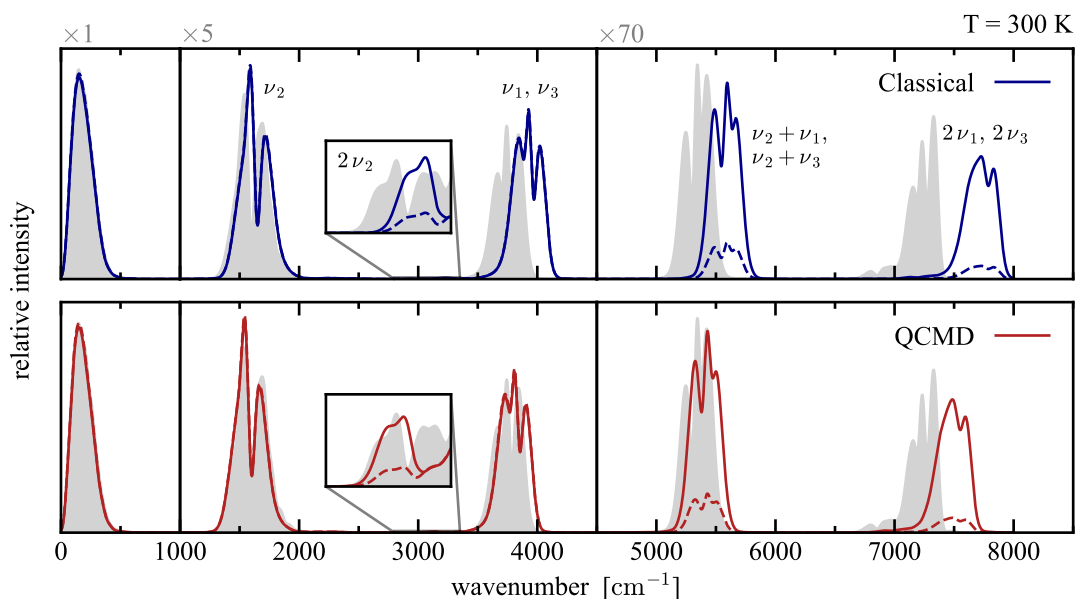
$$\lim_{\hbar \rightarrow 0} I(\Omega_j \pm \Omega_k) = \frac{\epsilon^2 \pi}{3c\epsilon_0 v} \left| \frac{1}{4} \sum_i^N \frac{\boldsymbol{\mu}_i \eta_{ijk}/m_i}{(\Omega_j \pm \Omega_k)^2 - \Omega_i^2} \right|^2 \frac{1}{(2 - \delta_{jk})m_j m_k} \\ \times \frac{2}{\beta} \left( \frac{\Omega_j \pm \Omega_k}{\Omega_j \Omega_k} \right)^2 + \mathcal{O}(\epsilon^3). \quad (1.53)$$

Taking the ratio of the two gives a simple correction for the classical intensities<sup>32</sup>

$$\frac{I(\Omega_j \pm \Omega_k)}{\lim_{\hbar \rightarrow 0} I(\Omega_j \pm \Omega_k)} = \frac{\beta \hbar}{2} \frac{\Omega_j \Omega_k}{\Omega_j \pm \Omega_k} \left( \coth\left(\frac{\beta \hbar \Omega_k}{2}\right) \pm \coth\left(\frac{\beta \hbar \Omega_j}{2}\right) \right). \quad (1.54)$$

This correction factor depends only on temperature and the fundamental frequencies, and so can be applied as a final step to an already calculated spectrum. This is demonstrated in fig. 1.5.

Due to rotational structure and line broadening, bands are never perfectly sharp, meaning there is some ambiguity on how this correction factor should work in practice. In this thesis, the fundamental frequencies  $\Omega_j$  are taken to be at the most intense point of the corresponding band. Corrections are applied by defining a uniform window across an entire overtone/combination band, with the boundaries determined manually by inspection. To smooth the transition to the rest of the spectrum, this square window is convolved with the same line shape  $F(\omega)$  that was used with the rest of the spectrum. This *ad hoc* approach has proven sufficient for this work, but it is entirely feasible that a more general prescription could be developed.



**Figure 1.5:** The quantum (shaded), classical and QCMD spectrum of gas-phase  $\text{H}_2\text{O}$  at 300 K, before (dashed) and after (solid) the intensity correction is applied. Panels are scaled in the ratio 1:5:70. See section 1.6 for more details.

The simplicity of this approach comes with some limitations. Firstly, overtone/combination bands can overlap, leading to some uncertainty as to what correction should be applied over which region of the spectrum. The method is also limited to transitions that differ by at most two quanta, although an appropriate solution might be found with a more general perturbation, or a higher order perturbation theory. Finally, IR inactive modes can produce IR active combination bands. However the lack of fundamental band makes it impossible to deduce the input frequencies without resorting to complementary methods (Raman spectroscopy). All of these issues arise later on, in chapter 4.

## 1.6 SUMMARY OF METHODS

All of the spectra in this chapter were calculated using a Hann window function,

$$f(t) = \begin{cases} \cos^2(\frac{\pi t}{2\tau}), & -\tau \leq t \leq \tau, \\ 0, & \text{otherwise,} \end{cases} \quad (1.55a)$$

$$F(\omega) = \frac{\sin(\omega\tau)}{\omega(1 - (\omega\tau/\pi)^2)}, \quad (1.55b)$$

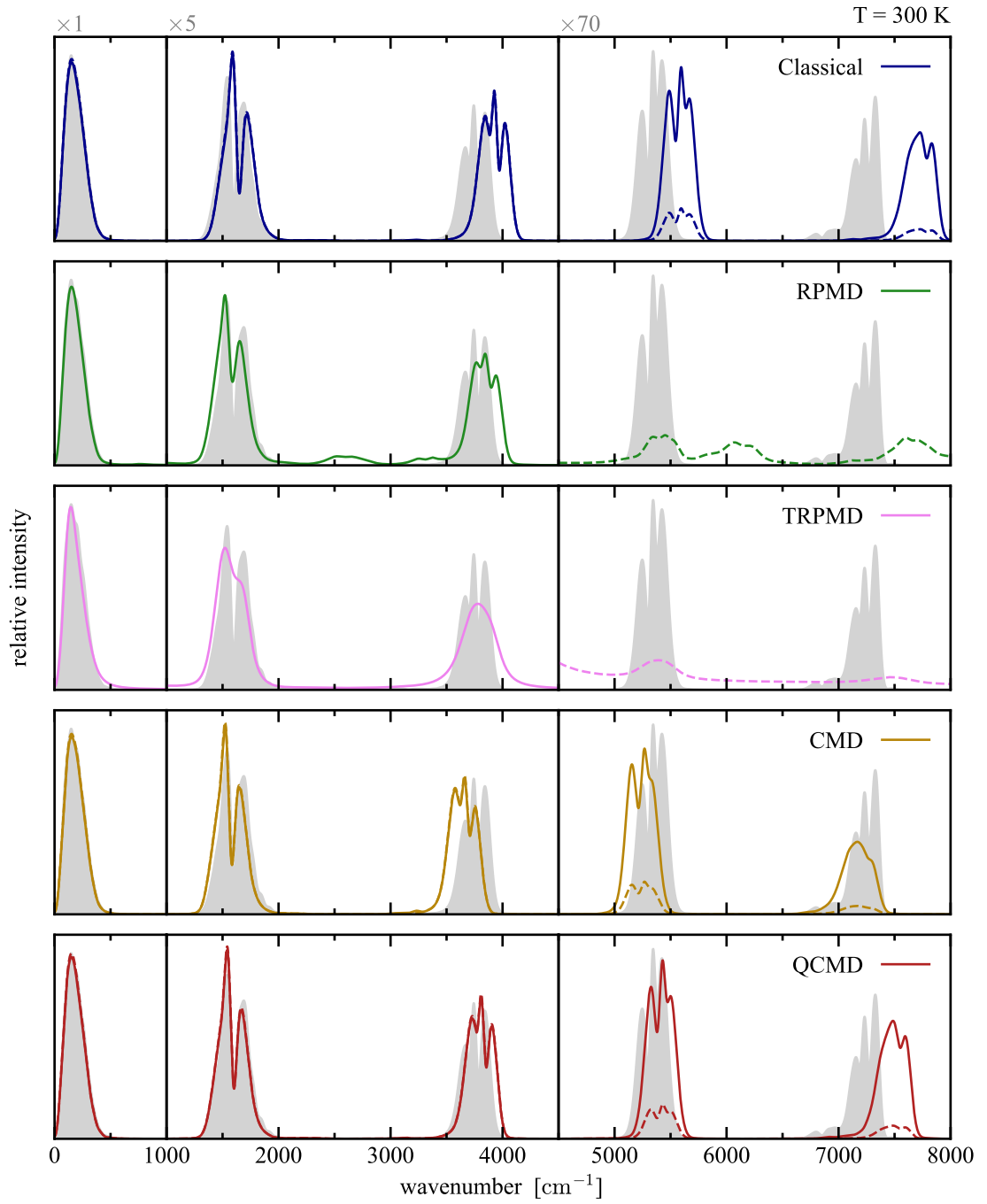
with  $\tau = 0.75$  ps.

The spectrum for  $\text{NH}_3$  in fig. 1.1 was obtained using the spectral linelist included in reference [33]. The same system is re-examined in greater detail in section 4.4.

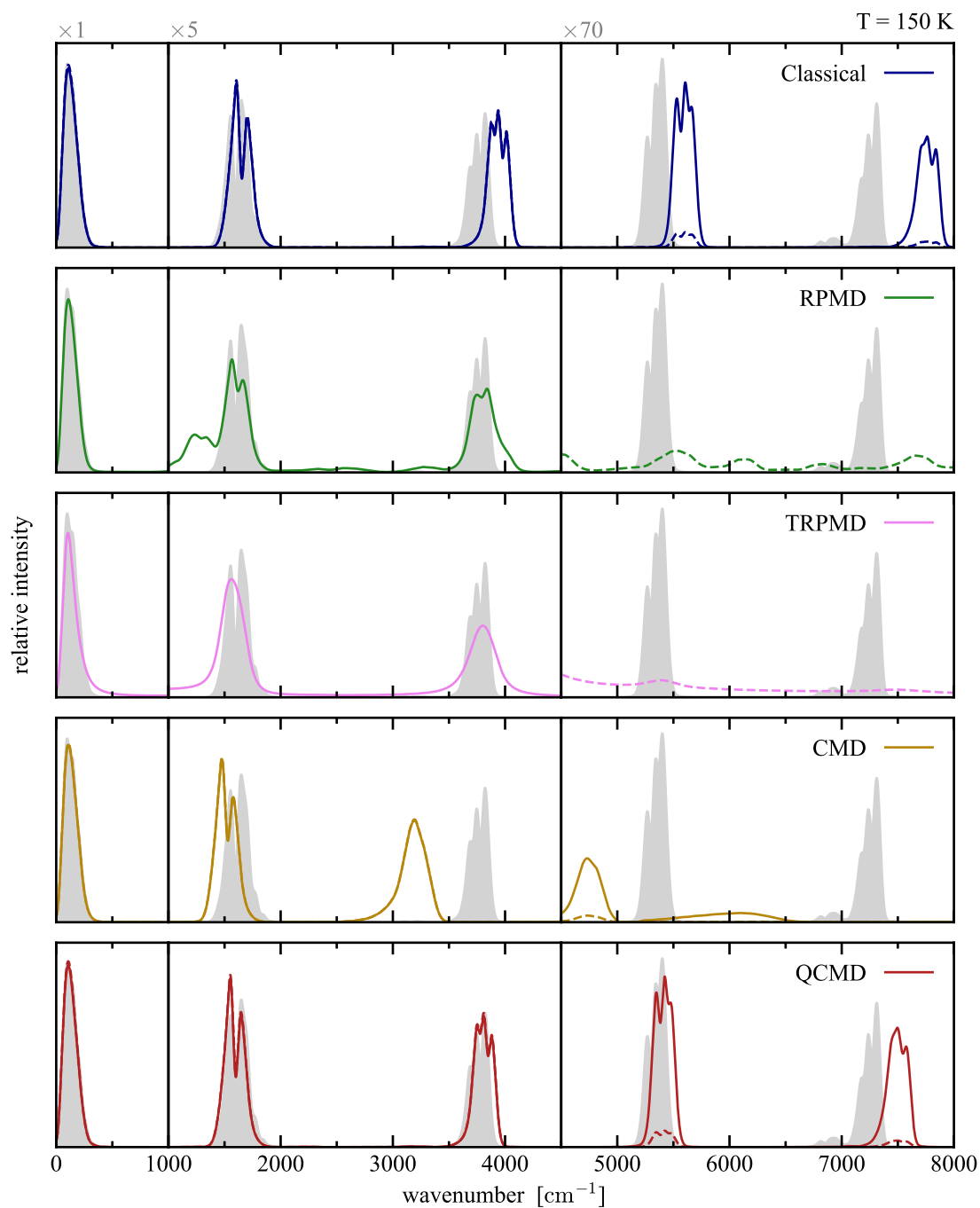
All calculations for  $\text{H}_2\text{O}$  use the potential energy surface<sup>34</sup> and dipole moment surface<sup>35</sup> published by Partridge and Schwenke. This is the same system studied by Benson, Trenins, Willat and Althorpe<sup>9,26</sup>, who kindly provided the quantum spectra. This system is also revisited in greater detail in section 4.2.

All classical and path integral dynamics were carried out with the PILE integration scheme<sup>14</sup> in its original OBABO ordering, with a time step of  $\Delta t = 1$  atomic time unit. All calculations were found to converge with  $n = 32$  at 300 K, and  $n = 64$  at 150 K. All CMD and QCMD calculations are taken directly from section 4.2. Both TRPMD calculations used the recommended friction coefficient of  $\gamma^{(k)} = \omega^{(k)}$ .<sup>22</sup> Due to the spurious resonances of RPMD, and the thermal line broadening of TRPMD, the combination/overtone region of the (T)RPMD spectra were poorly resolved. As a result, the intensity correction factor could not be accurately applied, and was omitted.

A comparison of all key methods applied to gas-phase  $\text{H}_2\text{O}$  is shown in figs. 1.6 and 1.7



**Figure 1.6:** The IR spectrum of gas-phase  $\text{H}_2\text{O}$  at 300 K, calculated exactly (shaded), and with various approximate methods. Where applicable, spectra are shown before (dashed) and after (solid) intensity correction. Panels are scaled in the ratio 1:5:70. See section 1.6 for more details.



**Figure 1.7:** The IR spectrum of gas-phase  $\text{H}_2\text{O}$  at 150 K, calculated exactly (shaded), and with various approximate methods. Where applicable, spectra are shown before (dashed) and after (solid) intensity correction. Panels are scaled in the ratio 1:5:70. See section 1.6 for more details.

## CHAPTER 2

# QUASI-CENTROID MOLECULAR DYNAMICS

With the development of path integrals, Feynman established a crucial link between quantum mechanics and the extended classical phase space of a ring polymer. However his ideas did not stop there, realising that this relationship could be taken all the way to a single classical particle: the path integral centroid.<sup>36</sup> Since those early days, the centroid formulation has been derived again and again,<sup>25,37,38</sup> and is now a robust formalism that, in theory, is completely equivalent to quantum mechanics.<sup>25</sup>

The problem, as with path integrals, is that computing the exact dynamics of the centroid is practically impossible. The response was centroid molecular dynamics (CMD),<sup>39,40</sup> an approximate method which has managed to successfully capture quantum effects in a variety of applications.<sup>41–43</sup> Unfortunately, CMD fails, sometimes quite drastically, to predict vibrational spectra.<sup>9,21,23,26,44,45</sup>

The reason for this is the ‘curvature problem’. This arises because the centroid is a Cartesian average of a ring polymer, but the potential energy of molecular vibrations is best described by curvilinear coordinates (bond lengths and angles).<sup>21</sup> Tweaking CMD to use curvilinear averages instead results in quasi-centroid molecular dynamics (QCMD),<sup>26</sup> which is one of the most promising candidates for solving condensed phase spectroscopy.<sup>8,26</sup> In the short time since it was first suggested, QCMD has been presented as an *ad hoc* modification to CMD. However, this chapter shows that a full quasi-centroid formalism does actually exist, which is more general and just as robust as the centroid theory.

## 2.1 THE CENTROID FORMALISM

There are various ways to derive the centroid formalism of quantum mechanics.<sup>37</sup>

The simplest way, at least for static properties, is to insert the identities

$$1 = \int d\mathbf{P} \, \delta(\bar{\mathbf{p}} - \mathbf{P}) = \int d\mathbf{Q} \, \delta(\bar{\mathbf{q}} - \mathbf{Q}) \quad (2.1)$$

into the path integral expressions from section 1.3. It is also possible to avoid path integrals entirely, through a relationship to the linear response of a quantum system,<sup>38</sup> or by substituting the following expression for the Boltzmann operator<sup>25</sup>

$$e^{-\beta\hat{H}} = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} \, \hat{\Phi}(\mathbf{P}, \mathbf{Q}), \quad (2.2a)$$

$$\hat{\Phi}(\mathbf{P}, \mathbf{Q}) = \frac{(2\pi\hbar)^N}{(2\pi)^{2N}} \int d\mathbf{a} \int d\mathbf{b} \, e^{-\beta\hat{H} + i\mathbf{a}\cdot(\hat{\mathbf{p}} - \mathbf{P}) + i\mathbf{b}\cdot(\hat{\mathbf{q}} - \mathbf{Q})}. \quad (2.2b)$$

Either way, the key resulting quantity is the classical phase space distribution

$$\text{tr} [\hat{\Phi}(\mathbf{P}, \mathbf{Q})] = \lim_{n \rightarrow \infty} e^{-\beta\bar{H}_n(\mathbf{P}, \mathbf{Q})}, \quad (2.3)$$

where

$$e^{-\beta\bar{H}_n(\mathbf{P}, \mathbf{Q})} = \frac{(2\pi\hbar)^N}{(2\pi\hbar)^{nN}} \int d^n\mathbf{p} \int d^n\mathbf{q} \, e^{-\beta H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(\bar{\mathbf{p}} - \mathbf{P}) \delta(\bar{\mathbf{q}} - \mathbf{Q}). \quad (2.4)$$

Integrating over  $\mathbf{P}$  and  $\mathbf{Q}$  shows that the partition function of this distribution is the same as that of a full ring polymer,

$$Z_n = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} \, e^{-\beta\bar{H}_n(\mathbf{P}, \mathbf{Q})}. \quad (2.5)$$

The important conclusion here is that the isomorphism established by path integrals can be taken even further,<sup>25</sup> to a single classical particle with momentum  $\mathbf{P}$  and position  $\mathbf{Q}$ . From Eq. (2.4), it is clear that this particle corresponds to  $\bar{\mathbf{p}}$  and  $\bar{\mathbf{q}}$ : the

centroid of a ring polymer.

By transforming to the normal mode representation (see section 1.3), the momentum integrals in Eq. (2.4) can be solved

$$\int d^n \mathbf{p} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(\bar{\mathbf{p}} - \mathbf{P}) = \left[ \prod_{i=1}^N \frac{\Lambda_i}{2\pi\hbar} e^{-\beta P_i^2/2m_i} \right] \int d^n \mathbf{p} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})}, \quad (2.6)$$

where  $\Lambda_i$  is the de Broglie thermal wavelength for the  $i^{\text{th}}$  degree of freedom

$$\Lambda_i = \sqrt{\frac{2\pi\beta\hbar^2}{m_i}}. \quad (2.7)$$

The centroid Hamiltonian therefore takes the following form

$$\bar{H}_n(\mathbf{P}, \mathbf{Q}) = \sum_{i=1}^N \frac{P_i^2}{2m_i} + \bar{V}_n(\mathbf{Q}), \quad (2.8)$$

where the effective centroid potential is defined according to

$$e^{-\beta \bar{V}_n(\mathbf{Q})} = \left[ \prod_{i=1}^N \Lambda_i \right] \frac{1}{(2\pi\hbar)^{nN}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(\bar{\mathbf{q}} - \mathbf{Q}). \quad (2.9)$$

This definition also provides a helpful insight into how  $\bar{V}_n(\mathbf{Q})$  relates to a ring polymer, as Eq. (2.9) is proportional to the probability of finding a ring polymer with its centroid at  $\mathbf{Q}$ :

$$\bar{\rho}_n(\mathbf{Q}) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{nN}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(\bar{\mathbf{q}} - \mathbf{Q}). \quad (2.10)$$

In words, the effective centroid potential is defined such that the probability of finding the particle at  $\mathbf{Q}$  is identical to the probability of finding the centroid of a ring polymer at  $\mathbf{Q}$ .

As with path integrals, within the centroid isomorphism, static quantum prop-

erties map exactly onto classical averages,<sup>25</sup>

$$\frac{1}{Z} \text{tr} \left[ e^{-\beta \hat{H}} A(\hat{\mathbf{q}}) \right] = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \mathbf{Q})} \bar{A}_n(\mathbf{Q}) + \mathcal{O}(\beta_n^2). \quad (2.11)$$

Instead of the path integral estimator, we now have  $\bar{A}_n(\mathbf{Q})$ , the centroid function corresponding to  $A(\hat{\mathbf{q}})$ , defined by

$$\text{tr} \left[ \hat{\Phi}(\mathbf{P}, \mathbf{Q}) A(\hat{\mathbf{q}}) \right] = \lim_{n \rightarrow \infty} e^{-\beta \bar{H}_n(\mathbf{P}, \mathbf{Q})} \bar{A}_n(\mathbf{Q}). \quad (2.12)$$

Provided  $A$  is a function of position only, a standard path integral discretisation gives  $\bar{A}_n(\mathbf{Q})$  as

$$\bar{A}_n(\mathbf{Q}) = \left\langle \frac{1}{n} \sum_{j=1}^n A(\mathbf{q}^{(j)}) \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}}, \quad (2.13)$$

where the brackets indicate a path integral average, such that the centroid is constrained at  $\mathbf{Q}$

$$\left\langle \dots \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}} = \frac{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\mathbf{q}} - \mathbf{Q}) (\dots)}{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\mathbf{q}} - \mathbf{Q})}. \quad (2.14)$$

In principle, a centroid function  $\bar{A}_n(\mathbf{Q})$  could be calculated from a path integral calculation of the corresponding estimator. However, the typical approach is to simply use the classical observable, evaluated at the centroid,  $\bar{A}_n(\mathbf{Q}) \approx A(\mathbf{Q})$ .<sup>46</sup> The justification behind this comes from expanding the function  $A(\mathbf{q})$  about  $\mathbf{Q}$ ,

$$\bar{A}_n(\mathbf{Q}) = A(\mathbf{Q}) + \frac{1}{2} \sum_{ii'}^N \frac{\partial^2 A}{\partial Q_i \partial Q_{i'}} \left\langle \frac{1}{n} \sum_{j=1}^n \left( q_i^{(j)} - Q_i \right) \left( q_{i'}^{(j)} - Q_{i'} \right) \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}} + \dots \quad (2.15)$$

This assumption is nothing more than the lowest-order term of this expansion, with higher-order terms accounting for fluctuations of the ring polymer beads about the centroid. From the above, this approximation is clearly valid in two cases:

- $A(\mathbf{Q})$  is a linear function of  $\mathbf{Q}$ .
- Fluctuations about the centroid are small, such that higher-order terms are negligible.

### 2.1.1 Centroid Molecular Dynamics

The centroid framework established so far is, fundamentally, completely equivalent to quantum mechanics. As a result, exact centroid dynamics is basically the same problem as quantum dynamics and is just as difficult,<sup>25</sup> so any practical application requires an approximation to be made.

To this end, the classical force generated by  $\bar{V}_n(\mathbf{Q})$  is

$$-\frac{\partial \bar{V}_n}{\partial Q_i} = -\left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial q_i^{(j)}} \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}}. \quad (2.16)$$

This shows that  $\bar{V}_n(\mathbf{Q})$  is in fact a potential of mean force. More specifically, mean force in this context refers to the average force on the beads of a ring polymer, such that the centroid is at  $\mathbf{Q}$ . As  $\mathbf{P}$  also corresponds to the mean momentum of the ring polymer, Hamilton's equations with  $\bar{H}_n(\mathbf{P}, \mathbf{Q})$  results in Ehrenfest-like dynamics.<sup>25,39</sup>

Centroid molecular dynamics (CMD) is exactly this: classical dynamics under  $\bar{H}_n$ .<sup>25,39</sup> The CMD approximation to Eq. (1.3) is given by

$$\tilde{C}(t) \approx \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \mathbf{Q})} \bar{\boldsymbol{\mu}}(\mathbf{P}, \mathbf{Q}) \cdot \dot{\boldsymbol{\mu}}(\mathbf{P}(t), \mathbf{Q}(t)). \quad (2.17)$$

Note also that in order to give an approximation to the Kubo-transformed correlation function, we have made the assumption  $\bar{\boldsymbol{\mu}}_n \approx \boldsymbol{\mu}^{25,47}$  — the error associated with this was discussed previously. The CMD approximation is thus correct at  $t \rightarrow 0$  only in the case of a linear dipole moment. Additionally, if  $V$  is harmonic, or in the high temperature ( $\beta \rightarrow 0$ ) limit, then  $\bar{V}_n \rightarrow V$  and CMD is exact at all times.<sup>25</sup>

Nevertheless, when it comes to spectroscopy, the available evidence suggests that when CMD does a good job, it does so even with non-linear dipole moments.<sup>2,10,48</sup> However the story may not be that simple. As discussed in the following section, the only time CMD performs well is when the ring polymer distribution is compact, which is the same condition under which  $\bar{\boldsymbol{\mu}}_n \approx \boldsymbol{\mu}$  is valid. If the distribution is not compact, the error in CMD comes overwhelmingly from the dynamics. The dipole assumption may well break down as well, but further investigation would be required to demonstrate this.

## 2.2 THE CURVATURE PROBLEM

At least for static properties, the centroid formalism correctly captures nuclear quantum effects, and it is worth considering why this is. As an example, this chapter uses the Morse potential

$$V_{\text{OH}}(r) = D_e \left(1 - e^{-a(r-r_e)}\right)^2. \quad (2.18)$$

The parameters  $D_e = 0.18748$ ,  $a = 1.1605$ ,  $r_e = 1.8324$  and mass of 1741.1 (all in atomic units) give a typical O–H stretching potential.<sup>26</sup>

Suppose a ring polymer is on a potential energy surface. When the centroid is at the potential energy minimum, the beads will be spread up the sides of the surface, and the mean potential experienced by the replicas is higher than the potential at the centroid itself,

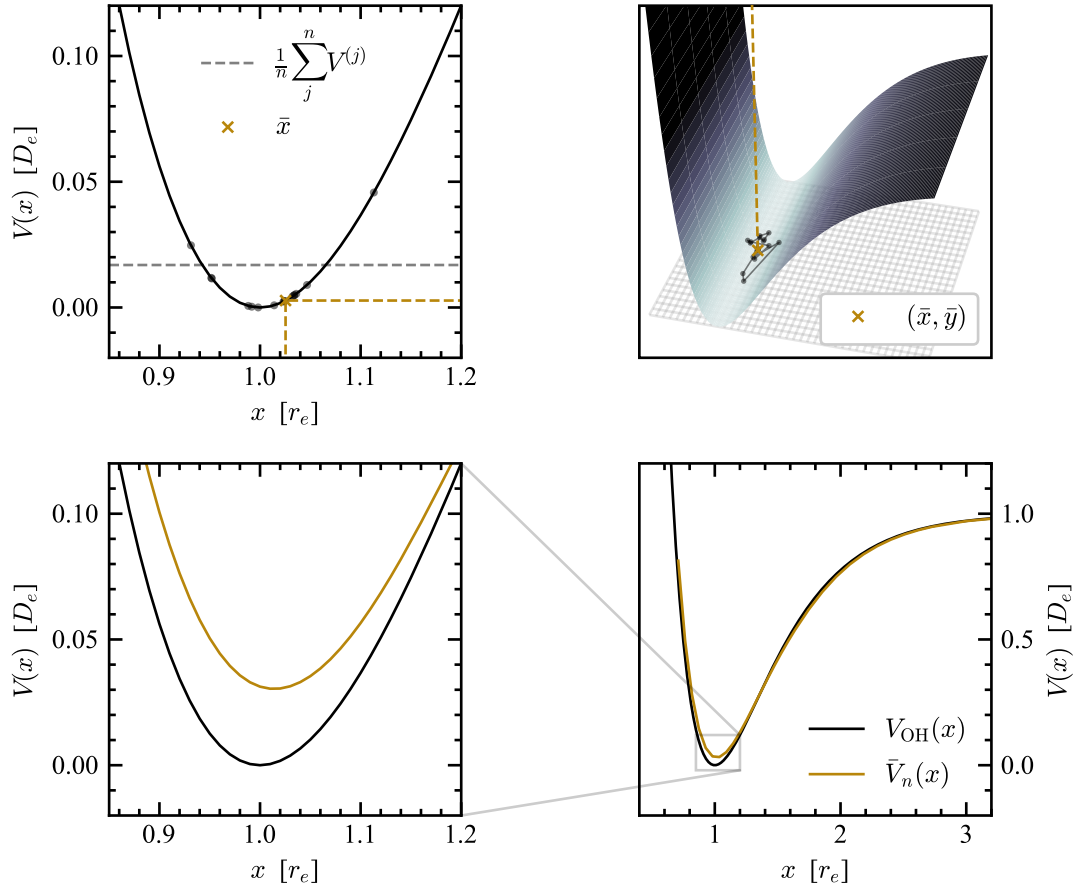
$$\frac{1}{n} \sum_{j=1}^n V^{(j)} > V(\bar{\mathbf{q}}).$$

As a result, the centroid of a ring polymer is less likely to be at the potential minimum than a classical particle,  $\bar{\rho}_n(\mathbf{Q}) < \bar{\rho}_1(\mathbf{Q})$ . In terms of the centroid potential, relative to the original surface,  $\bar{V}_n(\mathbf{Q})$  destabilises potential energy minima, which accounts for zero-point energy effects. The exact opposite is observed for potential

maxima, which are stabilised relative to the original surface, capturing tunnelling effects.

This effect is demonstrated in fig. 2.1, which shows an exemplary ring polymer configuration and the centroid potential, for a simple two-dimensional model with  $V(x, y) = V_{\text{OH}}(x)$ . Note that the  $y$  coordinate is in fact redundant in this model. However, as will be shown later, the breakdown of CMD cannot occur with just one dimension. This model is a multi-dimensional case that *does* work as intended, and is analogous to the later model (fig. 2.2) that fails.

In anharmonic cases such as this, starting from equilibrium the potential increases more gradually as  $x$  increases rather than decreases. As a result, the ring



**Figure 2.1:** The model system  $V(x, y) = V_{\text{OH}}(x)$  at 300 K.

**Top** An exemplary ring polymer, its centroid, and the mean potential of its beads.

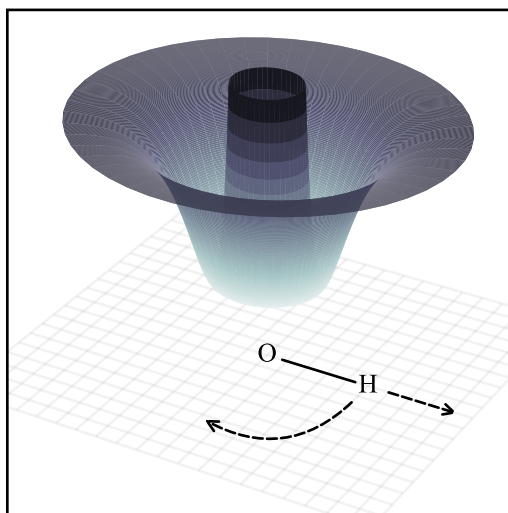
**Bottom** The effective centroid potential.

polymer beads generally sit at larger values of  $x$ , pulling the equilibrium centroid position in the same direction, out of the classical minimum. Lower temperatures cause the ring polymer to expand, exaggerating this effect.

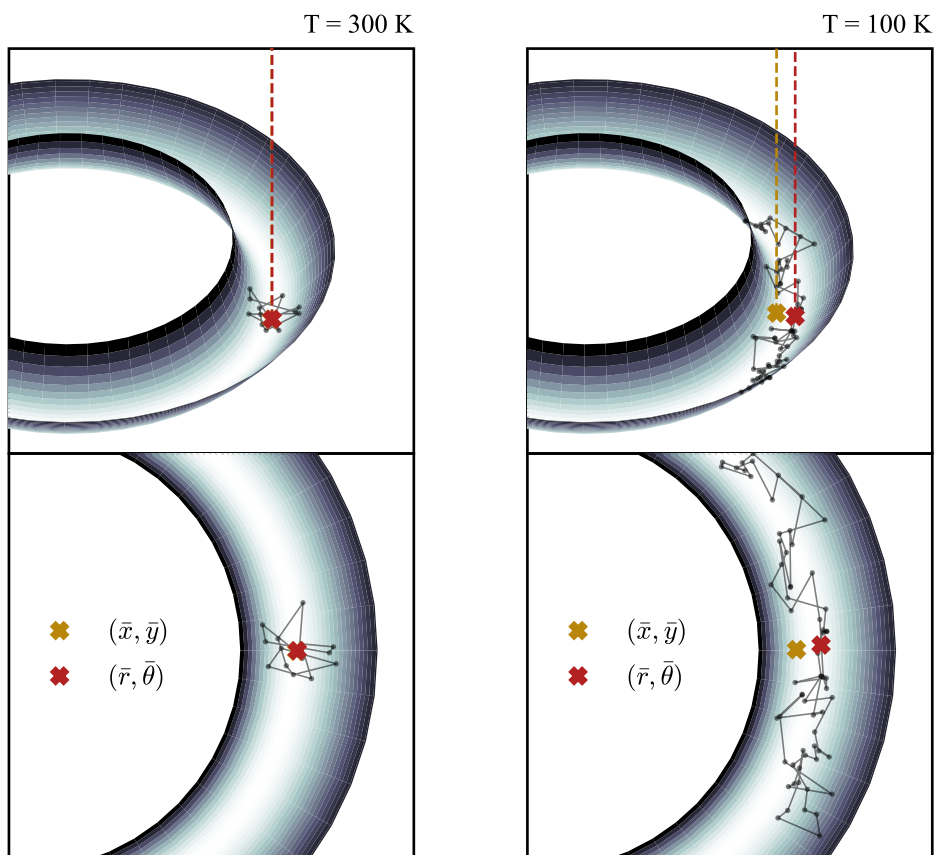
In some applications, CMD is able to correctly leverage this behaviour to reproduce quantum effects, and has been used successfully to calculate thermal conductivity,<sup>41</sup> reaction rates<sup>42</sup> and diffusion coefficients.<sup>43</sup> In a particularly interesting study it is argued, and then validated numerically, that CMD correctly predicts the first excitation energy of a 1-dimensional anharmonic oscillator as  $\beta \rightarrow \infty$ .<sup>38,49</sup>

At high temperatures, CMD also works well for vibrational spectroscopy.<sup>2,10,23,48,50</sup> However major issues arise at low temperatures where CMD predicts significant band broadening and red-shifting.<sup>9,21,23,26,44,45</sup>

Understanding the cause of this requires a model system that better reflects an actual molecule. Suppose the O (or any heavy) atom is fixed at the origin. This is bonded to a H atom that is free to rotate in the  $xy$  plane. As shown in fig. 2.2, the resulting potential energy surface is the O–H stretching potential swept around the  $z$ -axis:  $V(x, y) = V_{\text{OH}}(r)$ .



**Figure 2.2:** The radial O–H model potential energy surface,  $V(x, y) = V_{\text{OH}}(r)$ . The bond can stretch in the  $r$  direction, and rotate in the  $xy$  plane.



**Figure 2.3:** Exemplary ring polymer configurations for the model  $V(x, y) = V_{\text{OH}}(r)$ . Also shown are the centroid and quasi-centroid.

Now consider how a ring polymer behaves on the radial potential energy surface, as shown in fig. 2.3. At low temperatures, the polymer expands around the potential minimum, forming a crescent. This shape has the centroid closer to the origin than the beads themselves, inside the equilibrium bond length.<sup>45</sup> The result is an artificial stabilisation of short centroid bond lengths, the exact opposite to what one might expect of this anharmonic potential (see fig. 2.1). For  $\bar{V}_n$ , this manifests as a softening of the short-range repulsive interaction, giving lower frequency (red-shifted) oscillations.

If the H atom is allowed to rotate in 3D space, this phenomenon is the same, although it is harder to visualise. Now the ring polymer will lie on the surface of a sphere with radius  $r_e$ , and the centroid is within this sphere, giving the same result.<sup>21</sup> The best description of an OH group in a real molecule would be something between

the 2D and 3D models: rotation of the OH bond, but a more limited bending motion out of the plane.

Fundamentally, this behaviour occurs because molecular potential energy surfaces are curved, earning it the name: ‘the curvature problem’.<sup>21</sup>

If the ring polymer distribution is more compact, or the potential energy surface is less curved, it follows that the curvature problem is less severe. This is why CMD works as intended at higher temperatures, as the more compact ring polymer does not ‘feel’ the curvature of the surface, and instead recovers the paradigm of fig. 2.1.

More quantitatively, for a radial model such as this, Trenins argues for the existence of a cutoff radius  $r_c$  satisfying<sup>27</sup>

$$r_c = -\frac{V'(r_c)}{m} \left( \frac{\beta \hbar}{2\pi} \right)^2, \quad (2.19)$$

such that the curvature problem arises if radii shorter than  $r_c$  are thermally accessible. As  $r_c$  must be positive,  $V'(r_c)$  must be negative; therefore  $r_c$  must be shorter than the classical equilibrium  $r_e$ . Increasing  $\beta$  or decreasing  $m$  (which makes the system less classical) brings  $r_c$  closer to  $r_e$ . This makes the area inside  $r_c$  more accessible, and the curvature problem is more pronounced.

A more subtle point about this equation is that the left-hand side increases as  $r_c$  moves forward from 0, but the right-hand side increases as  $r_c$  moves backwards from  $r_e$ . The point at which the two sides meet thus depends on  $r_e$ , the equilibrium bond length. The shorter  $r_e$  is in the first place, then the closer  $r_c$  will be to it when the equation is satisfied. Again, this makes the area inside  $r_c$  more accessible, and the curvature problem more severe. Qualitatively, this is the same argument as before. A bond that is short at equilibrium has a tightly curved potential energy surface, that suffers severely from the curvature problem. Bonds that are longer at equilibrium have less tightly curved potentials that more closely resemble fig. 2.1.

### 2.3 POLAR CONSTRAINTS

This section continues to focus on the curvature problem for  $V(x, y) = V_{\text{OH}}(r)$ .

Ignoring the spring forces for now, each bead of the ring polymer is really just a classical particle subject to the potential  $V_{\text{OH}}(r)$ . We therefore expect the radius of the beads to be close to equilibrium. We can also expect this to be the case including the effect of the springs. At low temperatures, the weak spring forces leave the external potential as the dominant effect, whereas at high temperatures, the strong springs cause the beads to bunch up and behave more like a single classical particle subject to  $V_{\text{OH}}(r)$ .

Looking at fig. 2.3, this is indeed the case. Despite a wide spread in the  $x$  and  $y$  coordinates, each bead has a fairly reasonable radius. In fact, the mean radius,

$$\bar{r} = \frac{1}{n} \sum_{j=1}^n r^{(j)}, \quad (2.20a)$$

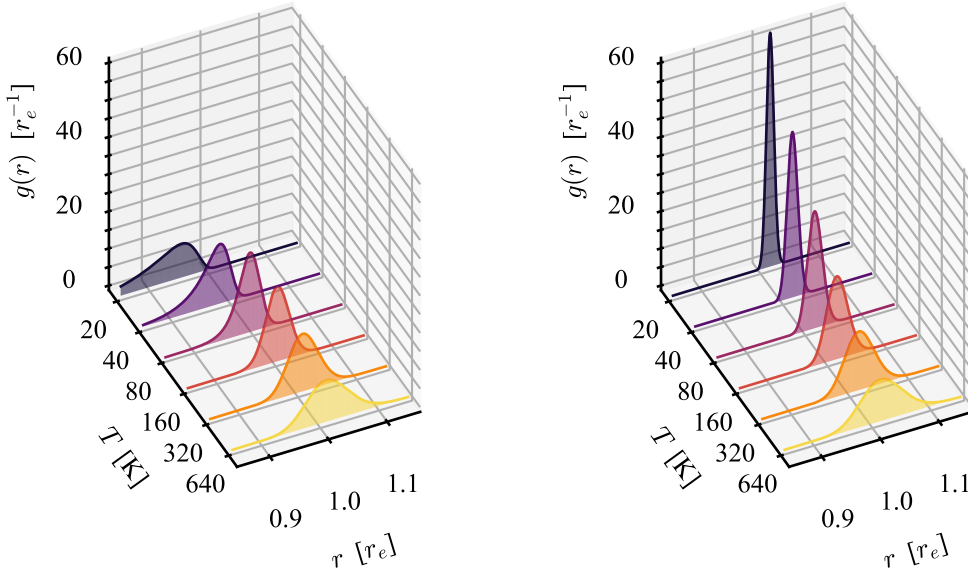
is also at a sensible value. Taking this further, fig. 2.4 shows the distribution function of the centroid radius  $\sqrt{\bar{x}^2 + \bar{y}^2}$ , and the mean radius  $\bar{r}$ , at various temperatures. At low temperatures, the centroid radius tends towards the origin (i.e.: the curvature problem). However,  $\bar{r}$  remains near the classical equilibrium at every temperature.

The curvature problem is the consequence of applying a Cartesian constraint to a system that is best described by curvilinear (polar) coordinates. So with the observations made above, an intuitive suggestion would be to simply replace the Cartesian averages with appropriate polar ones:  $\bar{r}$  and

$$\bar{\theta} = \frac{1}{n} \sum_{j=1}^n \theta^{(j)}. \quad (2.20b)$$

This gives<sup>26,27</sup>

$$e^{-\beta \bar{V}_n(R, \Theta)} = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \frac{\delta(\bar{r} - R) \delta(\bar{\theta} - \Theta)}{R}. \quad (2.21)$$



**Figure 2.4:** Path integral radial distribution functions for  $V(x, y) = V_{\text{OH}}(r)$  at various temperatures.

**Left** Centroid radius:  $\sqrt{\bar{x}^2 + \bar{y}^2}$

**Right** Mean radius:  $\bar{r}$ .

By design, the radial distribution of this new potential is the same as that of  $\bar{r}$  — already an improvement over the centroid potential. But the real test is, of course, the dynamics.

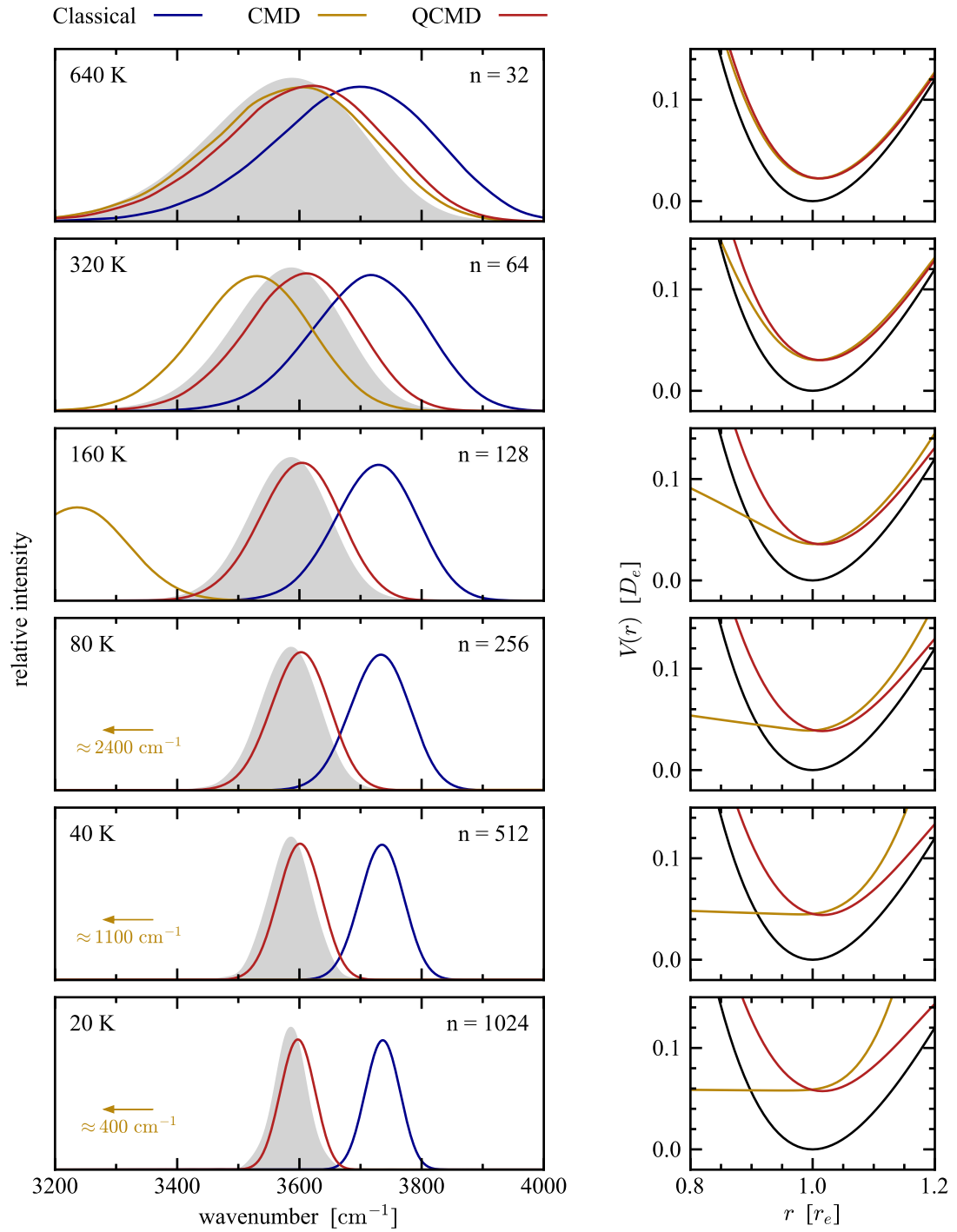
Classical dynamics, original CMD and the modification in Eq. (2.21) were used to calculate the vibrational spectrum of  $V(x, y) = V_{\text{OH}}(r)$ , using a linear dipole moment,

$$\boldsymbol{\mu}(\mathbf{q}) = \boldsymbol{\mu}_0 + \alpha \mathbf{q}. \quad (2.22)$$

The dipole-derivative autocorrelation functions were damped with a Hann window function,

$$f(t) = \begin{cases} \cos^2(\frac{\pi t}{2\tau}), & -\tau \leq t \leq \tau, \\ 0, & \text{otherwise,} \end{cases}$$

with  $\tau = 0.75$  ps. The resulting spectra and potential energy surfaces are shown in fig. 2.5, at a range of temperatures.



**Figure 2.5:** CMD, QCMD and classical dynamics applied to  $V(x, y) = V_{\text{OH}}(r)$ .  
**left** IR spectrum stretching band, compared to quantum (shaded) results.  
**right** Centroid and quasi-centroid potentials, compared to  $V_{\text{OH}}(r)$  (black).

As expected, classical dynamics gives a spectrum with a moderate blue-shift of  $\approx 150 \text{ cm}^{-1}$  at every temperature. CMD performs well at 640 K, but the curvature problem is noticeable at every temperature below this; by 80 K the spectrum becomes unrecognisable and the band frequency is off by thousands of  $\text{cm}^{-1}$ . This is easily understood from the effective centroid potential. The curvature problem causes the repulsive potential ( $r < r_e$ ) to soften, and by 20 K, it is practically non-existent. On the other hand, the modified potential Eq. (2.21) is sensible, and produces accurate spectra, right down to 20 K.

Because it is a linear function of position, the CMD dipole moment  $\bar{\mu}_n = \mu(\bar{x}, \bar{y})$  is exact, meaning the severe error in CMD comes entirely from the dynamics. On the other hand, the modified method clearly does very well, but does approximate the dipole moment (see section 2.4). This may account for some discrepancy in the line shape, but overall, this is strong evidence that it is much more important to get the dynamics right, rather than the dipole moment.

An interesting observation from fig. 2.5 is that the modified potential, Eq. (2.21), is basically identical at all temperatures. An explanation for this is that the vibrational excitation energy is around  $3600 \text{ cm}^{-1}$ , much larger than the thermal energy ( $k_B T \approx 445 \text{ cm}^{-1}$  at 640 K, see fig. 2.6). Even at 640 K, the system is mainly in its vibrational ground state, meaning the effective potential, which is intended to reflect the true quantum statistics, is already at its low temperature limit.<sup>51</sup>

Furthermore, at 640 K, the two effective potentials are very similar, yet clearly different from the classical limit. This suggests that 640 K is cold enough to capture the quantum ground state, but not cold enough for the curvature problem to take effect. This observation has recently inspired the temperature elevation ( $T_e$ ) modification to CMD,<sup>51</sup> where the effective centroid potential is calculated at a high temperature, but it is used for low temperature dynamics to get spectra without the curvature problem. Because it was published so recently, this thesis does not investigate this new method, but it will probably receive more focus in the future.

## 2.4 THE QUASI-CENTROID FORMALISM

The CMD modification in the previous section is a clear success for the model system  $V(x, y) = V_{\text{OH}}(r)$ . This section expands on this idea, to give a complete and general theory.

The key to the success of section 2.3 was using polar coordinates, which better described the underlying potential energy surface. We can extend this to a general system with a transformation to a set of curvilinear coordinates  $\boldsymbol{\xi}(\mathbf{q}) \equiv (\xi_1, \dots, \xi_N)$  with the Jacobian

$$J(\boldsymbol{\xi}) = \left\| \frac{\partial(q_1, \dots, q_N)}{\partial(\xi_1, \dots, \xi_N)} \right\|. \quad (2.23)$$

Using these new coordinates, the identity from section 2.1 now reads as

$$1 = \int d\boldsymbol{\Xi} \, \delta(\bar{\boldsymbol{\xi}} - \boldsymbol{\Xi}) = \int d\mathbf{Q} \, \frac{\delta(\bar{\boldsymbol{\xi}} - \boldsymbol{\Xi})}{J(\boldsymbol{\Xi})}, \quad (2.24)$$

where  $\boldsymbol{\Xi} \equiv \boldsymbol{\xi}(\mathbf{Q})$ , and  $\bar{\boldsymbol{\xi}}$  is the path integral estimator for  $\boldsymbol{\xi}$ ,

$$\bar{\boldsymbol{\xi}} = \frac{1}{n} \sum_{j=1}^n \boldsymbol{\xi}^{(j)}. \quad (2.25)$$

Alternatively, modifying the operator-based approach<sup>25</sup> in section 2.1 gives

$$\hat{\Phi}(\mathbf{P}, \boldsymbol{\Xi}) = \frac{(2\pi\hbar)^N}{(2\pi)^{2N}} \frac{1}{J(\boldsymbol{\Xi})} \int d\mathbf{a} \int d\mathbf{b} \, e^{-\beta\hat{H} + i\mathbf{a}\cdot(\hat{\mathbf{p}} - \mathbf{P}) + i\mathbf{b}\cdot(\hat{\boldsymbol{\xi}} - \boldsymbol{\Xi})}. \quad (2.26)$$

Following the exact same steps as section 2.1 then gives a theory that is more-or-less identical to the centroid formalism, except that the effective Hamiltonian and potential energy surface is now

$$\bar{H}_n(\mathbf{P}, \boldsymbol{\Xi}) = \sum_{i=1}^n \frac{P_i^2}{2m_i} + \bar{V}_n(\boldsymbol{\Xi}), \quad (2.27)$$

with

$$e^{-\beta \bar{V}_n(\Xi)} = \left[ \prod_{i=1}^N \Lambda_i \right] \frac{1}{(2\pi\hbar)^{nN}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \frac{\delta(\bar{\boldsymbol{\xi}} - \Xi)}{J(\Xi)}. \quad (2.28)$$

The momentum  $\mathbf{P}$  still refers to the mean Cartesian momentum of the ring polymer, but the position  $\mathbf{Q}$  is now defined such that  $\Xi \equiv \boldsymbol{\xi}(\mathbf{Q})$  is the mean  $\boldsymbol{\xi}$  of each of the replicas. This is expected to be close to, but not exactly the same as, the centroid, and has thus been named the *quasi-centroid*.<sup>26,27</sup>

The centroid and quasi-centroid theories are very similar. In fact, the quasi-centroid formalism is more general, with the centroid theory being the specific case that  $\boldsymbol{\xi}(\mathbf{q}) = \mathbf{q}$ . To distinguish between the two, the convention used is that centroid properties are functions of  $\mathbf{Q}$  and quasi-centroid properties are functions of  $\Xi \equiv \boldsymbol{\xi}(\mathbf{Q})$ . It is worth remembering that this is just a notation choice — a function of  $\Xi$  is also a function of the quasi-centroid  $\mathbf{Q}$ .

As with the centroid, the quasi-centroid formalism exactly recovers  $Z_n$  and static properties,

$$Z_n = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \Xi)}, \quad (2.29)$$

$$\frac{1}{Z} \text{tr} \left[ e^{-\beta \hat{H}} A(\hat{\mathbf{q}}) \right] = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \Xi)} \bar{A}_n(\Xi) + \mathcal{O}(\beta_n^2). \quad (2.30)$$

By the same logic as section 2.1, the quasi-centroid version of an operator  $A(\hat{\mathbf{q}})$  is

$$\bar{A}_n(\Xi) = \left\langle \frac{1}{n} \sum_{j=1}^n A(\mathbf{q}^{(j)}) \right\rangle_{\bar{\boldsymbol{\xi}}=\Xi}, \quad (2.31)$$

where the brackets indicate a path integral average where  $\bar{\boldsymbol{\xi}}$  is constrained at  $\Xi$ ,

$$\left\langle \dots \right\rangle_{\bar{\boldsymbol{\xi}}=\Xi} = \frac{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\boldsymbol{\xi}} - \Xi) (\dots)}{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\boldsymbol{\xi}} - \Xi)}. \quad (2.32)$$

The arguments at the end of section 2.1 are also completely analogous, and

expanding  $A(\mathbf{q})$  in terms of  $\boldsymbol{\xi}$  gives

$$\bar{A}_n(\boldsymbol{\Xi}) = A(\boldsymbol{\Xi}) + \frac{1}{2} \sum_{ii'}^N \frac{\partial^2 A}{\partial \Xi_i \partial \Xi_{i'}} \left\langle \frac{1}{n} \sum_{j=1}^n \left( \xi_i^{(j)} - \Xi_i \right) \left( \xi_{i'}^{(j)} - \Xi_{i'} \right) \right\rangle_{\bar{\boldsymbol{\xi}}=\boldsymbol{\Xi}} + \dots \quad (2.33)$$

Thus the approximation  $\bar{A}_n(\boldsymbol{\Xi}) \approx A(\boldsymbol{\Xi})$ , or equivalently  $\bar{A}_n(\mathbf{Q}) \approx A(\mathbf{Q})$  in terms of the quasi-centroid, is valid when the ring polymer is compact, or if  $A$  is a linear function of  $\boldsymbol{\xi}$ .

This also reveals more about the possible error in fig. 2.5 from the quasi-centroid approximation to the dipole moment, which really should be

$$\bar{\boldsymbol{\mu}}_n(R, \Theta) = \boldsymbol{\mu}_0 + \alpha \left\langle \bar{\mathbf{q}} \right\rangle_{(\bar{r}, \bar{\theta})=(R, \Theta)}. \quad (2.34)$$

Due to the curvature problem, when a ring polymer is constrained at  $\bar{r} \leq r_e$ , the centroid should be at a shorter radius than  $\bar{r}$  on average. By the same logic, when the quasi-centroid is inside the equilibrium radius,  $\bar{\boldsymbol{\mu}}_n(R, \Theta)$  is closer to  $\boldsymbol{\mu}_0$  than  $\boldsymbol{\mu}(R, \Theta)$  is. This may well have an effect on the resulting spectrum, especially at low temperatures.

In the current work, the approximation  $\bar{\boldsymbol{\mu}}_n \approx \boldsymbol{\mu}$  has been used throughout, with no investigation into the validity of this assumption. However, this may be an interesting topic for future work.

### 2.4.1 Quasi-Centroid Molecular Dynamics

Exact dynamics in the quasi-centroid formalism is just as difficult as with the centroid — in fact the latter is just a specific case of the former. We thus require a more general approximate method: quasi-centroid molecular dynamics (QCMD).<sup>26</sup> This is exactly the same as CMD, but using the quasi-centroid potential, Eq. (2.28),

$$\tilde{C}(t) \approx \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \boldsymbol{\Xi})} \dot{\boldsymbol{\mu}}(\mathbf{P}, \mathbf{Q}) \cdot \dot{\boldsymbol{\mu}}(\mathbf{P}(t), \mathbf{Q}(t)). \quad (2.35)$$

QCMD also maintains most of the desirable properties of CMD:<sup>27</sup>

- QCMD dynamics is consistent with the quantum Boltzmann distribution.
- It reduces to classical dynamics ( $\bar{V}_n \rightarrow V$ ) in the  $\beta \rightarrow 0$  or  $n = 1$  limit.
- Provided  $\boldsymbol{\mu}$  is a linear function of  $\boldsymbol{\xi}$ , it is exact in the  $t \rightarrow 0$  limit.

CMD has the additional benefit of correct behaviour in the harmonic limit. An analogous limit for QCMD may well exist, but nothing has been found yet; solving path integrals and quantum mechanics in curvilinear coordinates is usually easier said than done.<sup>52–54</sup> Nevertheless, the method can certainly be justified based on its performance for spectroscopy, where it has been used successfully on a variety of gas-phase molecules<sup>8,24,26</sup>, liquid water and ice.<sup>9,26</sup>

Compared to CMD, the forces from a quasi-centroid potential are substantially more complicated

$$-\frac{\partial \bar{V}_n}{\partial \Xi_i} = - \left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial H_n}{\partial \xi_i^{(j)}} + \frac{1}{\beta} \frac{\partial}{\partial \Xi_i} \ln J(\Xi) - \frac{1}{\beta} \sum_{j=1}^n \frac{\partial}{\partial \xi_i^{(j)}} \ln J(\boldsymbol{\xi}^{(j)}) \right\rangle_{\bar{\boldsymbol{\xi}}=\Xi}, \quad (2.36a)$$

and contain contributions from both the Jacobian term, and the ring polymer springs

$$\frac{\partial H_n}{\partial \xi_i^{(j)}} = \frac{\partial V^{(j)}}{\partial \xi_i^{(j)}} + \omega_n^2 \sum_{k=1}^N m_k \frac{\partial q_k^{(j)}}{\partial \xi_i^{(j)}} \left( 2q_k^{(j)} - q_k^{(j+1)} - q_k^{(j-1)} \right). \quad (2.36b)$$

Chapter 3 has more detail on how these quasi-centroid forces are calculated.

## 2.5 COMPUTATIONAL DETAILS

All path integral averages were calculated using PIMD and the PILE integration scheme<sup>14</sup> in its original OBABO ordering, with a time step of  $\Delta t = 1$  atomic time unit. All classical, CMD and QCMD dynamics was done with the  $n = 1$  limit of the PILE integrator, which is the same thing as velocity Verlet integration. A time step of  $\Delta t = 1$  atomic time unit was also used.

### 2.5.1 Distribution Functions

Figure 2.4 shows radial distribution functions (RDFs) of the centroid and quasi-centroid for the model  $V(x, y) = V_{\text{OH}}(r)$ . Respectively, these are

$$g(R) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\sqrt{\bar{x}^2 + \bar{y}^2} - R). \quad (2.37a)$$

$$g(R) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{r} - R). \quad (2.37b)$$

Both of these RDFs can be calculated precisely using the techniques explained in chapter 5. For this problem, the reduced variance estimators can be constructed from

$$\frac{\partial}{\partial r} \ln J = \frac{1}{r}, \quad \frac{\partial \mathbf{q}}{\partial r} = \frac{\mathbf{q}}{r}.$$

### 2.5.2 Effective Potential Energy Surfaces

The radial potential energy surface  $V(x, y) = V_{\text{OH}}(r)$  is a function of  $r$  only. By symmetry the effective (quasi-)centroid potential must also be a function of  $r$  only, and as just one dimension is important, it is feasible to calculate it without approximation.

For the centroid potential, starting from Eq. (2.9),

$$e^{-\beta \bar{V}_n(X, Y)} = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{x} - X) \delta(\bar{y} - Y),$$

and constraining  $\sqrt{X^2 + Y^2}$  gives

$$\begin{aligned} \int dX \int dY e^{-\beta \bar{V}_n(X,Y)} \delta(\sqrt{X^2 + Y^2} - R) \\ = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int dX \int dY \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n} \\ \times \delta(\bar{x} - X) \delta(\bar{y} - Y) \delta(\sqrt{X^2 + Y^2} - R). \end{aligned} \quad (2.38a)$$

The integrals over  $X$  and  $Y$  can be solved exactly. On the left-hand side, this is done by transforming to polar coordinates and noting that  $\bar{V}_n$  is independent of  $\theta$ . On the right-hand side, the delta functions simply constrain  $X$  and  $Y$  to  $\bar{x}$  and  $\bar{y}$ ,

$$2\pi R e^{-\beta \bar{V}_n(R)} = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\sqrt{\bar{x}^2 + \bar{y}^2} - R). \quad (2.38b)$$

Finally, dividing both sides by  $Z_n$  then gives

$$\frac{2\pi R}{Z_n} e^{-\beta \bar{V}_n(R)} = \Lambda^2 g(R), \quad (2.38c)$$

where  $g(R)$  is the centroid RDF, as defined in the previous section.

Similarly, for the quasi-centroid potential, we start from Eq. (2.28)

$$e^{-\beta \bar{V}_n(R)} = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \frac{\delta(\bar{r} - R) \delta(\bar{\theta} - \Theta)}{R}.$$

Integrating over  $\Theta$  on both sides,

$$2\pi e^{-\beta \bar{V}_n(R)} = \frac{\Lambda^2}{(2\pi\hbar)^{2n}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \frac{\delta(\bar{r} - R)}{R}, \quad (2.39a)$$

and dividing by  $Z_n$  gives

$$\frac{2\pi}{Z_n} e^{-\beta \bar{V}_n(R)} = \frac{\Lambda^2}{R} g(R), \quad (2.39b)$$

where this time  $g(R)$  is the quasi-centroid RDF, as defined in the previous section.

Both effective potentials have nearly identical forms, and can both be obtained

from their corresponding RDFs

$$\bar{V}_n(R) = -\frac{1}{\beta} \ln \left( \frac{g(R)}{R} \right) - \frac{1}{\beta} \ln \left( \frac{Z_n \Lambda^2}{2\pi} \right). \quad (2.40)$$

In practice, the RDFs are evaluated on a grid of  $R$  values, so a continuous function for the potential was obtained by fitting a cubic spline through  $\ln g(R)$ .

### Absolute Values

The RDFs are all that is practically required for (Q)CMD. This is because the final term in Eq. (2.40) is independent of  $R$ , and can be set to an arbitrary constant value with no effect on the dynamics or statistics. However, this term has been included in figs. 2.1 and 2.5, to give a completely faithful representation of the potential as it has been defined.

Calculating the centroid potential absolutely (including the extra term) is possible, but this is a slow approach for the whole surface. However doing this calculation at one particular geometry gives an absolute reference value, and shifting the surface gained by RDFs to pass through this point then gives the absolute surface, at all values of  $R$ .

To this end, we can introduce the modified Hamiltonian  $H_n^{(\lambda)}$ ,

$$H_n^{(\lambda)}(\underline{\mathbf{p}}, \underline{\mathbf{q}}) = \sum_{j=1}^n \sum_{i=1}^N \frac{(p_i^{(j)})^2}{2m_i} + \sum_{j=1}^n \sum_{i=1}^N \frac{1}{2} m_i \omega_n^2 (q_i^{(j)} - q_i^{(j+1)})^2 + \lambda \sum_{j=1}^n V^{(j)}. \quad (2.41)$$

This is just the standard ring polymer Hamiltonian, but the strength of the external potential has been scaled by a parameter  $\lambda$ . The benefit of this is that the difference in centroid potential from two different values of  $\lambda$  is given by a path integral average

$$e^{-\beta[\bar{V}_n^{(\lambda')} - \bar{V}_n^{(\lambda)}]} = \frac{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n^{(\lambda)}} \delta(\bar{\mathbf{q}} - \mathbf{Q}) e^{-\beta_n (H_n^{(\lambda')} - H_n^{(\lambda)})}}{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n^{(\lambda)}} \delta(\bar{\mathbf{q}} - \mathbf{Q})}, \quad (2.42a)$$

$$= \left\langle e^{-(\lambda' - \lambda)\beta_n \sum_j^n V^{(j)}} \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}}^{(\lambda)}. \quad (2.42b)$$

This is an average in the ensemble of  $H_n^{(\lambda)}$ , with the constraint that  $\bar{\mathbf{q}} = \mathbf{Q}$ . Equations of this form are typically referred to as ‘free energy perturbation’.<sup>13</sup>

Another key observation is that with no external potential, there is no centroid potential of mean force,  $\bar{V}_n^{(0)} = 0$ .\* Thus, a free energy perturbation from  $\lambda = 0 \rightarrow 1$  provides a route to the absolute value of the centroid potential,

$$e^{-\beta \bar{V}_n(\mathbf{Q})} = \left\langle e^{-\beta_n \sum_j^n V^{(j)}} \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}}^{(0)}. \quad (2.43)$$

For fig. 2.5, this calculation was done at a single point on the classical minimum,  $(X_0, Y_0) = (r_e, 0)$ . This gets an absolute reference value of the centroid potential at  $R = r_e$ . The extra term in Eq. (2.40) is therefore

$$\begin{aligned} -\frac{1}{\beta} \ln \left( \frac{Z_n \Lambda^2}{2\pi} \right) &= \frac{1}{\beta} \ln \left( \frac{g(r_e)}{r_e} \right) + \bar{V}_n(r_e) \\ &= \frac{1}{\beta} \ln \left( \frac{g(r_e)}{r_e} \right) - \frac{1}{\beta} \ln \left\langle e^{-\beta_n \sum_j^n V(\mathbf{q}_j)} \right\rangle_{\mathbf{q}_n=\mathbf{Q}_0}^{(0)}, \end{aligned} \quad (2.44)$$

where  $g(R)$  is the centroid radial distribution function.

Again, it should be stressed that this step is not necessary for (Q)CMD, because shifting the whole surface up or down has no effect on the dynamics. This is done here purely to give plots of  $\bar{V}_n$  that reflect how it has been defined earlier.

The surface in fig. 2.1 is purely illustrative, and does not need to be precise enough for dynamics. It also extends well beyond the region in which a distribution function will converge. Therefore, this surface was evaluated by free energy

---

\*This can be proven by substituting  $H_n^{(0)}$  into Eq. (2.9).

perturbation at a set of  $X$  values, and smoothed through by a cubic spline.

With the PILE integration scheme, all updates to the positions are done in the normal mode representation, so the constraint on the centroid simply amounts to not updating  $\tilde{x}^{(0)}$  and  $\tilde{y}^{(0)}$ . Free energy perturbation only converges for small changes in the distributions.<sup>13</sup> Following standard practice, this calculation was therefore split into 20 equal increments,

$$\begin{aligned} \left\langle e^{-\beta_n \sum_j^n V^{(j)}} \right\rangle_{\mathbf{q}_n = \mathbf{Q}_0}^{(0)} &= \left\langle e^{-\beta_n \sum_j^n V^{(j)}/20} \right\rangle_{\mathbf{q}_n = \mathbf{Q}_0}^{(0)} \times \left\langle e^{-\beta_n \sum_j^n V^{(j)}/20} \right\rangle_{\mathbf{q}_n = \mathbf{Q}_0}^{(0.05)} \cdots \\ &\quad \cdots \times \left\langle e^{-\beta_n \sum_j^n V^{(j)}/20} \right\rangle_{\mathbf{q}_n = \mathbf{Q}_0}^{(0.9)} \times \left\langle e^{-\beta_n \sum_j^n V^{(j)}/20} \right\rangle_{\mathbf{q}_n = \mathbf{Q}_0}^{(0.95)}. \end{aligned}$$

### 2.5.3 Quantum Spectra

The eigenstates of the radial model  $V(x, y) = V_{\text{OH}}(r)$  are parameterised by a vibrational number  $\nu$ , and a rotational number  $l$ . At the temperatures considered, only  $v = 0$  states are thermally accessible, and only  $v = 0 \rightarrow 1$  transitions are considered. Using Eq. (1.6) and the line shape function  $F(\omega)$ , the quantum spectrum is therefore given by

$$n(\omega)\alpha(\omega) \propto \sum_{l,l'} \omega_{l'l} e^{-\beta E_{0l}} (1 - e^{-\beta \hbar \omega_{l'l}}) |\langle 0, l | \hat{\boldsymbol{\mu}} | 1, l' \rangle|^2 F(\omega - \omega_{l'l}). \quad (2.45)$$

The  $v$  indices have been omitted from the frequencies  $\omega_{l'l}$  for conciseness.

#### Eigenstates

Using the position representation and polar coordinates, the Schrödinger equation for this model is

$$E_{\nu l} \Psi_{\nu l} = -\frac{\hbar^2}{2m} \left( \frac{\partial^2 \Psi_{\nu l}}{\partial r^2} + \frac{1}{r} \frac{\partial \Psi_{\nu l}}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \Psi_{\nu l}}{\partial \theta^2} \right) + V_{\text{OH}}(r) \Psi_{\nu l}. \quad (2.46)$$

The total wavefunction can be decomposed into a radial component  $R_{\nu l}(r)$ , which describes vibrations of the bond, and an angular component  $\Theta_l(\theta)$ , which describes rotation about the origin

$$\Psi_{\nu l}(r, \theta) = \frac{R_{\nu l}(r)\Theta_l(\theta)}{\sqrt{r}}. \quad (2.47)$$

Substituting this into the Schrödinger equation and rearranging then gives

$$E_{\nu l}R_{\nu l}\Theta_l = -\frac{\hbar^2}{2m}\frac{d^2R_{\nu l}}{dr^2}\Theta_l - \frac{\hbar^2}{2mr^2}\frac{d^2\Theta_l}{d\theta^2}R_{\nu l} + \left(V_{\text{OH}}(r) - \frac{\hbar^2}{8mr^2}\right)R_{\nu l}\Theta_l. \quad (2.48)$$

The angular differential equation in the above can be identified as the textbook ‘particle-on-a-ring’ problem, with the eigenstates

$$\Theta_l(\theta) = \frac{e^{il\theta}}{\sqrt{2\pi}}, \quad l \in \mathbb{Z}. \quad (2.49)$$

Substituting this result,

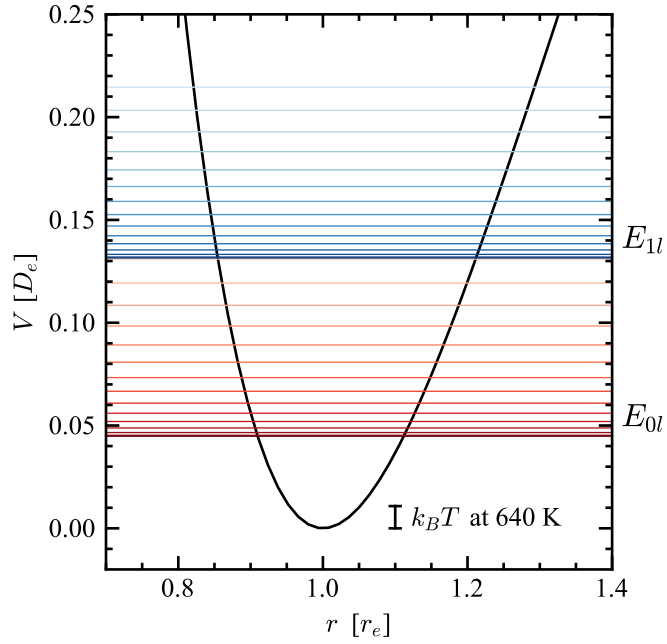
$$E_{\nu l}R_{\nu l}\Theta_l = -\frac{\hbar^2}{2m}\frac{d^2R_{\nu l}}{dr^2}\Theta_l + \frac{\hbar^2l^2}{2mr^2}R_{\nu l}\Theta_l + \left(V_{\text{OH}}(r) - \frac{\hbar^2}{8mr^2}\right)R_{\nu l}\Theta_l.$$

and cancelling out  $\Theta_l$  gives a differential equation for  $R_{\nu l}(r)$

$$E_{\nu l}R_{\nu l}(r) = -\frac{\hbar^2}{2m}\frac{d^2R_{\nu l}}{dr^2} + \left(V_{\text{OH}}(r) + \frac{\hbar^2}{2mr^2}\left(l - \frac{1}{2}\right)\left(l + \frac{1}{2}\right)\right)R_{\nu l}(r). \quad (2.50)$$

This is simply a one-dimensional Schrödinger equation with an effective potential, which can be solved with standard numerical methods. In this case, a sinc function discrete variable representation<sup>55</sup> was used between  $0.3 r_e$  and  $3 r_e$ , with a grid spacing of  $\Delta r = 0.001 r_e$ . This was used to find the  $v = 0$  and  $v = 1$  eigenstates for each value of  $l$  where  $E_{0l} < E_{10}$ . Beyond this, the  $|0, l\rangle$  states are thermally inaccessible, and need not be included.

As a final aside, the three-dimensional version of this model can be solved in an



**Figure 2.6:** Quantum energy levels,  $E_{0l}$  and  $E_{1l}$ , for  $V(x, y) = V_{\text{OH}}(r)$ . Also shown is the thermal energy,  $k_B T$  at 640 K.

entirely analogous way, by decomposing the wavefunction as

$$\Psi_{\nu lm}(r, \theta, \phi) = \frac{R_{\nu lm}(r) Y_{lm}(\theta, \phi)}{r}. \quad (2.51)$$

This time, the functions  $Y_{lm}(\theta, \phi)$  turn out to be the spherical harmonics, and in fact the whole system is completely analogous to the textbook problem of the hydrogen atom.

### Transition Dipole Moment

For a  $v = 0 \rightarrow 1$  transition, the matrix elements of a linear dipole moment  $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0 + \alpha \hat{\mathbf{q}}$  are

$$\langle 0, l | \hat{\boldsymbol{\mu}} | 1, l' \rangle = \alpha \int dx \int dy \Psi_{0l}^* \begin{pmatrix} x \\ y \end{pmatrix} \Psi_{1l'}, \quad (2.52a)$$

which is separable when expressed in polar coordinates

$$\langle 0, l | \hat{\boldsymbol{\mu}} | 1, l' \rangle = \alpha \int_0^\infty dr R_{0l}^*(r) r R_{1l'}(r) \int_0^{2\pi} d\theta \Theta_l^*(\theta) \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \Theta_{l'}(\theta). \quad (2.52b)$$

The angular wavefunctions are known, and this integral can be solved analytically. First, the trigonometric terms are expanded using Euler's formula

$$\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = \frac{e^{i\theta}}{2} \begin{pmatrix} 1 \\ -i \end{pmatrix} + \frac{e^{-i\theta}}{2} \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Then using the relationship  $\Theta_{l\pm 1} = e^{\pm i\theta} \Theta_l$ , and the orthogonality of the wavefunctions gives

$$\int_0^{2\pi} d\theta \Theta_l^*(\theta) \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} \Theta_{l'}(\theta) = \frac{\delta_{l,l'+1}}{2} \begin{pmatrix} 1 \\ -i \end{pmatrix} + \frac{\delta_{l,l'-1}}{2} \begin{pmatrix} 1 \\ i \end{pmatrix}. \quad (2.53)$$

Using this result, the square magnitude of the transition dipole moment is

$$|\langle 0, l | \hat{\boldsymbol{\mu}} | 1, l' \rangle|^2 \propto (\delta_{l,l'+1} + \delta_{l,l'-1}) \left| \int_0^\infty dr R_{0l}^*(r) r R_{1l'}(r) \right|^2. \quad (2.54)$$

To summarise, the only effect of the angular wavefunction is a  $\Delta l = \pm 1$  selection rule. The radial wavefunctions,  $R_{\nu l}$ , have already been found numerically, and the remaining integral can be evaluated using the discrete variable representation. Finally, in fig. 2.5, the intensity is scaled to give a normalised spectrum. Therefore any other coefficients can be ignored, and the right-hand side of the above equation can be used as is.

## CHAPTER 3

# FAST QUASI-CENTROID MOLECULAR DYNAMICS

With the theory of QCMD laid out, we now consider how to implement it, which raises a few questions.

The first and most fundamental question is what coordinates should be used to define the quasi-centroid? This is quite difficult, because the valence of the molecule (bond lengths and angles) has to be used to avoid the curvature problem, but this gives far too many coordinates for a proper transformation.

Then even if an appropriate coordinate transformation can be found, there is still the problem of performance. CMD is technically just classical dynamics, but the computational cost is far higher. The reason is that the potential energy surface is defined by a path integral average. Naively, this constitutes a full-scale path integral calculation at every single time step, and an overall process that is orders of magnitude slower than, for example, RPMD. Fortunately, considerable effort has gone into finding ways around this, and the longevity of CMD is not just owed to the theory, but to these improvements in its implementation. However the tricks that make CMD a viable approach do not translate well to QCMD, and a new approach is needed if it is ever going to be a useful tool.

The answer to both questions comes from coarse-graining, which turns out to share remarkable similarities with the (quasi-)centroid formalism. It is also a thriving field of study, with plenty of tried-and-tested ideas that can be taken and applied directly to (Q)CMD. Ultimately, this leads to the key idea of this thesis: fast quasi-centroid molecular dynamics, or f-QCMD.

### 3.1 CHOOSING COORDINATES

The first question that needs answering is how should the quasi-centroid coordinates be picked to give the best results for vibrational dynamics? Similar questions exist in many other areas of chemistry too,<sup>54,56,57</sup> including vibrational analysis and geometry optimisation, which provides a useful insight into the pros and cons of various ideas when it comes to QCMD.

A handful of the coordinates are obvious. Suppose a molecule has  $N_a$  atoms, and  $N = 3N_a$  degrees of freedom overall. Translation and rotation of the entire molecule will each account for 3 degrees of freedom (assuming that the molecule is non-linear), and neither will affect the potential energy. This leaves  $3N_a - 6$  coordinates to describe vibrations, still to be chosen.

The naive choice would be the ubiquitous ‘normal coordinates’,<sup>56</sup> which are just the eigenvectors of the (mass-weighted) potential energy Hessian

$$H_{ij} = \frac{1}{\sqrt{m_i m_j}} \frac{\partial^2 V}{\partial q_i \partial q_j}, \quad (3.1)$$

at an equilibrium geometry. However these coordinates are not curvilinear, they are rectilinear, i.e.: linear combinations of Cartesian coordinates,

$$\xi_i = \sum_j a_{ij} q_j. \quad (3.2a)$$

Consequently, constraining the ring polymer average of these coordinates,

$$\bar{\xi}_i = \sum_j a_{ij} \bar{q}_j, \quad (3.2b)$$

is really just the same thing as constraining the centroid, and is no different from CMD.

Outside of QCMD, these normal coordinates are already known to be problematic

because they are only valid for very small displacements from equilibrium, and considerable coupling between coordinates can occur at large separations.<sup>56,57</sup> This is basically the same reason they are invalid for QCMD, except this time the large displacements arise from the dispersion of the ring polymer.

Generally speaking, it has long been acknowledged that the best coordinate choice tends to be based on the bonding of the molecule, i.e.: bond lengths, bond angles and torsion angles.<sup>57–59</sup> These give the most natural and physically motivated picture of a molecule, and significantly reduce coupling between coordinates. As discussed in chapter 2, the same logic applies to QCMD.

The main problem with valence-based coordinates is that there are just far too many of them. For example, even without considering angles, the  $N_a(N_a - 1)/2$  interatomic distances in a molecule already outnumber  $3N_a - 6$  if there are more than 4 atoms. For a better understanding of this redundancy, it is useful to consider an infinitesimal change in a valence coordinate,<sup>56</sup>

$$B_{ij} = \frac{\partial r_i}{\partial q_j}, \quad (3.3)$$

where  $r_i \equiv r_i(\mathbf{q})$  is one of the valence coordinates, and the derivatives are evaluated at an equilibrium geometry. Typically, there are many more  $r_i$  coordinates than  $q_j$ , and a valence coordinate is redundant if its row in the  $\mathbf{B}$  matrix is linearly dependent on some other rows.<sup>60</sup>

With the above, simply finding and eliminating some redundant variables is possible,<sup>59</sup> but the question still remains as to what non-redundant subset gives the best description. This is difficult even in some of the simplest imaginable scenarios; for example, consider a molecule of  $\text{H}_2\text{O}$  with 3 vibrational coordinates to pick. The two O–H bond lengths probably influence the potential most, and are sensible choices for the first two coordinates, but what about the third? The H–H distance and H–O–H bond angle will both influence the potential — the water model used

here is a function of them both<sup>34</sup> — so which is the best choice?

This is the same problem that arises using the popular Z-matrix coordinate system. This is basically just a convention for choosing the non-redundant subset, where any molecule is described by  $N_a - 1$  bond lengths,  $N_a - 2$  bond angles and  $N_a - 3$  dihedral angles. However the exact choice is still ambiguous, and is quite unlikely to actually reflect the bonding.<sup>57</sup> For example, it is difficult to see how any dihedral angle is important to  $\text{NH}_3$ , yet one must be included in the Z-matrix convention.

Another option to remove the redundancy is to diagonalise the matrix  $\mathbf{B}\mathbf{B}^T$  (or  $\mathbf{B}\mathbf{M}^{-1}\mathbf{B}^T$ , where  $M_{ij} = m_i\delta_{ij}$  is the diagonal matrix of Cartesian masses). Of the resulting eigenvectors,  $3N_a - 6$  will have non-zero eigenvalues, and the corresponding linear combinations of the  $r_i$  are taken as the vibrational coordinates.<sup>59–61</sup> This approach has the benefit of incorporating every valence coordinate rather than outright ignoring them. The downside is that linear combinations are much harder to interpret physically, especially because an individual coordinate can become ‘delocalised’ across the molecule.<sup>59</sup> Another disadvantage here is that the  $\mathbf{B}$  matrix has to be evaluated at a specific equilibrium geometry. This could be problematic in fluxional molecules where a single global minimum is hard to define, but also means that non-redundancy is not guaranteed for large displacements from equilibrium.

For the most part, the impact of different coordinate choices on QCMD has not been tested, because the implementation presented later on does not actually need a choice to be made. However, this would certainly be an interesting topic for future investigation. From the handful of ideas presented here, we might suspect that using linear combinations of valence coordinates (as described above) is best suited for QCMD, i.e.:

$$\xi_i = \sum_{j=1}^{N_r} a_{ij} r_j. \quad (3.4a)$$

The reason is that the quasi-centroid is defined by the mean of the  $\xi_i$  coordinates,

and the mean of a linear combination is equal to the linear combination of the means,

$$\bar{\xi}_i = \sum_{j=1}^{N_r} a_{ij} \bar{r}_j. \quad (3.4b)$$

Thus the quasi-centroid is influenced by the average structure of the entire molecule, and no aspect has to be outright ignored.

### 3.2 IMPLEMENTING CENTROID MOLECULAR DYNAMICS

In this section, the brackets

$$\left\langle \dots \right\rangle_{\bar{f}=F} = \frac{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{f} - F) (\dots)}{\int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{f} - F)},$$

are used to represent a path integral average where a mean property of the ring polymer,

$$\bar{f} = \frac{1}{n} \sum_{j=1}^n f(\mathbf{q}^{(j)})$$

is constrained at a value  $F$ .

For (Q)CMD, the target property is not the effective potential itself, but the gradient, which gives the forces needed for dynamics. The quasi-centroid forces are given in Eq. (2.36),

$$-\frac{\partial \bar{V}_n}{\partial \Xi_i} = - \left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial H_n}{\partial \xi_i^{(j)}} + \frac{1}{\beta} \frac{\partial}{\partial \Xi_i} \ln J(\Xi) - \frac{1}{\beta} \sum_{j=1}^n \frac{\partial}{\partial \xi_i^{(j)}} \ln J(\xi^{(j)}) \right\rangle_{\bar{\xi}=\Xi}.$$

In general the forces contain contributions from the polymer springs in  $H_n$ , and from  $J$ , the Jacobian for the  $\mathbf{q} \rightarrow \xi$  transform,

$$J(\xi) = \left\| \frac{\partial(q_1, \dots, q_N)}{\partial(\xi_1, \dots, \xi_N)} \right\|.$$

This is a fairly major complication because even if a coordinate transform can be defined, there is no guarantee that these extra force terms will be easy to calculate. In fact, when QCMD was originally proposed, these terms were neglected entirely,

$$-\frac{\partial \bar{V}_n}{\partial \Xi_i} \approx - \left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial \xi_i^{(j)}} \right\rangle_{\bar{\xi}=\Xi}, \quad (3.5)$$

on the basis that they largely cancel out anyway, provided the ring polymer is compact.<sup>26,27</sup>

A benefit of CMD is that these extra terms do not appear in the forces, Eq. (2.16),

$$-\frac{\partial \bar{V}_n}{\partial Q_i} = - \left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial q_i^{(j)}} \right\rangle_{\bar{q}=\mathbf{Q}},$$

because there is no transformation ( $\boldsymbol{\xi} = \mathbf{q}$  and  $J = 1$ ), and the ring polymer springs are invariant to translations of the centroid. This simplicity means that CMD is much easier to implement, and allows for a few creative tricks that can accelerate the calculations. This section gives an overview of these methods, and explains why nothing so far can be adapted well to QCMD.

To summarise the above equations in words, the forces at any point are a path integral average, with the (quasi-)centroid constrained. Naturally, evaluating these forces is the most expensive part of a (Q)CMD calculation: at first glance this would be a full-scale  $\mathcal{O}(nN)$  path integral simulation at every single time step. An efficient (Q)CMD implementation therefore has two main aims: speeding up these path integral averages, and minimising the number of full-scale force calculations required.

### 3.2.1 Faster Averages

The first main approach to efficient CMD is speeding up the path integral averages that give the centroid force.

### Gaussian Approximation

An ideal scenario would be if  $\bar{V}_n$  could be evaluated analytically, and require no path integral simulation at all. For CMD, this is possible if the classical potential is harmonic<sup>25</sup>, or at least harmonic on the length scale of the ring polymer

$$V(\mathbf{q}) \approx V(\mathbf{Q}) + \sum_{i=1}^N \frac{\partial V(\mathbf{Q})}{\partial Q_i} (q_i - Q_i) + \frac{1}{2} \sum_{ii'}^N \frac{\partial^2 V(\mathbf{Q})}{\partial Q_i \partial Q_{i'}} (q_i - Q_i) (q_{i'} - Q_{i'}). \quad (3.6a)$$

If this expansion is done about the centroid,  $\mathbf{Q} = \bar{\mathbf{q}}$ , then the ring polymer Hamiltonian, Eq. (1.15), is described entirely by the centroid and quadratic fluctuations of the  $k > 0$  normal modes

$$H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}}) \approx nH(\bar{\mathbf{p}}, \bar{\mathbf{q}}) + \sum_{k=1}^{n-1} \sum_{i=1}^N \frac{(\tilde{p}_i^{(k)})^2}{2m_i} + \sum_{k=1}^{n-1} \sum_{ii'}^N \frac{1}{2} \left[ m_i (\omega^{(k)})^2 \delta_{ii'} + \frac{\partial^2 V(\bar{\mathbf{q}})}{\partial \bar{q}_i \partial \bar{q}_{i'}} \right] \tilde{q}_i^{(k)} \tilde{q}_{i'}^{(k)}. \quad (3.6b)$$

With this approximation,  $e^{-\beta_n H_n}$  becomes Gaussian and Eq. (2.9) can be solved analytically to give the centroid potential.<sup>62,63</sup>

The main problem with this approach is that it requires the ring polymer to be compact enough that a harmonic approximation is valid. This means the method is unsuitable for low temperature, where the need for a lot of beads makes a performance boost most beneficial. It also only works because CMD is exact in the harmonic limit, so it is difficult to see how this could be adapted to QCMD where no analogous limit is known.

### Adiabatic Integration

A more generally applicable approach requires an actual path integral simulation to calculate the forces, with the aim to sample as much of the ring polymer distribution as possible, as quickly as possible. For PIMD, this comes down to two factors: a

simulation length just short enough to explore the whole distribution, and a time step just short enough to properly integrate the dynamics.

With this in mind, to properly integrate dynamics of frequency  $\omega$ , a time step should be chosen such that  $\Delta t \ll \omega^{-1}$ . However, the physical frequencies of a ring polymer range massively: the slowest and fastest normal modes are at  $\omega^{(1)} = 2\pi/(\beta\hbar)$  and  $\omega^{(n/2)} = 2n/(\beta\hbar)$  respectively. This means that the fast modes demand a small  $\Delta t$ , and then more steps are required to integrate the full dynamics of the slow modes.

Despite this issue, the dynamics of PIMD should not be taken literally; it is simply used to sample from the NVT ensemble. In fact, the momenta of the ring polymer only really exist to facilitate PIMD, as they can be integrated out analytically

$$\int dp e^{-\beta_n p^2/2m} = \sqrt{\frac{2\pi m}{\beta_n}}. \quad (3.7)$$

By extension, the ring polymer masses are also completely artificial, and the original choice<sup>64</sup> can be changed to be unique for each bead or normal mode,

$$\int d\tilde{p}_i^{(k)} e^{-\beta_n (\tilde{p}_i^{(k)})^2/2m_i} = \sqrt{\frac{m_i}{m_i^{(k)}}} \int d\tilde{p}_i^{(k)} e^{-\beta_n (\tilde{p}_i^{(k)})^2/2m_i^{(k)}}. \quad (3.8)$$

By adjusting these masses, the dynamics of individual normal modes can be accelerated or decelerated.

A common choice is to modify all the non-centroid ( $k > 0$ ) normal modes according to<sup>62</sup>

$$m_i(\omega^{(k)})^2 = m_i^{(k)}(\gamma\omega_n)^2, \quad (3.9)$$

to give

$$\int d^n \mathbf{p} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} = \left[ \prod_{k=1}^{n-1} \frac{\gamma\omega_n}{\omega^{(k)}} \right]^N \int d^n \mathbf{p} e^{-\beta_n H'_n(\mathbf{p}, \mathbf{q}; \gamma)}. \quad (3.10)$$

The new Hamiltonian, which now depends on the  $\gamma$  parameter, is

$$H'_n(\underline{\mathbf{p}}, \underline{\mathbf{q}}; \gamma) = \sum_{i=1}^N \frac{(\tilde{p}_i^{(0)})^2}{2m_i} + \sum_{k=1}^{n-1} \sum_{i=1}^N \left[ \frac{(\tilde{p}_i^{(k)})^2}{2m_i^{(k)}} + \frac{1}{2} m_i^{(k)} (\gamma \omega_n)^2 (\tilde{q}_i^{(k)})^2 \right] + \sum_{j=1}^n V^{(j)}. \quad (3.11)$$

The key difference compared to Eq. (1.15) is that every normal mode now has the same frequency,  $\omega^{(k)} \rightarrow \gamma \omega_n$ , which means they can be sampled on the same time scale as each other.\*

Because the masses of the  $k = 0$  mode were not altered, the dynamics of the centroid is unchanged. Furthermore, the remaining normal modes can be tuned to any frequencies, which can be much faster than the centroid. This allows for an adiabatic separation of the centroid and normal mode motion, and the centroid forces can be calculated on-the-fly.<sup>62</sup> The adiabaticity parameter  $\gamma$  controls the frequency of the normal modes, and thus the extent of adiabatic separation. (Note that some authors use  $\gamma^{-1}$  instead of  $\gamma$  — this is purely a convention.)

‘Adiabatic CMD’, or ACMD, comes with some costs. For one, the force on the centroid is instantaneous rather than an average, relying on the adiabatic separation to give mean force dynamics over time. This naturally leads to some statistical noise that is exacerbated with fewer beads.<sup>66</sup> However the bigger issue is that the faster modes need an integration time step that is about  $\gamma$  times smaller than what is acceptable for the centroid, meaning the overall number of time steps increases by a factor of  $\gamma$ .<sup>27</sup>

Generally, the  $\gamma$  parameter should be just large enough to achieve sufficient adiabatic separation — typically no greater than 100. However in the condensed phase, multiple studies show that one can actually get away with a surprisingly small  $\gamma$ .<sup>2,23,67</sup> One way to interpret this is as ‘partially adiabatic CMD’ (PACMD) where the normal modes are not necessarily ergodic.<sup>67</sup> Another is that the motion of the

---

\*Because the frequencies of the polymer will also be affected by the external potential, the relevant normal mode frequencies are actually  $\Omega^{(k)} = \sqrt{(\omega^{(k)})^2 + \Omega^2}$ , where  $\Omega$  is a frequency characteristic of the external potential. A better mass substitution is  $m_i(\Omega^{(k)})^2 = m_i^{(k)}(\gamma \omega_n)^2$ , which gives a better sampling of the total ensemble.<sup>65</sup>

ring polymer is closer to (T)RPMD except that the normal mode frequencies, which would usually contaminate the spectrum, have been shifted out of the frequency range of interest.<sup>2</sup>

Extending the adiabatic algorithm to QCMD is also possible, albeit with a few added difficulties.<sup>26,65</sup> Separating the centroid from the normal modes is very simple as the centroid is itself (proportional to) one of them. This is much more difficult in QCMD, where special algorithms are needed to constrain the quasi-centroid. Adiabatic CMD is already quite slow because of the need for small time steps, but it is even slower for QCMD as it fails to take advantage of partial adiabaticity: numerical tests benchmark it at  $8\times$  slower at simulating ice at 150 K. Finally, as published, it approximates the forces on the quasi-centroid, neglecting the Jacobian and spring terms.

### 3.2.2 Fewer Averages

Despite the improvements offered by adiabatic integration, evaluating the forces with path integrals at every time step is still a slow process. But in theory (Q)CMD is just classical dynamics, so these full-scale simulations should not really be necessary; if  $\bar{V}_n$  is known ahead of time, it really is just a classical simulation. This would take  $\mathcal{O}(N)$  computations per time step, instead of  $\mathcal{O}(Nn\gamma)$  offered by ACMD. Of course, some effort is required to generate  $\bar{V}_n$ , but it is then found once-and-for-all, massively speeding up future calculations.

This was the approach used in chapter 2, where the effective potential was calculated on a grid of points before any dynamics. Generally the size of the grid would grow exponentially with the number of degrees of freedom, so this technique is only appropriate for very low dimensional systems.

### Pseudo-potentials

One way to effectively lower the dimensionality of the system is to assume that the centroid potential is only pairwise additive,

$$\bar{V}_n(\mathbf{Q}) \approx \frac{1}{2} \sum_a^{N_a} \sum_{b \neq a}^{N_a} \bar{v}_n(|\mathbf{Q}_a - \mathbf{Q}_b|), \quad (3.12)$$

where  $\bar{v}_n(R)$  is the pair potential and  $\mathbf{Q}_a$  is the position vector of atom  $a$ . This assumption is generally untrue, even if the classical potential is also pairwise additive, but it does mean the problem of finding  $\bar{V}_n(\mathbf{Q})$  is reduced to finding a single one-dimensional function,  $\bar{v}_n(R)$ .

Using Eq. (2.16), the force from  $\bar{v}_n$  between atoms  $a$  and  $b$  is

$$\bar{v}'_n(|\mathbf{Q}_a - \mathbf{Q}_b|) = \frac{\mathbf{Q}_a - \mathbf{Q}_b}{|\mathbf{Q}_a - \mathbf{Q}_b|} \cdot \left( \frac{\partial \bar{V}_n}{\partial \mathbf{Q}_a} - \frac{\partial \bar{V}_n}{\partial \mathbf{Q}_b} \right), \quad (3.13a)$$

$$= \left\langle \frac{\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b}{|\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b|} \cdot \frac{1}{n} \sum_{j=1}^n \left( \frac{\partial V^{(j)}}{\partial \mathbf{q}_a^{(j)}} - \frac{\partial V^{(j)}}{\partial \mathbf{q}_b^{(j)}} \right) \right\rangle_{\bar{\mathbf{q}}=\mathbf{Q}}. \quad (3.13b)$$

The pair potential is dependent only on the scalar distance between two atoms. This allows for an extra simplification, because the path integral average only needs to constrain the scalar distance between the two atom centroids,

$$\bar{v}'_n(R) = \left\langle \frac{\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b}{|\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b|} \cdot \frac{1}{n} \sum_{j=1}^n \left( \frac{\partial V^{(j)}}{\partial \mathbf{q}_a^{(j)}} - \frac{\partial V^{(j)}}{\partial \mathbf{q}_b^{(j)}} \right) \right\rangle_{|\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b|=R}. \quad (3.13c)$$

One way to evaluate this would be to run a set of path integral simulations, each with a given bond constrained at a different value of  $R$ .<sup>62,68</sup> A faster solution is to generate the averages at different  $R$  values simultaneously, by running a simpler *un*-constrained simulation. Because  $a$  and  $b$  were chosen arbitrarily, at any given time step each of the  $N_a(N_a - 1)/2$  bonds can contribute to the average at their corresponding  $R$  value.<sup>41</sup>

The obvious limitation of this approach is the pairwise approximation, which will only be reasonable for a dilute system of identical particles.<sup>62</sup> In principle, a modification for QCMD might be possible, if bond distances were used as the curvilinear coordinates. However this runs into the problem outlined in section 3.1 as the number of pairs,  $N_a(N_a - 1)/2$ , quickly outnumbers the  $3N_a - 6$  required coordinates.

### Force Matching

More generally, the effective potential can take a functional form that depends on a set of parameters,  $\alpha$ . A good starting point is the classical limit, which is actually the original surface  $V(\mathbf{Q})$ . Quantum effects are then encapsulated by a small deviation,<sup>69</sup>

$$\bar{V}_n(\mathbf{Q}; \alpha) = V(\mathbf{Q}) + \Delta\bar{V}_n(\mathbf{Q}; \alpha) \quad (3.14)$$

which should be easier to fit than the full surface. The true potential should get the centroid forces correct, so the best approximation is the set of  $\alpha$  that minimises the error function<sup>69,70</sup>

$$\chi(\alpha)^2 = \left\langle \frac{1}{N} \sum_{i=1}^N \left( f_i^{\text{ref}}(\mathbf{Q}) + \frac{\partial \bar{V}_n(\mathbf{Q}; \alpha)}{\partial Q_i} \right)^2 \right\rangle. \quad (3.15)$$

The reference forces  $\mathbf{f}^{\text{ref}}(\mathbf{Q})$  are full path integral calculations, Eq. (2.16).

In early studies, the suggested functional form was a linear combination of basis functions,

$$\Delta\bar{V}_n(\mathbf{Q}; \alpha) = \sum_{i=1}^{N_\alpha} \alpha_i g_i(\mathbf{Q}), \quad (3.16)$$

because this reduces the problem to a set of linear equations which can be solved with singular value decomposition.<sup>69,71</sup> Since then, the meteoric rise of machine learning has enabled more sophisticated functional forms, and is likely to remain a very popular way of solving this problem.<sup>51,72</sup>

The issue with this approach, as with all the others considered so far, is that they require the centroid forces, and usually centroid-constrained path integral simulations. The quasi-centroid versions of these are inherently much more complicated, and so adapting these methods to QCMD seems difficult, even with additional approximations.

### 3.3 COARSE-GRAINING

To summarise sections 3.1 and 3.2, a straightforward extension of CMD to QCMD is practically impossible. The very first problem to solve, finding a set of curvilinear coordinates, is itself very difficult. Then even if a transformation can be found, the quasi-centroid forces are so complicated that adapting a current fast CMD method is difficult without some level of approximation. Really, QCMD needs to be implemented in a way that does not need forces from a path integral calculation, and confusingly, does not even define the quasi-centroid at all.

Even if they are unsuitable ‘out-of-the-box’, there are useful aspects of the fast CMD methods. In section 3.2.2, by assuming the effective potential is pairwise, the CMD problem is reduced to finding a one-dimensional function: the pair potential. This was solved through matching to path integral forces, but there is another way thanks to ‘Henderson’s theorem’:<sup>73</sup>

“...for quantum and classical fluids with only pairwise interactions, and under given conditions of temperature and density, the pair potential  $v(r)$  which gives rise to a given radial distribution function  $g(r)$  is unique up to a constant.”

What is even more interesting is the reverse implication: the ‘inverse Henderson problem’. Given a radial distribution function (DF), it is theoretically possible to find the unique pair potential that generated it.

In this case, Henderson’s theorem relates to the effective centroid pair potential

$\bar{v}_n(R)$ , and the centroid radial DF,<sup>†</sup>

$$\begin{aligned} & \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta\bar{H}_n(\mathbf{P},\mathbf{Q})} \frac{1}{2} \sum_a^{N_a} \sum_{b \neq a}^{N_a} \delta(|\mathbf{Q}_a - \mathbf{Q}_b| - R) \\ &= \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n\mathbf{p} \int d^n\mathbf{q} e^{-\beta_n H_n(\mathbf{p},\mathbf{q})} \frac{1}{2} \sum_a^{N_a} \sum_{b \neq a}^{N_a} \delta(|\bar{\mathbf{q}}_a - \bar{\mathbf{q}}_b| - R), \quad (3.17) \end{aligned}$$

which can be calculated with PIMD. Moreover, if this radial DF is known, no further path integral results are needed, and  $\bar{v}_n(R)$  can be deduced without ever calculating the force on the centroid. This was exactly how the CMD results in chapter 2 were obtained.

Solving the inverse Henderson problem is already a well-established technique in coarse-graining, which is the field dedicated to finding low-dimension (coarse-grained) models that reproduce certain properties of full-scale systems.<sup>74,75</sup> In essence, this is exactly what the (quasi-)centroid formalism is as well, where the  $(n \times N)$ -dimensional path integral is reduced to an  $N$ -dimensional model. This link has been known for some time;<sup>76,77</sup> for example, force matching was originally a coarse-graining method that was later applied to CMD.<sup>70,78,79</sup> Nevertheless, it seems that the coarse-graining literature would be very applicable to (Q)CMD, yet it remains relatively untapped.

Section 2.1 shows more specifically how the centroid formalism is coarse-graining, as  $\bar{V}_n(\mathbf{Q})$  is defined such that it reproduces the distribution of the centroid, Eq. (2.10),

$$\frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} e^{-\beta\bar{H}_n(\mathbf{P},\mathbf{Q})} = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n\mathbf{p} \int d^n\mathbf{q} e^{-\beta_n H_n(\mathbf{p},\mathbf{q})} \delta(\bar{\mathbf{q}} - \mathbf{Q}).$$

Of course, the full distribution of the centroid is an  $N$ -dimensional function, and practically impossible to calculate. But a useful consequence is that  $\bar{V}_n(\mathbf{Q})$  must

---

<sup>†</sup>An extra factor of  $4\pi R^2$  is optional.

reproduce *any* DF of the centroid,

$$\begin{aligned} \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta\bar{H}_n(\mathbf{P},\mathbf{Q})} \delta(f(\mathbf{Q}) - F) \\ = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{N_n}} \int d^n\mathbf{p} \int d^n\mathbf{q} e^{-\beta_n H_n(\mathbf{p},\mathbf{q})} \delta(f(\mathbf{q}) - F), \end{aligned} \quad (3.18)$$

and low-dimension DFs can be calculated. A carefully chosen set of these can be used as the target properties when constructing  $\bar{V}_n(\mathbf{Q})$ , without ever directly calculating the centroid force. This is really just a logical extension of the inverse Henderson problem, and is already a tried-and-tested approach in coarse-graining.<sup>74,75</sup>

On that point, it has been argued that if a potential contains at most  $n$ -body terms, then all DFs up to  $n^{\text{th}}$  order are needed to completely determine the system.<sup>80</sup> Immediately, this presents a computational trade-off. One-dimensional DFs are easy to calculate (see chapter 5) but lack correlations; higher order DFs include correlations, but are exponentially harder to calculate, let alone invert. Thankfully, this concern can be mitigated by borrowing another good idea from previous CMD implementations and basing  $\bar{V}_n(\mathbf{Q})$  on the underlying classical potential,

$$\bar{V}_n(\mathbf{Q}) = V(\mathbf{Q}) + \Delta\bar{V}_n(\mathbf{Q}). \quad (3.19)$$

As  $\Delta\bar{V}_n(\mathbf{Q})$  is a small perturbation, it should not be highly correlated, with most of the correlations captured by the classical surface. We therefore expect that one-dimensional DFs should be enough.

### 3.3.1 Fast Quasi-Centroid Molecular Dynamics

The naive extension of the above to the quasi-centroid is

$$\begin{aligned} \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \Xi)} \delta(f(\Xi) - F) \\ = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(f(\bar{\Xi}) - F). \end{aligned}$$

This does provide a route to  $\bar{V}_n(\Xi)$  without needing to calculate the quasi-centroid forces, but still needs the curvilinear transformation. However a much more interesting result occurs when  $f(\Xi)$  is one of the quasi-centroid coordinates,  $f(\Xi) \equiv \xi_i$ ,

$$\begin{aligned} \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{P} \int d\mathbf{Q} e^{-\beta \bar{H}_n(\mathbf{P}, \Xi)} \delta(\Xi_i - F) \\ = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} \delta(\bar{\xi}_i - F). \quad (3.20) \end{aligned}$$

By definition,  $\bar{V}_n(\Xi)$  must reproduce the DF of  $\bar{\xi}_i$ : the mean value of  $\xi_i$  averaged around the ring polymer. But more importantly, this relationship does not need any of the other quasi-centroid coordinates to be known, just that  $\xi_i$  is one of them.

This suggests a new procedure for finding the effective potential:<sup>8</sup>

1. Pick a set of functions  $\{\xi_1(\mathbf{q}), \xi_2(\mathbf{q}), \dots\}$ .
2. For each of these functions, calculate the distribution function of  $\bar{\xi}_i$ .
3. Fit the effective surface  $\bar{V}_n$  such that it reproduces these distribution functions.

Of course the  $\xi$  functions still have to be chosen, but there is much less restriction now because each DF depends on its corresponding function only. In other words,  $\xi$  does not have to be a coordinate transformation of  $\mathbf{q}$ , just any set of functions. In fact, there is not even a constraint on the number of  $\xi$  variables, because each one is just another DF used to fit the potential.

As explained in section 3.1, the logical choice for  $\xi$  is to base them on the valence of the molecule, i.e.: bond lengths, bond angles and torsion angles. However with this new approach, the previously mentioned redundancy is not a problem — we are free to use more than  $3N_a - 6$  coordinates. To recap, the logic behind valence-based coordinates is that they are most associated with the potential energy of the molecule, which has two main benefits. Firstly, these properties are least likely to suffer from the curvature problem, and will produce sensible DFs. Secondly, these DFs will parameterise  $\bar{V}_n$  in an effective and meaningful way. Another benefit is that it might be possible to exploit molecular symmetry and reduce the work even more. For example,  $\text{CH}_4$  has 4 C–H bond lengths and 6 H–C–H bond angles, but only 1 bond length distribution and only 1 bond angle distribution.

To date, this is the only general way to implement QCMD. It is also the fastest, because all the distribution functions can be obtained efficiently using the techniques in chapter 5 from a single unconstrained path integral simulation. Therefore, this approach is named ‘fast quasi-centroid molecular dynamics’, or f-QCMD.<sup>8</sup>

### 3.3.2 The Inverse Problem

The final piece of this puzzle is the actual inverse Henderson problem, i.e.: how to invert the DFs to get a potential. Fortunately, as it is *the* central problem to coarse-graining, there are already multiple solutions in the literature.<sup>74,75</sup>

The simplest and most widely used method is iterative Boltzmann inversion (IBI),<sup>80–82</sup> where the potential is iteratively improved,

$$V^{(k+1)}(\mathbf{Q}) = V^{(k)}(\mathbf{Q}) - \frac{1}{\beta} \sum_i^{N_\xi} \alpha_i \ln \left( \frac{G_i(\Xi_i)}{g_i^{(k)}(\Xi_i)} \right). \quad (3.21a)$$

using the classical DF at each stage of the iteration

$$g_i^{(k)}(\Xi) = \frac{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})} \delta(\xi_i(\mathbf{q}) - \Xi)}{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})}}. \quad (3.21b)$$

The process starts with the classical potential,  $V^{(0)}(\mathbf{Q}) \equiv V(\mathbf{Q})$ , and ends once it reproduces  $G_i$ : the target DFs. In this case, these are the distributions of  $\bar{\xi}_i$ .

In this work, all f-QCMD calculations were done with IBI. It is simple to implement, in this case with cubic splines. The iterations are by no means trivial, but are still fast classical calculations, especially paired with the techniques of chapter 5. In the wider field of coarse-graining, IBI can converge slowly<sup>83</sup> — the  $\alpha$  parameters are there to assist with this. However this did not turn out to be the case for f-QCMD, probably because it is only fitting a small perturbation to the starting classical potential. In all calculations, all  $\alpha$  parameters were set to 1.

Updating one degree of freedom may affect the DF of another; for example, modifying the bond stretching potential may alter the bond angle distribution. The order of the IBI updates is therefore important.<sup>80</sup> In all calculations here, all degrees of freedom were updated at every iteration, which seems to be most general approach.

Another consideration is that IBI fits the log of the distribution functions, which will go to zero when calculated numerically. Here, this was overcome by fitting a cubic spline through the log of the non-zero data points, which could then be extrapolated.

The other main approach to the inverse Henderson problem is inverse Monte Carlo, or IMC.<sup>84</sup> Suppose there is a small change made to the potential,  $\delta V(\mathbf{Q})$ .

The change in a DF due this perturbation is approximated as

$$\delta g_i^{(k)}(\Xi) \approx \beta \frac{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})} \delta V(\mathbf{q}) \left[ g_i^{(k)}(\Xi) - \delta(\xi_i(\mathbf{q}) - \Xi) \right]}{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})}}. \quad (3.22)$$

This is then used in an exact Newton method. The benefit of this approach is the exact update scheme, which can give faster convergence. It also fits the DFs directly, instead of their logs. The major drawback is the need to calculate  $\delta g_i$ , because unlike a one-dimensional DF, these contain cross correlations and take much longer to calculate.<sup>85</sup>

This is just a small handful of the ways the inverse Henderson problem can be solved, and new ideas are constantly put forward as the field of coarse-graining evolves.<sup>86,87</sup> These may be worth investigating in the future, but for now IBI has been completely sufficient.



## CHAPTER 4

# APPLICATION TO GAS-PHASE MOLECULES

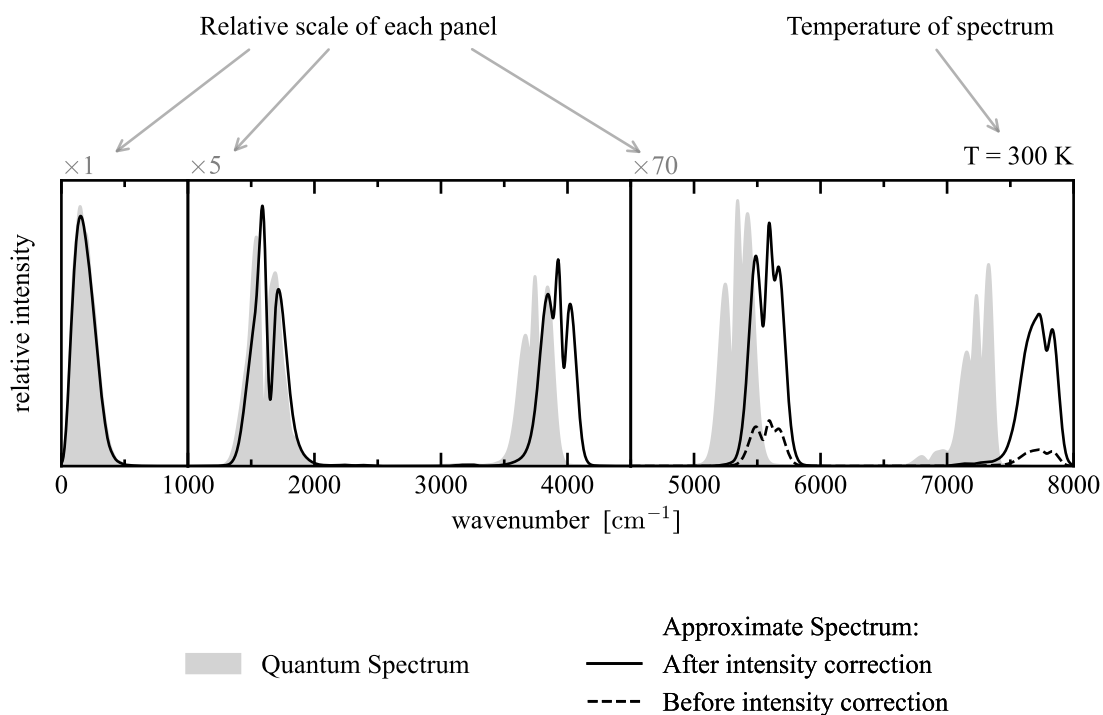
With the theoretical framework fully laid out, this chapter applies f-QCMD to the spectroscopy of small gas-phase molecules. These systems are simple enough that exact quantum calculations are possible, allowing a true benchmark of this methodology. In addition, for every molecule considered here, classical dynamics is used to assess how important quantum effects are, and CMD gives a reference for how severe the curvature problem is.

H<sub>2</sub>O and NH<sub>3</sub> are the only two molecules to have been studied with QCMD so far, which provides a point of comparison for the new fast implementation. However this already reaches the limits of QCMD. Trying to define a quasi-centroid for CH<sub>4</sub> runs into ambiguities, because there are too many bond lengths and angles, and the problem is overdetermined.<sup>8</sup> This is no problem for f-QCMD and accurate spectra can be easily obtained. Finally, H<sub>2</sub>O<sub>2</sub> presents an interesting test involving torsional motion, which comes with its own set of challenges.

This chapter also provides a good range of tests for the intensity correction scheme<sup>32</sup> from section 1.5. This allows for a more critical assessment of just how viable this technique would be beyond small model systems.

## EXAMPLE SPECTRUM

All the spectra presented in this chapter are formatted in exactly the same way — an annotated example is shown below. Any exceptions to this formatting are made obvious in the figures and their captions.



Quantum spectra are shown as shaded grey regions. Approximate (classical/path integral) calculations are shown as lines: dashed lines are the unaltered spectra, and solid lines are after the intensity correction of section 1.5 has been applied.

All spectra are normalised to give an integrated area of unity. This normalisation was also done after the intensity scaling, i.e: both the solid and dashed curves have the same integrated area. Since the overtone/combination bands tend to be much weaker and contribute very little to the overall intensity, the scaling has no noticeable effect on the fundamental regions.

## 4.1 COMPUTATIONAL DETAILS

Details of the fast implementation of CMD and QCMD used in this chapter are given in chapter 3, but the method can be summarised as follows:

1. Pick a set of functions  $\{\xi_1(\mathbf{q}), \xi_2(\mathbf{q}), \dots\}$ , which describe the valence of the molecule. In this work, the choice was made case-by-case for each molecule.
2. For each of these functions, calculate the distribution function of  $\bar{\xi}_i$ ,

$$\bar{\xi}_i = \frac{1}{n} \sum_{j=1}^n \xi_i(\mathbf{q}^{(j)}),$$

and  $\xi_i(\bar{\mathbf{q}})$ , where  $\bar{\mathbf{q}}$  is the centroid

$$\bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}^{(j)}.$$

These can be calculated efficiently and precisely using path integral molecular dynamics (PIMD) and the techniques described in chapter 5. Throughout, the PILE integration scheme<sup>14</sup> was used, in its original OBABO ordering and with a time step of  $\Delta t = 1$  atomic time unit. For every molecule considered here, convergence was achieved with 32 ring polymer beads at 300 K, and 64 beads at 150 K.

3. Fit an effective surface  $\bar{V}_n$  such that it reproduces these distribution functions. Specifically, the effective quasi-centroid potential should reproduce the distribution functions of  $\bar{\xi}_i$ , and the effective centroid potential should reproduce the distribution functions of  $\xi_i(\mathbf{q})$ .

This fitting was done using iterative Boltzmann inversion (IBI),<sup>80–82</sup> where the classical potential,  $V^{(0)}(\mathbf{Q}) \equiv V(\mathbf{Q})$ , is iteratively updated,

$$V^{(k+1)}(\mathbf{Q}) = V^{(k)}(\mathbf{Q}) - \frac{1}{\beta} \sum_i^{N_\xi} \ln \left( \frac{G_i(\Xi_i)}{g_i^{(k)}(\Xi_i)} \right).$$

Here,  $G_i(\Xi_i)$  are the distribution functions to be reproduced, and  $g_i^{(k)}(\Xi_i)$  are the classical distribution functions at each stage of the iteration

$$g_i^{(k)}(\Xi) = \frac{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})} \delta(\xi_i(\mathbf{q}) - \Xi)}{\int d\mathbf{q} e^{-\beta V^{(k)}(\mathbf{q})}},$$

which can also be calculated efficiently using the methods in chapter 5.

The correction terms were implemented by fitting cubic splines to  $\ln G_i$  and  $\ln g_i^{(k)}$ , ignoring any data points where the distribution functions vanish. The process was deemed to converge once the set of  $g_i^{(k)}$  reproduced  $G_i$  to graphical accuracy.

IBI is probably the simplest way to achieve this fitting (see section 3.3.2). Nevertheless, it was completely sufficient for the work here, and never needed more than 4 iterations, presumably because the starting point of the classical potential is already close to the final effective potential energy surface. The iterations for f-QCMD are shown in full in appendix B — the corresponding CMD calculations are not shown but converged just as quickly, if not quicker.

4. Using classical dynamics on  $\bar{V}_n$ , calculate  $\tilde{C}(t)$ : the autocorrelation function of the dipole time derivative,  $\dot{\boldsymbol{\mu}}$ . This was done using the  $n = 1$  limit of the PILE integrator (which is the same thing as velocity Verlet integration) and a time step of  $\Delta t = 1$  atomic time unit. For comparison, classical results were also obtained the same way, but using the classical potential.

Calculating  $\dot{\boldsymbol{\mu}}$  directly requires the gradient of  $\boldsymbol{\mu}$ , which would be costly and/or complicated to evaluate. To avoid this,  $\boldsymbol{\mu}$  was calculated at each time step, and the time derivative was approximated by finite difference,

$$\dot{\boldsymbol{\mu}}(t + \Delta t/2) \approx \frac{\boldsymbol{\mu}(t + \Delta t) - \boldsymbol{\mu}(t)}{\Delta t}. \quad (4.1)$$

The leading error in this approximation is  $\mathcal{O}(\Delta t^2)$ , which is no worse than the error

in the velocity Verlet algorithm. Thus provided  $\Delta t$  is small enough for the dynamics to be integrated properly, this approximation is also valid.

5. Calculate the spectrum using  $\tilde{C}(t)$  and Eq. (1.10). Every correlation function was damped using a Hann window function,

$$f(t) = \begin{cases} \cos^2(\frac{\pi t}{2\tau}), & -\tau \leq t \leq \tau, \\ 0, & \text{otherwise,} \end{cases} \quad (4.2a)$$

with  $\tau = 0.75$  ps. This produces the following spectral line shape

$$F(\omega) = \frac{\sin(\omega\tau)}{\omega(1 - (\omega\tau/\pi)^2)}. \quad (4.2b)$$

Generally, quantum spectra were produced from their corresponding line lists, which are sets of quantum transitions generated with the same potential energy surface and dipole moment surface. These can be calculated once-and-for-all and published, for example, in the ExoMol database.<sup>5</sup> These quantum transitions are then Boltzmann weighted to the correct temperature, and convolved with the line shape function  $F(\omega)$  to give the complete spectrum.<sup>6</sup> The only exceptions are the quantum spectra of H<sub>2</sub>O, which are taken from reference [9]. The authors are acknowledged for kindly providing this data.

## 4.2 H<sub>2</sub>O

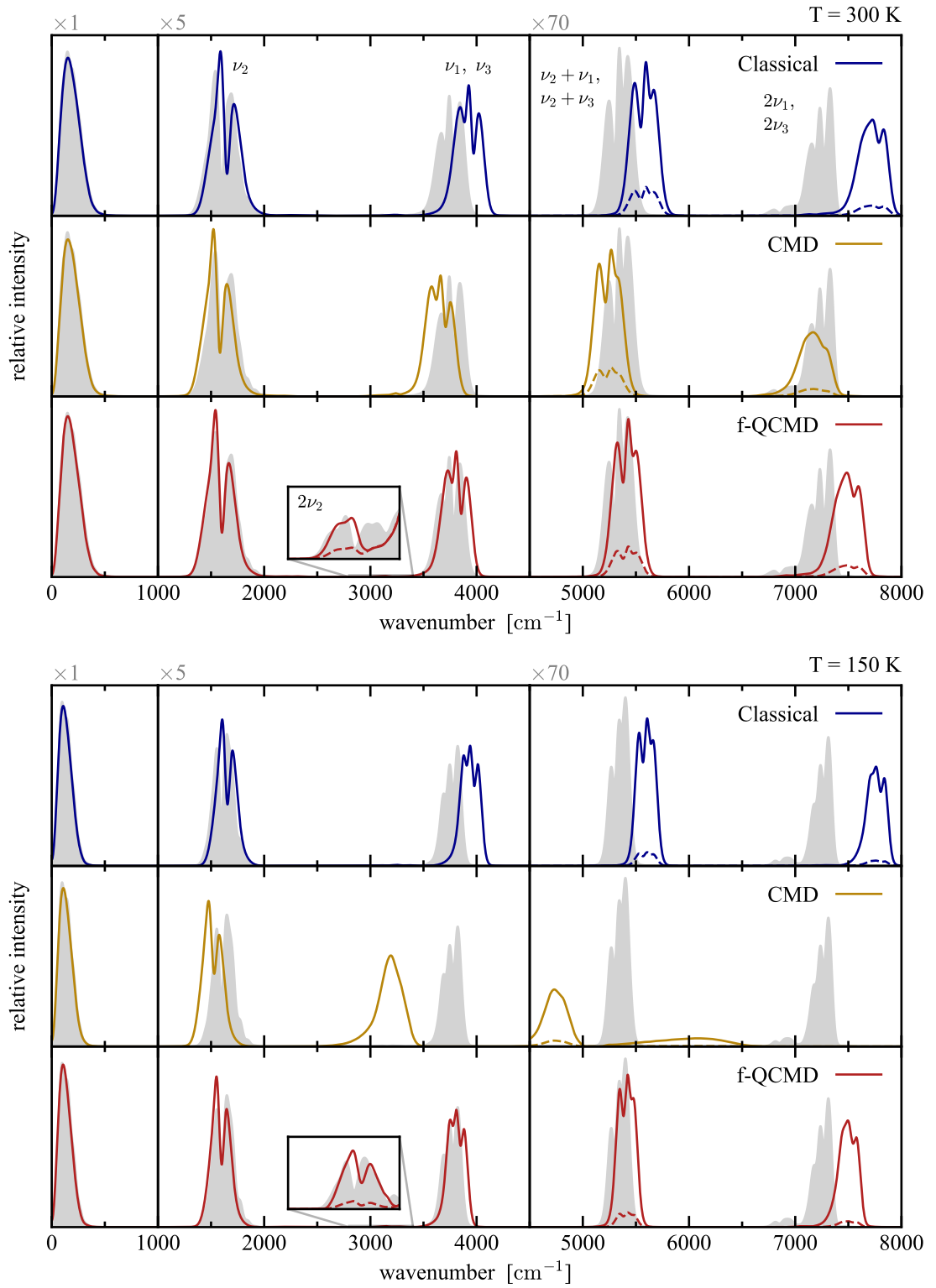
Gas-phase H<sub>2</sub>O was modelled using the potential energy surface from reference [34] and dipole moment surface from reference [35]. Quantum spectra and additional QCMD results were kindly provided by the authors of reference [9].

The O–H bond lengths and H–O–H bond angle were chosen as the  $\xi$  functions. This follows the original implementation of QCMD,<sup>26</sup> allowing for a fair comparison of the two methods. It is also, arguably, the most natural and intuitive description of the H<sub>2</sub>O valence. Nevertheless, the H–H bond length could have been used instead of the H–O–H angle — it turns out that the result is basically the same either way.<sup>88</sup>

Quantum and approximate spectra for gas-phase H<sub>2</sub>O at 300 K and 150 K are shown in fig. 4.1. In general, all the approximate methods get the spectral line shape correct. However the classical spectrum is blue-shifted slightly at both temperatures, and the CMD spectrum is red-shifted. This is only minor at 300 K, but much more severe at 150 K, and also much worse for the stretching modes ( $\nu_1$  and  $\nu_3$ ). This is consistent with the curvature problem, as explained in chapter 2. On the other hand, f-QCMD avoids both problems, and agrees well with quantum frequencies, especially in the fundamental region below 4500 cm<sup>-1</sup>.

f-QCMD does not perform so well in the overtone/combination region of the spectrum, especially with the stretching overtone bands ( $2\nu_1$  and  $2\nu_3$ ) where it overestimates the frequency considerably. Interestingly, the classical spectrum also blue-shifts this band much more than it does the fundamental, and at 300 K, CMD gets it about right despite underestimating the fundamental frequency. This suggests that classical and path integral methods systematically overestimate overtone frequencies.

On closer examination, the approximate methods place the overtone bands at exactly double the fundamental frequencies. This is consistent with the dynamics of a classical system, but *not* consistent with quantum mechanics. The energy



**Figure 4.1:** Spectra of gas-phase H<sub>2</sub>O at 300 K (top) and 150 K (bottom).

eigenvalues of an anharmonic system are spaced closer together at higher energies — for example, consider fig. 2.6. Therefore the frequency of an overtone ( $0 \rightarrow 2$ ) transition is not exactly double the fundamental ( $0 \rightarrow 1$ ) frequency, but slightly less. This effect can be seen in the quantum spectra. In essence, all (Q)CMD does is tweak the classical potential to change the fundamental frequencies, but as the dynamics itself is still classical, it neglects this higher-order anharmonic effect.

An interesting perspective on this problem is to use the textbook anharmonic potential: the Morse oscillator. With this potential, the frequency of a  $0 \rightarrow n$  transition has the form

$$\omega_{n0} = n\omega_e(1 - (n+1)x_e),$$

where  $\omega_e$  is the classical fundamental frequency, and  $x_e$  is a constant accounting for anharmonicity. As seen above, f-QCMD gives a good estimate of the fundamental frequency, i.e:

$$\omega_{10} = \omega_e(1 - 2x_e).$$

However, it predicts the overtone at twice this frequency, which does not coincide with the actual quantum energy levels,

$$2\omega_{10} = 2\omega_e(1 - 2x_e) > \omega_{20} = 2\omega_e(1 - 3x_e).$$

The discrepancy of  $2\omega_e x_e$  cannot be captured by f-QCMD alone. However, it can still be found with the tools at our disposal because it is exactly the amount that the classical fundamental frequency is blue-shifted<sup>8</sup>

$$\omega_e - \omega_{10} = 2\omega_e x_e.$$

Using fig. 4.1, the  $\nu_1/\nu_3$  fundamental bands are at about  $3930 \text{ cm}^{-1}$  with classical dynamics, and  $3810 \text{ cm}^{-1}$  with f-QCMD. It is by no means perfect, but red-shifting the f-QCMD overtones by  $120 \text{ cm}^{-1}$  would certainly give a much better prediction

of the true frequencies.

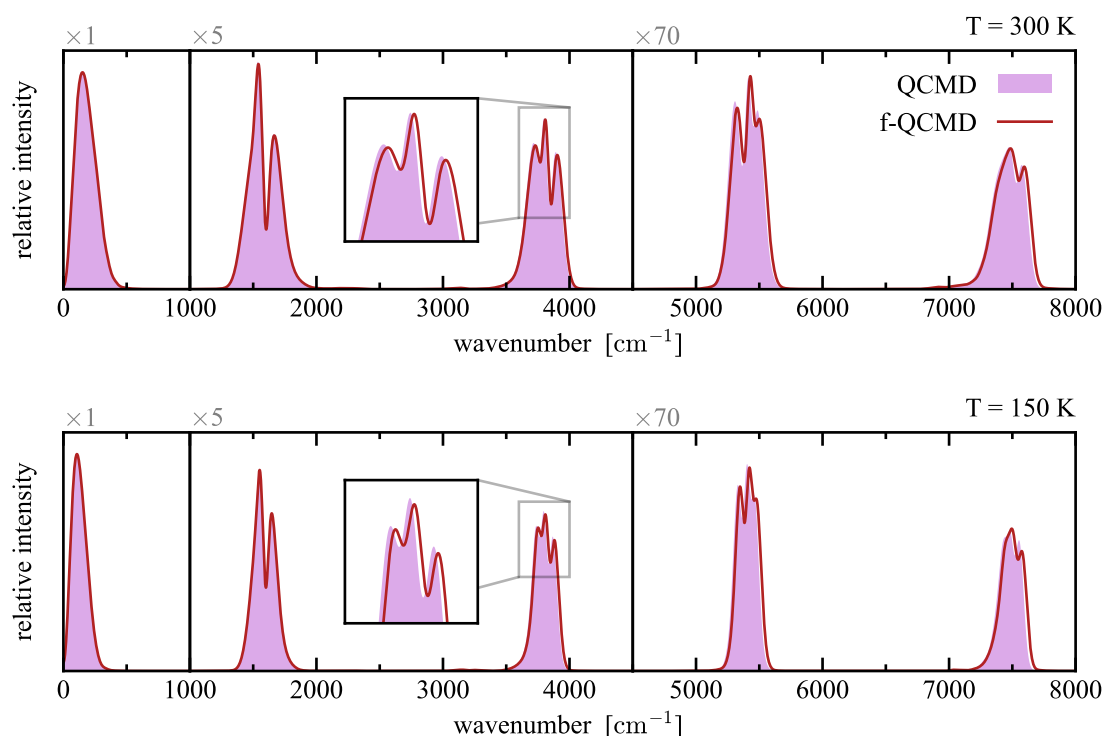
Combination bands tend to do better. Again, the approximate bands appear exactly at the sum of their fundamental frequencies. But provided the different modes are uncoupled from each other (and the doubly excited energy is just the sum of the two singly excited energies), this is also the case in quantum mechanics. Hence the frequency error for the approximate combination bands is just the combined error from the corresponding fundamentals.

On a similar point, fig. 4.1 demonstrates that classical and path integral methods systematically underestimate the intensity of the overtone and combination bands. Fortunately, the correction factor described in section 1.5 does an excellent job of fixing this issue.

Gas-phase H<sub>2</sub>O was the first molecule to be investigated by QCMD,<sup>9</sup> but some approximations were still made in its implementation. In these early studies, it is argued that, provided the ring polymer is sufficiently compact, the effect of the Jacobian and ring polymer springs on the quasi-centroid forces can be ignored<sup>26,27</sup>

$$-\frac{\partial \bar{V}_n}{\partial \Xi_i} \approx - \left\langle \frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial \xi_i^{(j)}} \right\rangle_{\bar{\xi}=\Xi}.$$

In contrast, f-QCMD does not assume this, but it does make an assumption about the form of the effective quasi-centroid potential (see section 3.3.1). A direct comparison of QCMD<sup>9</sup> and f-QCMD is shown in fig. 4.2. Despite their differences, the two methods produce practically identical spectra, which suggests both approximations are equally valid (for this problem).

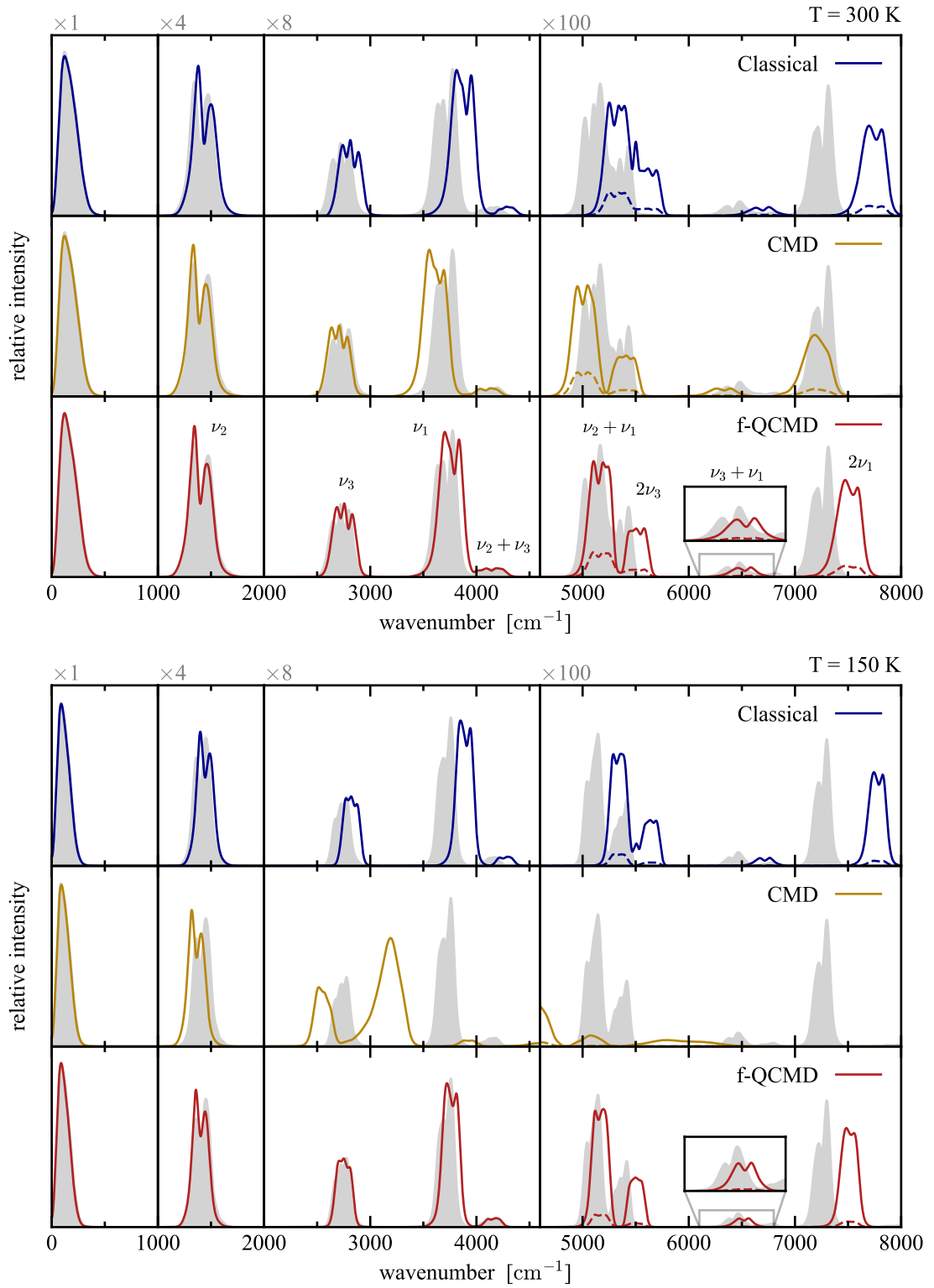


**Figure 4.2:** f-QCMD compared to an earlier implementation of QCMD,<sup>9</sup> for gas-phase  $\text{H}_2\text{O}$  at 300 K (top) and 150 K (bottom).

### 4.3 HDO

Gas-phase HDO was modelled using the potential energy surface from reference [89]. Quantum spectra were generated using the line list published in reference [90], which used this potential and the dipole moment surface from reference [91]. For the classical and path integral methods, this dipole moment failed to produce sensible spectral intensities, presumably because of bugs in the published source code. Instead, the previous dipole function<sup>35</sup> for  $\text{H}_2\text{O}$  was used; this replacement is actually quite similar in terms of its functional form, fitting and accuracy.<sup>91</sup>

Even though it is nearly identical to  $\text{H}_2\text{O}$ , quantum calculations of HDO are considerably more expensive, largely because the deuteration causes a reduction in symmetry.<sup>90</sup> f-QCMD calculations are also more complicated, but only slightly: instead of fitting both O–H potentials simultaneously, O–H and O–D are fit to separate quasi-centroid distribution functions. Everything else is unchanged.



**Figure 4.3:** Spectra of gas-phase HDO at 300 K (top) and 150 K (bottom).

Quantum and approximate spectra for gas-phase HDO at 300 K and 150 K are shown in fig. 4.3. For the most part, the key observations here are the same as for H<sub>2</sub>O. Classical dynamics over-estimates vibrational frequencies, whereas CMD under-estimates them due to the curvature problem, particularly at low temperatures. On the other hand, f-QCMD accurately predicts fundamental frequencies consistently. Overtone bands are systematically blue-shifted and not intense enough, but the latter can be fixed using the scaling described in section 1.5.

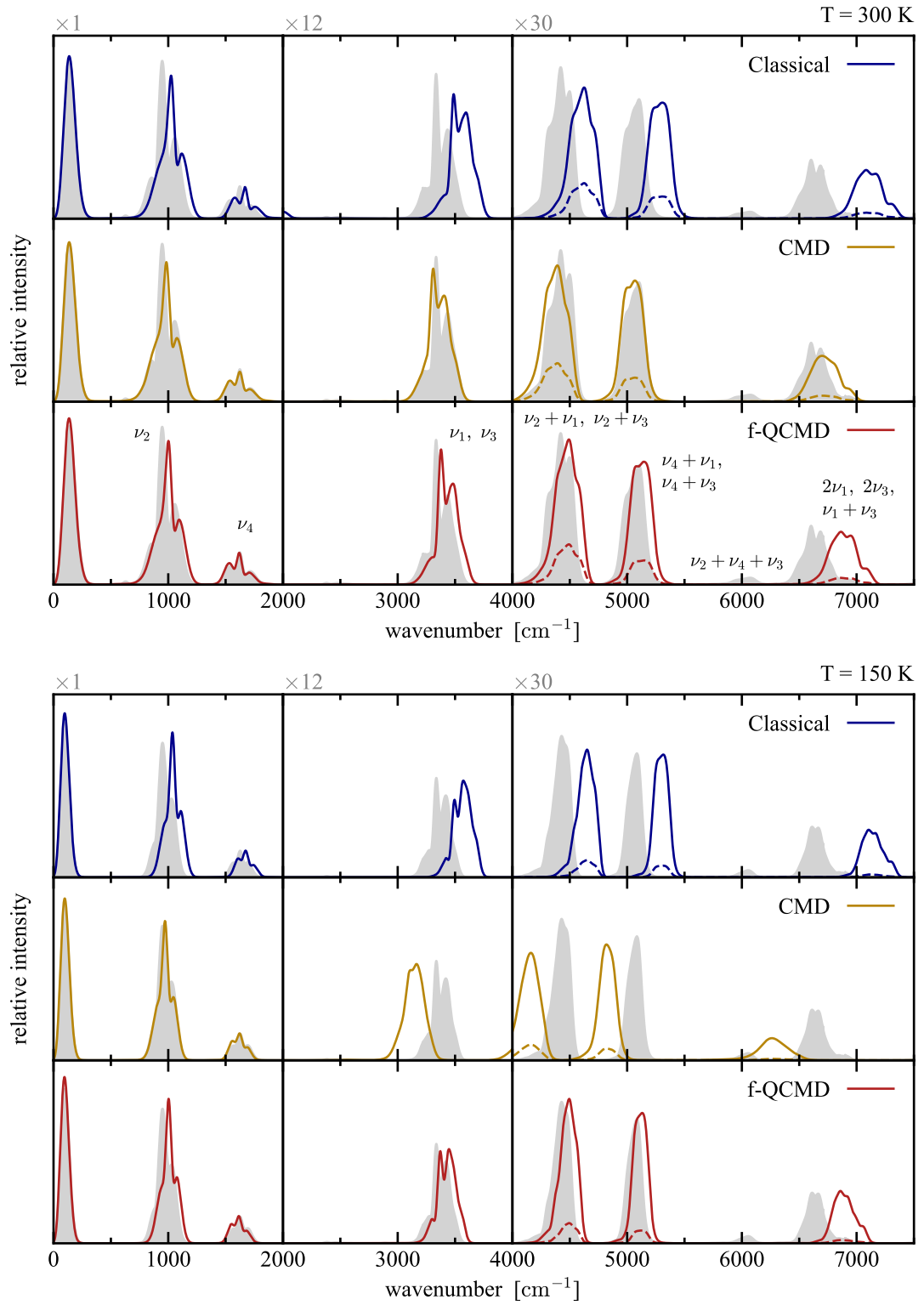
Compared to H<sub>2</sub>O, the O–H stretching band,  $\nu_1$ , is largely unchanged. However the increase in mass has significantly reduced the frequency of the O–D stretch,  $\nu_3$ , along with its combination and overtone bands. The increase in mass also has several interesting consequences for the approximate methods, because the dynamics of  $\nu_3$  is now more classical than  $\nu_1$ . This means the classical prediction of  $\nu_3$  is now noticeably better than  $\nu_1$ , and f-QCMD is almost perfect. The curvature problem is also much less pronounced for  $\nu_3$ , because the heavier mass produces stiffer ring polymer springs and a more compact bead distribution (see section 2.2).

In most cases, the width of vibrational bands is fairly accurate, but the detailed line shape is not predicted very well. This might be a consequence of using a different dipole moment surface to the reference quantum calculations.

## 4.4 NH<sub>3</sub>

Gas-phase NH<sub>3</sub> was modelled using the PES-2 potential energy surface and dipole moment surface from reference [33]. Quantum spectra were generated using the line list published in the same paper, which used these potential energy and dipole moment surfaces. Additional QCMD results are taken from reference [24], and were kindly provided by the authors.

f-QCMD extends naturally to NH<sub>3</sub>: the  $\xi$  functions are the three N–H bond lengths and three H–N–H bond angles. This also reflects previous QCMD studies of NH<sub>3</sub>,<sup>24</sup> providing a direct point of comparison.

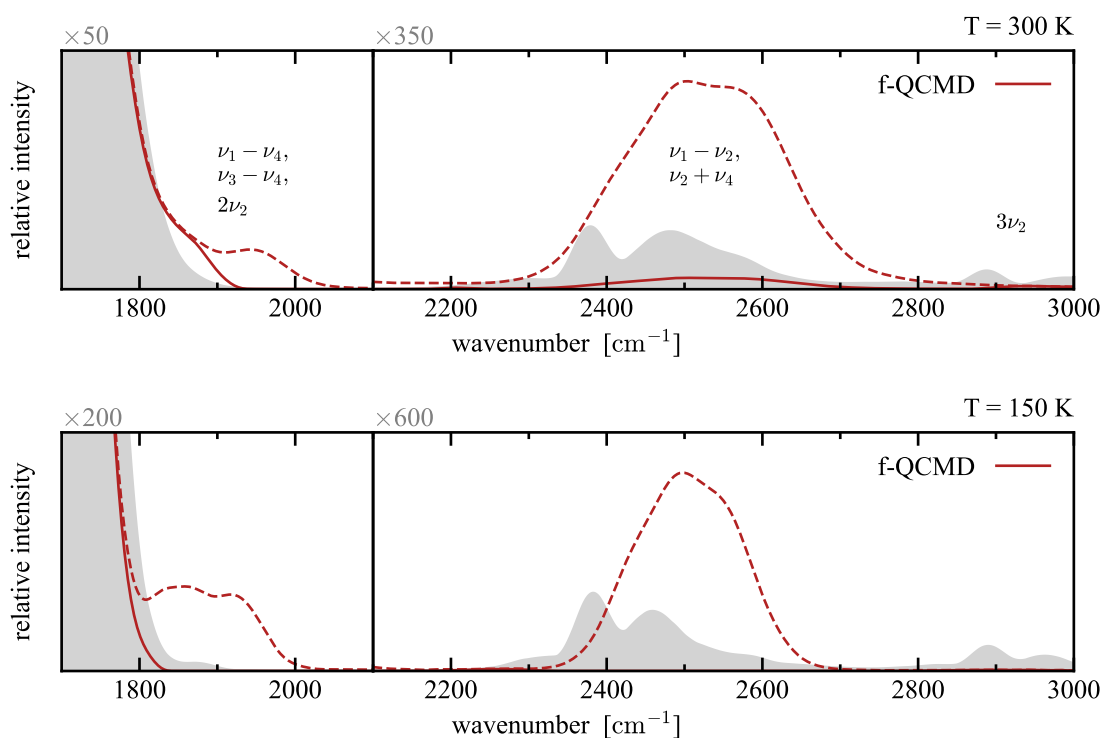


**Figure 4.4:** Spectra of gas-phase NH<sub>3</sub> at 300 K (top) and 150 K (bottom).

Quantum and approximate spectra for gas-phase  $\text{NH}_3$  at 300 K and 150 K are shown in fig. 4.4. As with  $\text{H}_2\text{O}$ , f-QCMD generally performs very well, and the same key observations about the spectrum can be made here too. At 150 K, f-QCMD is the clear winner as the curvature problem badly affects CMD. At 300 K, f-QCMD also performs well but CMD seems to do slightly better and agrees nearly perfectly with the quantum band positions. The reason for this is unclear; perhaps the curvature problem is gone entirely, or it might be present and provides a fortuitous amount of red-shift. Either way, this suggests that the temperature elevation modification to CMD<sup>51</sup> might perform very well here.

At first glance, the intensity correction factor also works well for  $\text{NH}_3$ , producing much better agreement for binary combination and overtone bands. However, closer inspection reveals some as-of-yet unseen limitations. For one, the bands at roughly  $2900\text{ cm}^{-1}$  and  $6000\text{ cm}^{-1}$  arise from transitions where more than two quanta change. No suitable correction factor has been derived for these cases, and the approximate intensities do not agree well with quantum results.

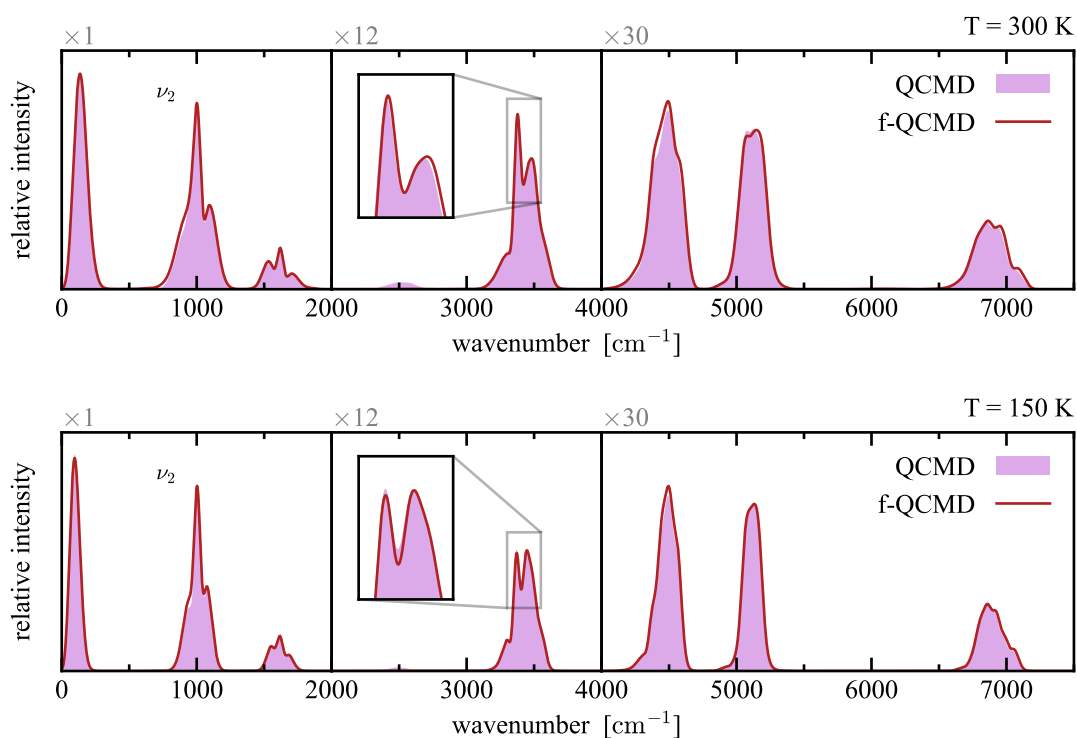
The  $\text{NH}_3$  spectrum also features noticeable difference bands. Opposite to combination bands, these should be very weak, but are actually too intense in classical and path integral predictions. For binary differences, the intensity correction factor can help with this and has been shown to work well when a difference band is taken in isolation.<sup>32</sup> The difficulty here is that the difference bands (which are too strong) overlap considerably with combination bands (which are too weak), and there is no clear answer on how to correct the intensity. This is demonstrated in fig. 4.5, which is a magnified look at the f-QCMD spectrum. The  $\nu_1 - \nu_4$ ,  $\nu_3 - \nu_4$  and  $2\nu_2$  bands overlap at about  $1900\text{ cm}^{-1}$ , and  $\nu_1 - \nu_2$  overlaps with  $\nu_2 + \nu_4$  at about  $2500\text{ cm}^{-1}$ . These were predicted to be too intense, and so the difference band correction was applied. At the very least this is a move in the right direction, but it tends to significantly overcompensate, presumably because of the combination bands that should have been amplified instead. Of course, knowing that the bands were too



**Figure 4.5:** A magnified view of fig. 4.4. The relative scale of each panel compared to fig. 4.4 is shown above.

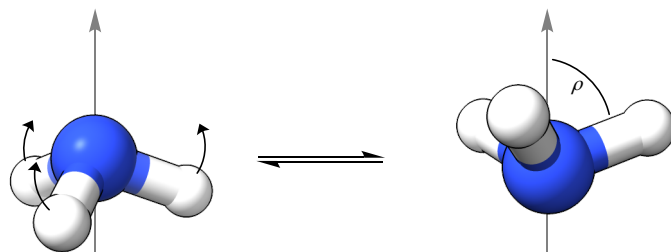
intense was only possible here because of a reference quantum spectrum. No such deduction would be possible *a priori*, and the best course of action would probably be to not attempt any kind of correction at all.

NH<sub>3</sub> is the only other molecule that the original QCMD method has been applied to,<sup>24</sup> and a comparison of these results to f-QCMD is shown in fig. 4.6. Here, QCMD was implemented using the adiabatic algorithm (see section 3.2.1), introducing further approximation on top of neglecting certain force terms. Nevertheless, both adiabatic and fast implementations produce nearly identical spectra. Although it is still very minor, the one noticeable discrepancy is the symmetric bending band,  $\nu_2$ , and by extension its combination bands. The reason for this is surprisingly elaborate and due to a particular quirk of NH<sub>3</sub> that makes a purely curvilinear approach difficult.



**Figure 4.6:** f-QCMD compared to an earlier implementation of QCMD,<sup>24</sup> for gas-phase  $\text{NH}_3$  at 300 K (top) and 150 K (bottom).

As demonstrated in fig. 4.7,  $\text{NH}_3$  is capable of inversion through a large-amplitude symmetric bend.<sup>92</sup> The adiabatic QCMD algorithm cannot handle large-amplitude motion such as this, and so the calculations in reference [24] simply discarded any trajectories that tried to invert. f-QCMD does not have to do this, which probably causes the slight discrepancy to QCMD. However the inversion problem is still not solved: looking at the  $\nu_2$  band in fig. 4.4, f-QCMD is still too high frequency and CMD gives surprisingly good agreement with the quantum spectrum.



**Figure 4.7:** Demonstration of the inversion of  $\text{NH}_3$  via large-amplitude symmetric bending ( $\nu_2$ ). Also shown is a geometric interpretation of the  $\rho$  coordinate.

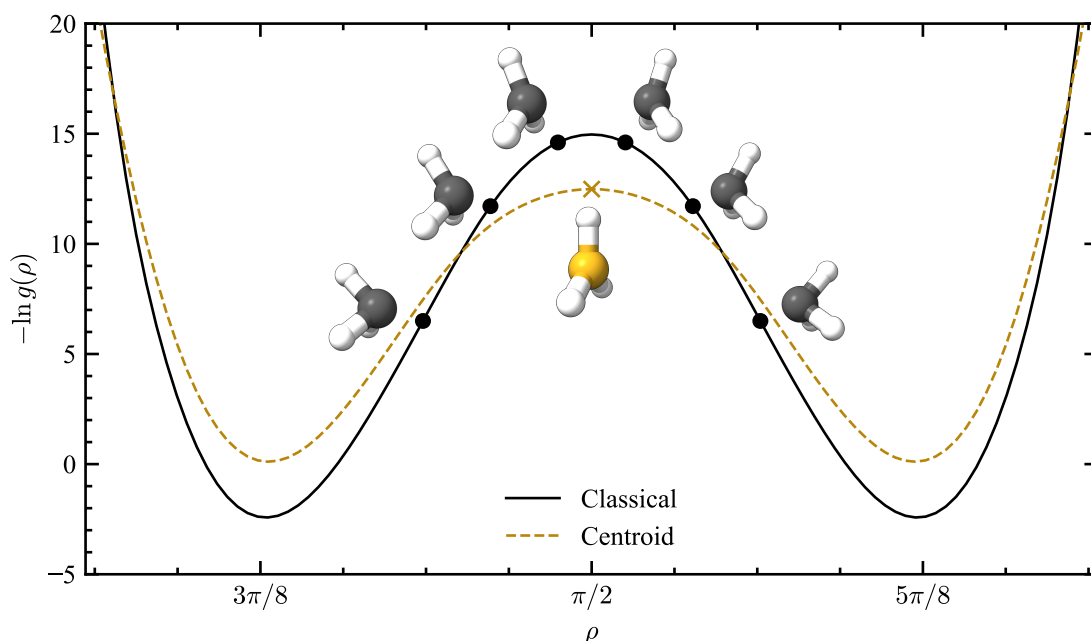
To better understand this, it is useful to consider the inversion coordinate  $\rho$ , defined such that

$$\sin \rho = \frac{2}{\sqrt{3}} \sin \left( \frac{\theta_{12} + \theta_{23} + \theta_{31}}{6} \right). \quad (4.3)$$

During the inversion, the molecule goes via a planar transition state, which occurs when  $\theta_{12} + \theta_{23} + \theta_{31} = 2\pi$  and  $\rho = \pi/2$ . Deciding whether a given molecule has  $\rho < \pi/2$  or  $\rho > \pi/2$  is just a matter of convention. For example, this could be decided by labelling each of the H atoms, and using the order of the atoms around the base of the NH<sub>3</sub> pyramid (which reverses during an inversion).

Suppose a ring polymer is at the transition state. Some beads have  $\rho < \pi/2$  and some have  $\rho > \pi/2$ , but on average the potential energy of the beads is lower than the transition state at  $\rho = \pi/2$ . This is demonstrated in fig. 4.8, where  $-k_B T \ln g(\rho)$  is used to represent the potential energy, and  $g(\rho)$  is the classical distribution function of  $\rho$ .

The Cartesian centroid of this ring polymer is also shown in fig. 4.8. The N–H



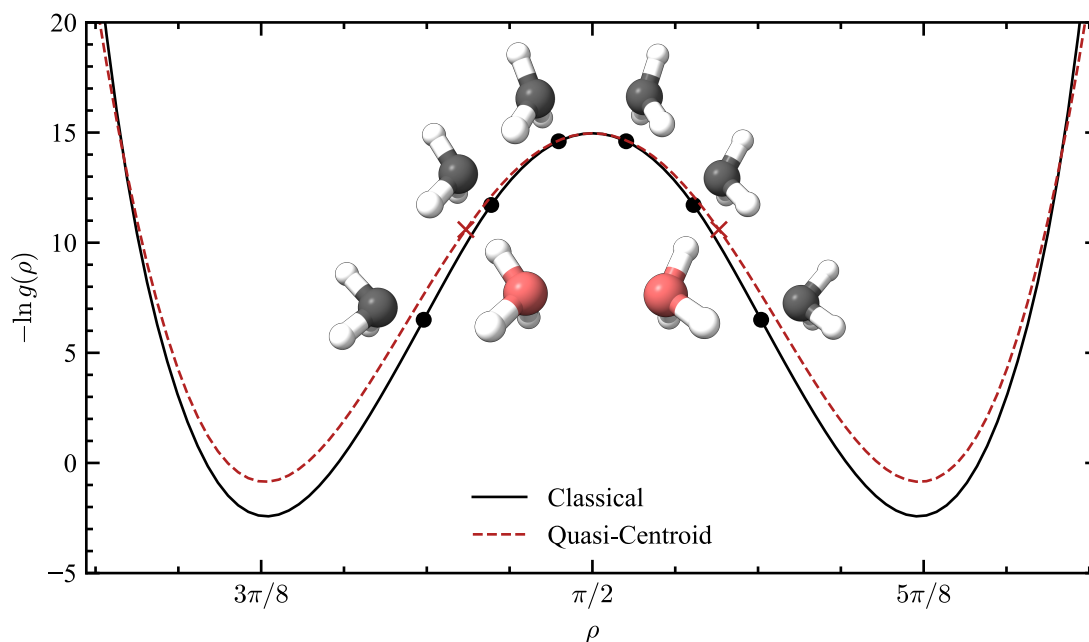
**Figure 4.8:** Demonstration of a ring polymer on the inversion pathway at 150 K.

**Black** An example ring polymer on the classical potential energy surface.

**Yellow** The ring polymer centroid and its effective potential energy surface.

bonds might be too short (i.e. the curvature problem), but the centroid is roughly planar. Thus, the effective centroid potential is stabilised at  $\rho = \pi/2$  and the barrier to inversion is lowered, which reflects the accelerated inversion due to tunnelling. This has also been represented by  $-k_B T \ln g(\rho)$ , but using the distribution of  $\rho(\bar{\mathbf{q}})$ . The end result for the spectrum is fairly impressive agreement for the  $\nu_2$  band.

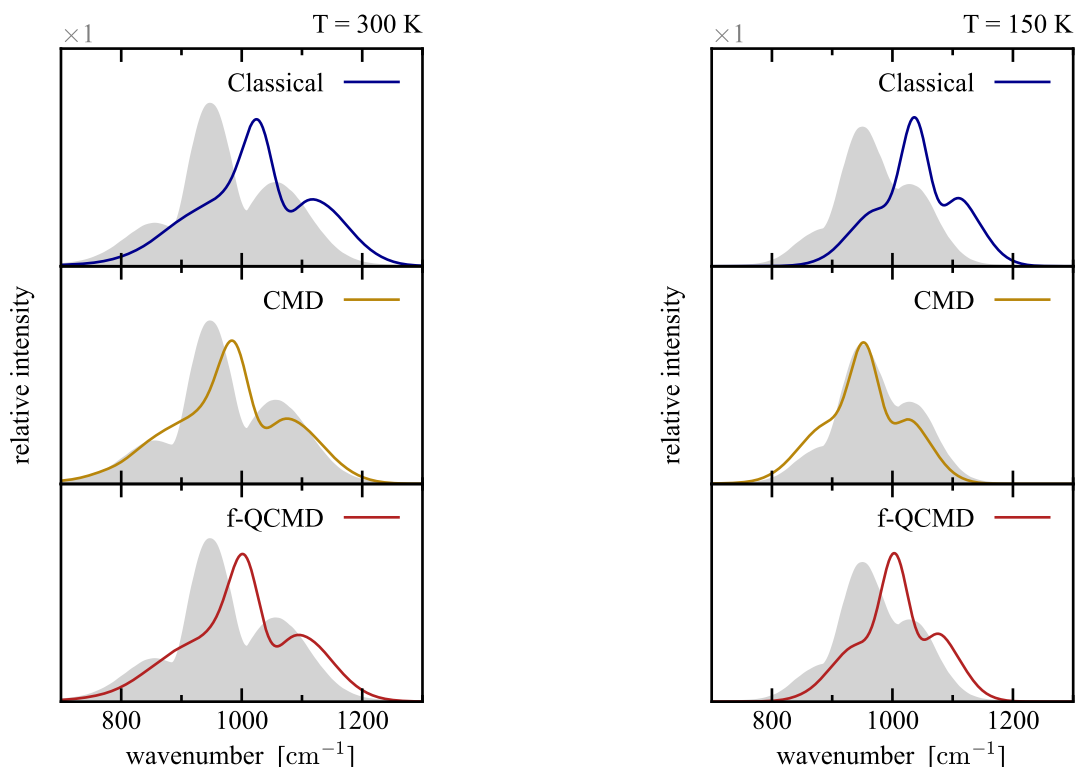
Although the centroid is fairly planar, all the ring polymer beads are in bent configurations. Therefore taking the mean of the bond angles gives a bent quasi-centroid, not the desired planar transition state. In fact, the only way the quasi-centroid can be planar is if all the beads are clustered together with  $\theta_{12} + \theta_{23} + \theta_{31} \approx 2\pi$ , effectively behaving like a single classical particle. The result is that the quasi-centroid potential barrier is not actually lowered as it should be. This is illustrated in fig. 4.9, where the distribution of  $\rho(\bar{\theta}_{12}, \bar{\theta}_{23}, \bar{\theta}_{31})$  is used to represent the quasi-centroid inversion potential. Note also that if we are provided with just the mean bond angles, it is impossible to determine which side of the inversion barrier the quasi-centroid is on.



**Figure 4.9:** Demonstration of a ring polymer on the inversion pathway at 150 K.

**Black** An example ring polymer on the classical potential energy surface.

**Red** The ring polymer quasi-centroid and its effective potential energy surface.



**Figure 4.10:** The  $\nu_2$  band from fig. 4.4, at 300 K (left) and 150 K (right).

Figure 4.10 takes a closer look at the symmetric bending band from fig. 4.4. Classical dynamics neglects quantum effects, and its spectrum is slightly blue-shifted. On the other hand, CMD somewhat captures the quantum effects of both zero-point energy and tunnelling, giving the best agreement with quantum. f-QCMD is somewhere in the middle: it describes zero-point energy, but not tunnelling.

The general idea behind QCMD is that, if a ring polymer is at equilibrium, its centroid might not resemble the bead geometries very well, but the quasi-centroid will. Interestingly, it is for this reason that QCMD fails to describe the barrier crossing, because the planar transition state does not resemble any of the beads.

The simplest attempt at fixing this would be to simply include  $\rho$  as another  $\xi$  function. However, as argued by Haggard et al.,<sup>24</sup> this is probably not worthwhile. For one thing, it is an added complication to a method that currently requires just two path integral distribution functions to work. Also, the  $\theta_{ij}$  angles and  $\rho$  are related non-linearly; hence it is uncertain how well f-QCMD would be able to

reproduce the distributions of  $\bar{\theta}_{ij}$  and  $\bar{\rho}$  simultaneously. Most importantly though, the classical  $\nu_2$  band is only out by about  $100\text{ cm}^{-1}$ , and the f-QCMD band is out by even less. This is a fairly minor discrepancy, especially considering how well f-QCMD performs across the wider spectrum.

## 4.5 CH<sub>4</sub>

Gas-phase CH<sub>4</sub> was modelled using the potential energy surface and dipole moment surface from reference [93]. Quantum spectra were generated using the line list published in reference [94], which used these potential energy and dipole moment surfaces.

During our path integral calculations, there were some instances of the CH<sub>4</sub> molecule dissociating. The reason is that the dispersion of the ring polymer pushed some replicas to highly non-equilibrium geometries. It is probably the case that these beads then experienced unphysical forces, because the potential energy surface was not parameterised to deal with such extreme situations. This is no fault of PIMD inherently, but is because the potential energy surface was used for a purpose it was never designed for. To solve this problem, if any replica dissociated, the trajectory was discarded and the ring polymer was reset to the classical equilibrium configuration. The PILE thermostat was then run briefly to equilibrate the beads before sampling of the distribution functions continued.

Following on from the previous sections, the logical next test of QCMD would be CH<sub>4</sub>. Even though there are just 5 atoms, this proves to be a significant challenge for the original implementation of QCMD. The reason is that there are 4 C–H bond lengths and 6 H–C–H bond angles, totalling 10 valence coordinates. However CH<sub>4</sub> only has  $(3 \times 5) - 6 = 9$  internal degrees of freedom, and the problem of defining the quasi-centroid is overdetermined (see section 3.1).

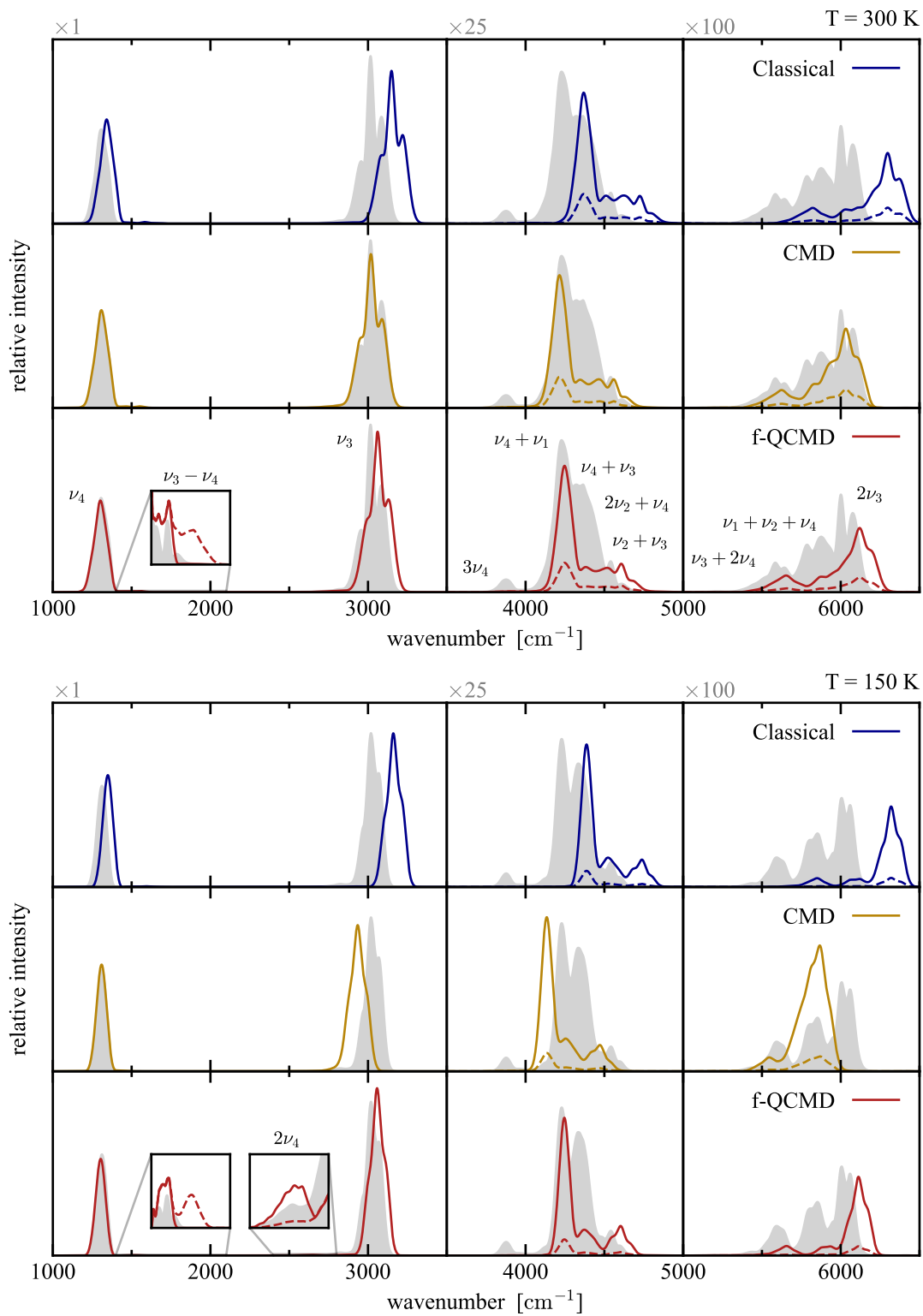
This is no problem for f-QCMD, because the quasi-centroid is never actually defined; instead, the 10 bond lengths and angles make up the  $\xi$  functions. It is also

worth noting that even though CH<sub>4</sub> is the most complex molecule considered so far, the implementation is just as simple, as the entire problem boils down to fitting just two distribution functions.

Quantum and approximate spectra for gas-phase CH<sub>4</sub> at 300 K and 150 K are shown in fig. 4.11. Methane does not have a permanent dipole moment, so the pure rotational part of the spectrum below 1000 cm<sup>-1</sup> is very weak and has been ignored. With regards to the band positions, any comments on fig. 4.11 are identical to previous sections. Once again, f-QCMD gives the most consistent agreement with quantum frequencies at both temperatures. Although it fails at low temperature, CMD does seem to do very well at 300 K. As with NH<sub>3</sub>, this may be because it is too hot for the curvature problem to have set in, or it may be because the curvature problem is weak and provides a fortuitous red-shift. Whatever the reason, further investigation, particularly with the temperature elevation method,<sup>51</sup> may be enlightening.

The fundamental region of fig. 4.11 is the simplest considered so far because only two modes are IR active: the  $\nu_4$  bending and the  $\nu_3$  asymmetric stretch. In contrast, the overtone/combination region is the most crowded and complicated, with many overlapping bands which makes it very difficult to apply the intensity correction. Furthermore, many of the significant bands come from tertiary combinations, which are beyond the scope of the simple correction factor in section 1.5. Even limiting ourselves to binary combinations, CH<sub>4</sub> has a more fundamental problem that IR inactive modes can produce IR active combination bands. With no detectable fundamental frequency, it is impossible to calculate the correction factor for these combination bands, or even predict where they are.

Here, these problematic bands were identified using the line list, as well as standard literature assignments.<sup>95</sup> For most of the bands above 3500 cm<sup>-1</sup>, intensity correction is impossible for at least one of the above reasons. The exceptions are  $\nu_4 + \nu_3$  and  $2\nu_3$ , and their correction factors were used to scale the entire spectrum in



**Figure 4.11:** Spectra of gas-phase  $\text{CH}_4$  at 300 K (top) and 150 K (bottom).

the regions 4000–5000 cm<sup>-1</sup> and 5000–6500 cm<sup>-1</sup> respectively. As one might expect, this does not give great agreement with quantum intensities. However it is a step in the right direction: combination bands are predicted to be too weak even if they are tertiary or come from IR inactive modes.

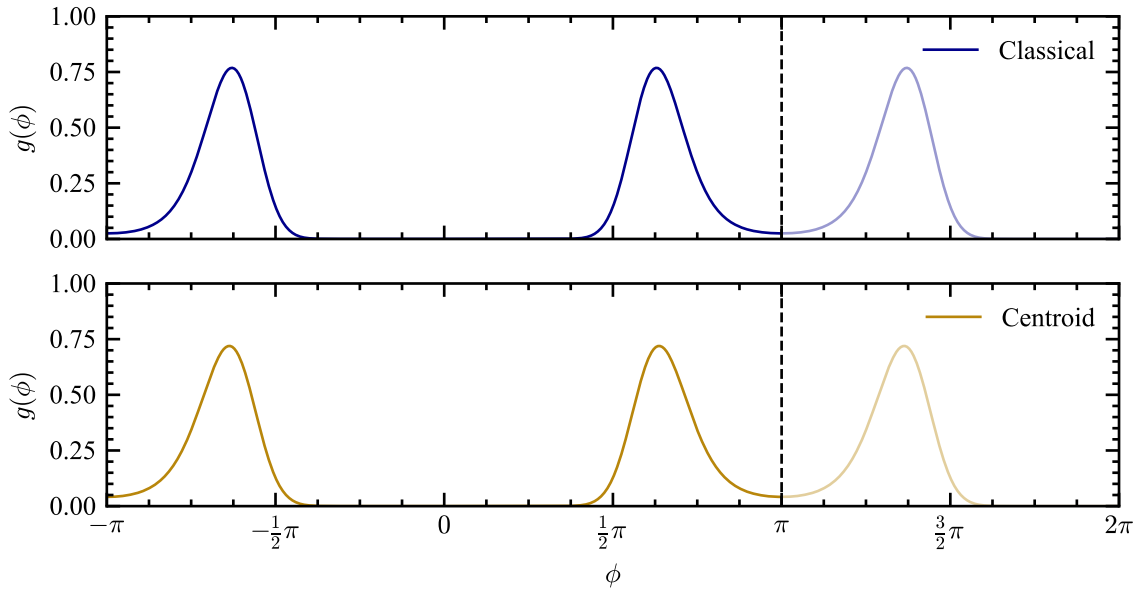
This underwhelming performance of the intensity correction is perhaps cause for concern. In the context of chemistry at large, CH<sub>4</sub> is still a very simple molecule and these problems are only going to get worse as the complexity grows. On reflection, this type of post-processing is probably best left to the simplest spectra with clear-cut binary combinations/overtone. In fact, fig. 4.11 has examples to support this, as the spurious intensities of  $\nu_3 - \nu_4$  and  $2\nu_4$  are easily identified and fixed with this formula.

## 4.6 H<sub>2</sub>O<sub>2</sub>

Gas-phase H<sub>2</sub>O<sub>2</sub> was modelled using the potential energy surface from reference [96] and dipole moment surface from reference [97]. Quantum spectra were generated using the line list published in reference [97], which used these potential energy and dipole moment surfaces.

H<sub>2</sub>O<sub>2</sub> is an interesting test as it is the simplest molecule to exhibit torsional motion. The  $\xi$  functions are as follows: the two O–H bond lengths, the two O–O–H bond angles, the O–O bond length and the H–O–O–H torsion angle.

Treating torsional motion with f-QCMD introduces some new problems. To demonstrate this, fig. 4.12 shows the classical and centroid torsion angle distributions at 150 K. Following convention<sup>98</sup> the torsion angle is measured in the interval  $[-\pi, \pi]$ , but the plots have been extended up to  $2\pi$ . As expected, fig. 4.12 shows that H<sub>2</sub>O<sub>2</sub> usually has a torsion angle of about  $\pm 2\pi/3$ . The two conformers can interchange via the trans geometry ( $\phi = \pi$ ), but virtually never by the cis ( $\phi = 0$ ).

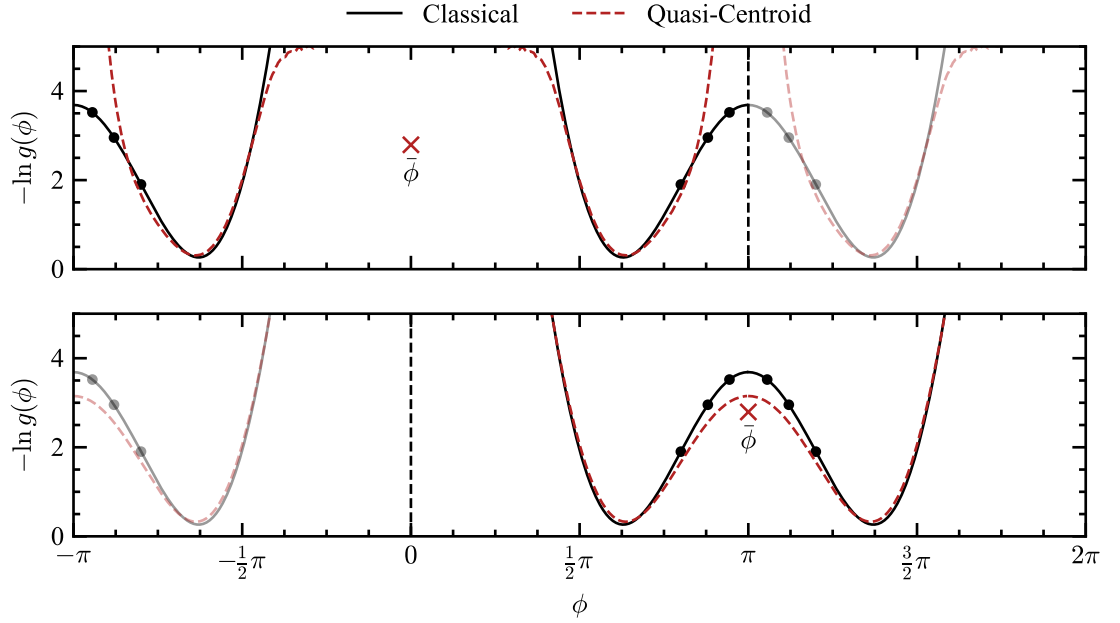


**Figure 4.12:** Distribution functions of the  $\text{H}_2\text{O}_2$  torsion angle at 150 K. The angle is measured in  $[-\pi, \pi]$  and the discontinuity is indicated on the plots.

Path integral methods should accelerate this interchange. Suppose a ring polymer has its centroid at  $\phi = \pi$ , at the top of the potential barrier. The beads of the ring polymer will be draped down either side of the barrier, at lower potential. This stabilises the centroid relative to the original potential, and effectively lowers the barrier to rotation. For more, see section 2.2.

This is not necessarily the case with the quasi-centroid. In the configuration described above, all the beads have  $\phi \approx \pm\pi$ , but some will have  $\phi > 0$  and some will have  $\phi < 0$ . Therefore the mean torsion angle,  $\bar{\phi}$ , will actually be close to 0, and is completely unrepresentative of the ring polymer. To demonstrate this, fig. 4.13 shows  $-k_B T \ln g(\phi)$ , which can be interpreted as a torsion potential. Also shown is an exemplary ring polymer in the trans geometry, and the location of  $\bar{\phi}$ .

In fact, actually achieving  $\bar{\phi} = \pi$  requires all the beads to be at  $\phi = \pi$  and on the same side of the discontinuity, which is practically impossible. The consequences for the quasi-centroid dynamics is a complete reversal of the proper behaviour of  $\text{H}_2\text{O}_2$ . Rotation via the trans geometry becomes impossible, and the cis barrier is slightly lowered because the quasi-centroid is spuriously found in the cis geometry



**Figure 4.13:** Demonstration of a ring polymer in the trans geometry. Also shown is the mean torsion angle,  $\bar{\phi}$ , and the effective quasi-centroid torsion potential.

**top**  $\phi$  measured in  $[-\pi, \pi]$

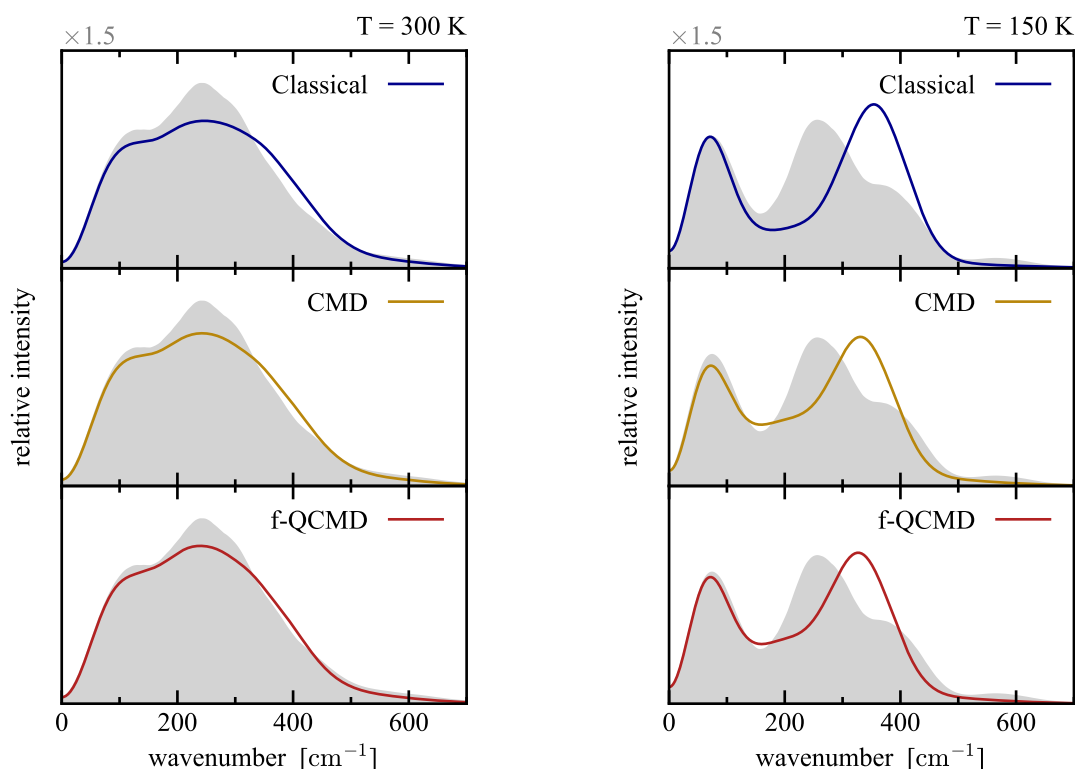
**bottom**  $\phi$  measured in  $[0, 2\pi]$

quite often. This is also visualised in fig. 4.13, where  $-k_B T \ln g(\phi)$  has also been plotted for the distribution of  $\bar{\phi}$ .

This problem arises because the torsion angle is periodic and therefore discontinuous, and the ring polymer stretches over this discontinuity. In the case of H<sub>2</sub>O<sub>2</sub>, this problem is easily solved by just measuring the angle in  $[0, 2\pi]$  instead, as rotation should never occur by the cis ( $\phi = 0$ ) geometry. More generally, a molecule might rotate freely such that there is no inaccessible angle. A solution here would be to measure the angles relative to the centroid,

$$\bar{\phi} = \phi(\bar{\mathbf{q}}) + \frac{1}{n} \sum_{j=1}^n (\phi^{(j)} - \phi(\bar{\mathbf{q}})) .$$

This ensures the measurement origin is within the ring polymer distribution, and the discontinuity is as far away as possible.



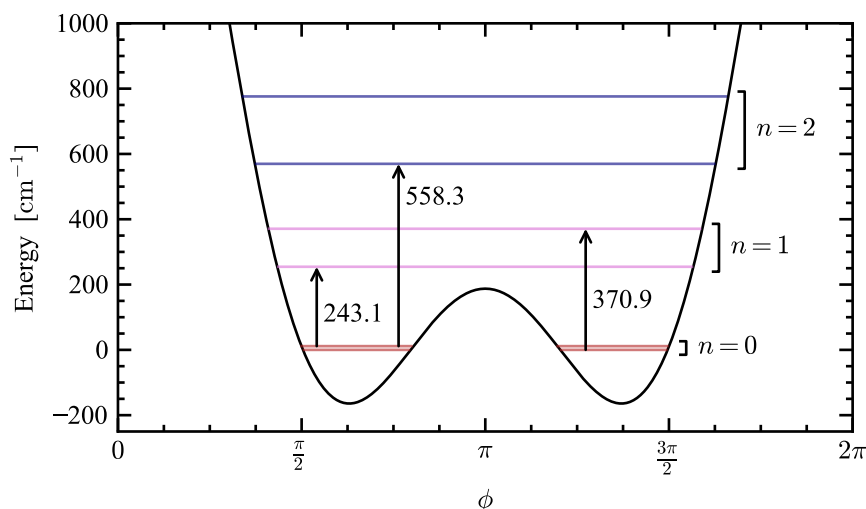
**Figure 4.14:** The rotation/torsion spectrum of gas-phase  $\text{H}_2\text{O}_2$  at 300 K (left) and 150 K (right). The indicated magnification is relative to fig. 4.16.

Classical dynamics, CMD and f-QCMD were used to generate the spectrum of  $\text{H}_2\text{O}_2$  at 300 K and 150 K. Figure 4.14 shows the lowest frequency part of the spectrum. The first peak arises from rotation, which is generally described quite well. The rest of the band arises from torsional ( $\nu_4$ ) motion.

At 150 K, the quantum torsion band has distinct peaks at about 250, 370 and 570  $\text{cm}^{-1}$ . This occurs because the torsion potential has two minima and two barriers connecting them: the cis and trans geometries. This causes each  $\nu_4$  level to split into two sub-levels,\* and gives a corresponding split in the excitation frequency.<sup>99,100</sup> This effect is shown in fig. 4.15.

On the other hand, all that the approximate methods do at this temperature is sample the bottom of either potential well, giving a single peak corresponding to this frequency. Both path integral methods capture some of the effects from zero-

\*Technically, there are 4 new levels but they are in pairs that are effectively degenerate.



**Figure 4.15:** Quantum torsional ( $\nu_4$ ) energies for H<sub>2</sub>O<sub>2</sub>, taken from reference [100]. The potential energy surface is represented by  $-k_B T \ln g(\phi)$  at  $T = 300$  K, where  $g(\phi)$  is the classical torsion angle distribution.

point energy and are slightly lower in frequency, but this does not really address the main issue. Put simply, the double well potential is extremely anharmonic, and path integral methods can never hope to capture all the quantum effects in the dynamics.

The quantum and approximate spectra change significantly upon increasing the temperature to 300 K. The quantum bands have broadened and aggregated into a single peak at about 250 cm<sup>-1</sup>. Presumably this is due to greater rotational broadening, but could also be due to new hot bands from the low-lying  $n = 1$  states. Torsional bands from the approximate methods are also much broader, but have now red-shifted into much better agreement with quantum. The reason is that the higher temperature facilitates barrier crossing, and much more of the anharmonicity of the potential is explored. All three approximate spectra are more similar than they are different, but f-QCMD seems to give the best agreement here, particularly around 400 cm<sup>-1</sup>.

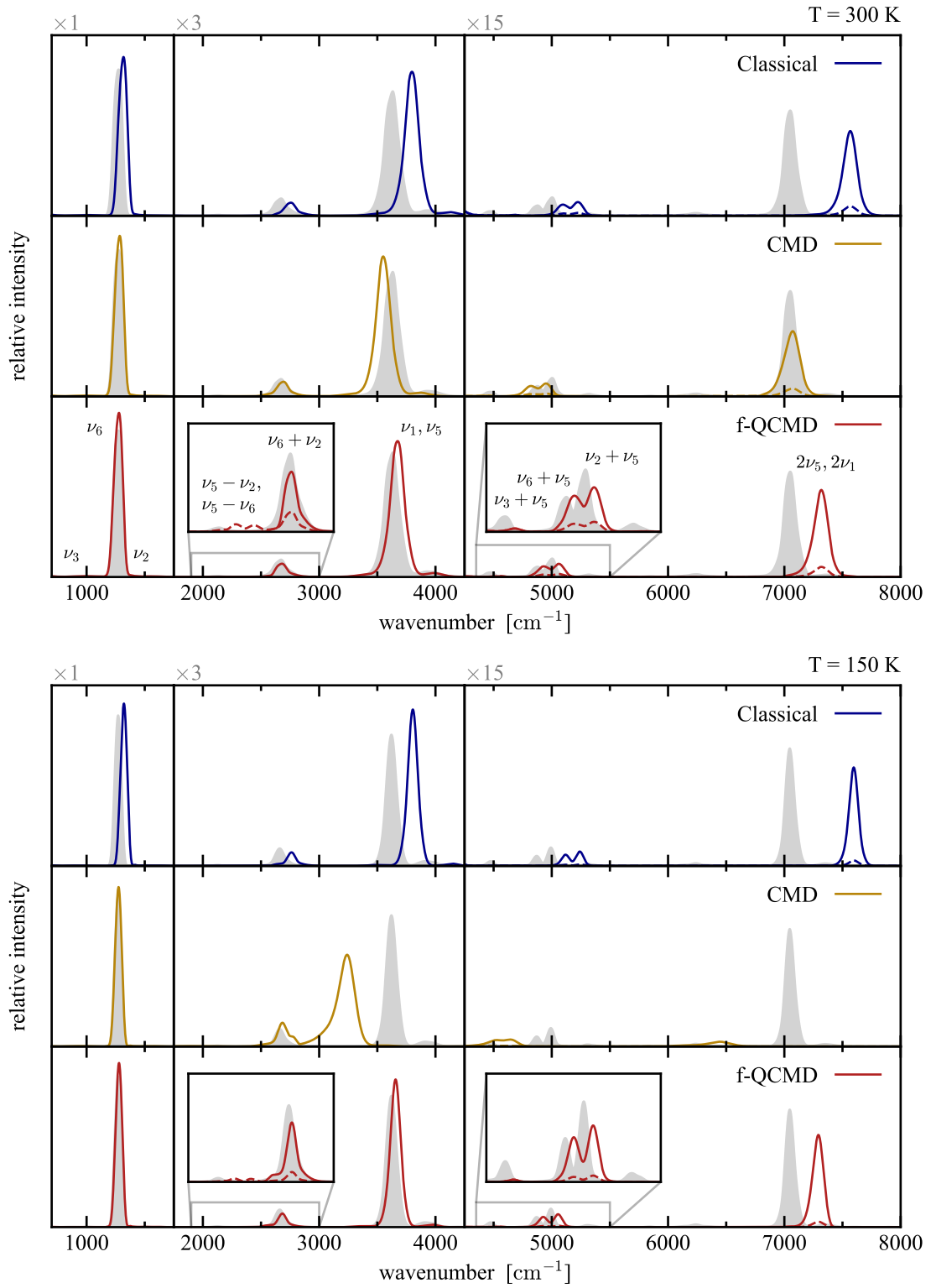
A final interesting point about the torsion in H<sub>2</sub>O<sub>2</sub> is that it is remarkably similar to the inversion ( $\nu_2$ ) motion of NH<sub>3</sub>, as both have double well potentials and problematic spectra. The difference is that the barrier to inversion in NH<sub>3</sub> is substantially higher, so the splitting of levels is much smaller<sup>92</sup> — path integrals

have no problem reproducing the quantum line shape in fig. 4.10, which suggests these effects are unnoticeable under these conditions. The failure of  $\text{NH}_3$  was down to the averaging of the curvilinear coordinates which meant the quasi-centroid was unrepresentative of the ring polymer distribution. Something very similar would have happened in  $\text{H}_2\text{O}_2$  had the torsion angle been measured in  $[-\pi, \pi]$ , but this problem is thankfully avoidable here (see above).

Yet another similarity is that  $\text{NH}_3$  and  $\text{H}_2\text{O}_2$  both have 4 atoms and their f-QCMD implementations used exactly  $(3 \times 4) - 6 = 6$  curvilinear coordinates. This suggests that an adiabatic QCMD implementation of  $\text{H}_2\text{O}_2$  is possible, as it was for  $\text{NH}_3$ . The caveat is that the adiabatic algorithm cannot handle large-amplitude motion and trajectories where  $\text{NH}_3$  inverted had to be discarded.<sup>24</sup> This was excusable because the barrier to inversion is high, making it a rare enough event to not impact the spectrum. However the isomerisation of  $\text{H}_2\text{O}_2$  is much easier and simply discarding the many trajectories that do so might not be possible, casting doubt on whether an adiabatic implementation is possible at all.

The remainder of the spectrum of  $\text{H}_2\text{O}_2$  is shown in fig. 4.16. Once again, f-QCMD gives the best and most consistent prediction of band frequencies. In particular the  $\nu_6$  band, which corresponds to O–O–H bending, is predicted perfectly. Presumably, this is because the presence of two heavy O atoms gives weaker quantum effects that are more easily captured by path integrals. One downside is that all approximate methods tend to produce fundamental bands that are too intense, which is likely because the torsional bands are underestimated and the overall spectra have been normalised to have the same area.

With 6 non-degenerate vibrational modes,  $\text{H}_2\text{O}_2$  has a complex overtone and combination spectrum. A big contributor to this is the  $\nu_4$  torsion band. As seen in fig. 4.15, this has multiple frequency components, and low-lying excited states make difference bands thermally accessible. To avoid this complication, no effort was made to locate or scale torsion-vibration interactions — these effects were treated more



**Figure 4.16:** Spectra of gas-phase H<sub>2</sub>O<sub>2</sub> at 300 K (top) and 150 K (bottom).

like rotational broadening than combination bands. This assumption is actually quite valid because a torsional frequency is much smaller than any other vibration,  $\Omega_4 \ll \Omega_j$ , such that Eq. (1.54) reduces to

$$\frac{I(\Omega_j \pm \Omega_4)}{\lim_{\hbar \rightarrow 0} I(\Omega_j \pm \Omega_4)} \approx 1 \pm \frac{\beta \hbar \Omega_4}{2} \coth\left(\frac{\beta \hbar \Omega_j}{2}\right) + \mathcal{O}((\beta \hbar \Omega_4)^2). \quad (4.4)$$

This tends to 1 if  $\beta \hbar \Omega_4$  is small (which it is), meaning the approximate dynamics should reproduce the correct intensities anyway.

As with  $\text{CH}_4$ , some of the other combinations arise from modes whose fundamental bands are absent in fig. 4.16.<sup>101</sup> The O–O stretching band,  $\nu_3$ , is very weak and should appear around  $863 \text{ cm}^{-1}$ ; the  $\nu_3 + \nu_5$  band was therefore not scaled. The  $\nu_2$  band (another O–O–H bend) should appear at  $1394 \text{ cm}^{-1}$ , but it is also weak and mostly covered by the  $\nu_6$  band. For intensity scaling, the peak frequency around  $1300 \text{ cm}^{-1}$  was used for both  $\Omega_6$  and  $\Omega_2$ . This explains why  $\nu_6 + \nu_2$  and  $\nu_2 + \nu_5$  are still quite low in intensity: using a slightly larger value of  $\Omega_2$  would have given a slightly larger scaling and better agreement with quantum intensities.

## CHAPTER 5

# ESTIMATORS FOR DISTRIBUTION FUNCTIONS

Distribution functions play an important role across chemistry, and are an essential component of f-QCMD. The naive approach to calculating distribution functions is by means of a histogram. However, this method is fundamentally flawed as the variance of the estimation is tied to its resolution: a higher resolution distribution needs more samples to properly converge. Because of this, and given the vital role of distribution functions, it is well worth taking the time to find a better way to calculate them.

The solution is a class of reduced variance estimators,<sup>102–104</sup> which are the primary focus of this chapter. These estimators are in fact remarkably general, and are applicable to any probability distribution with very little added complication. Specific formulations are derived for averages in both classical and path integral ensembles. For the sake of f-QCMD, these ideas are specialised to some of the most important molecular properties: bond lengths, bond angles and torsion angles. Finally, appendix C gives some consideration to how these estimators are best implemented computationally.

### 5.1 REDUCED VARIANCE ESTIMATORS

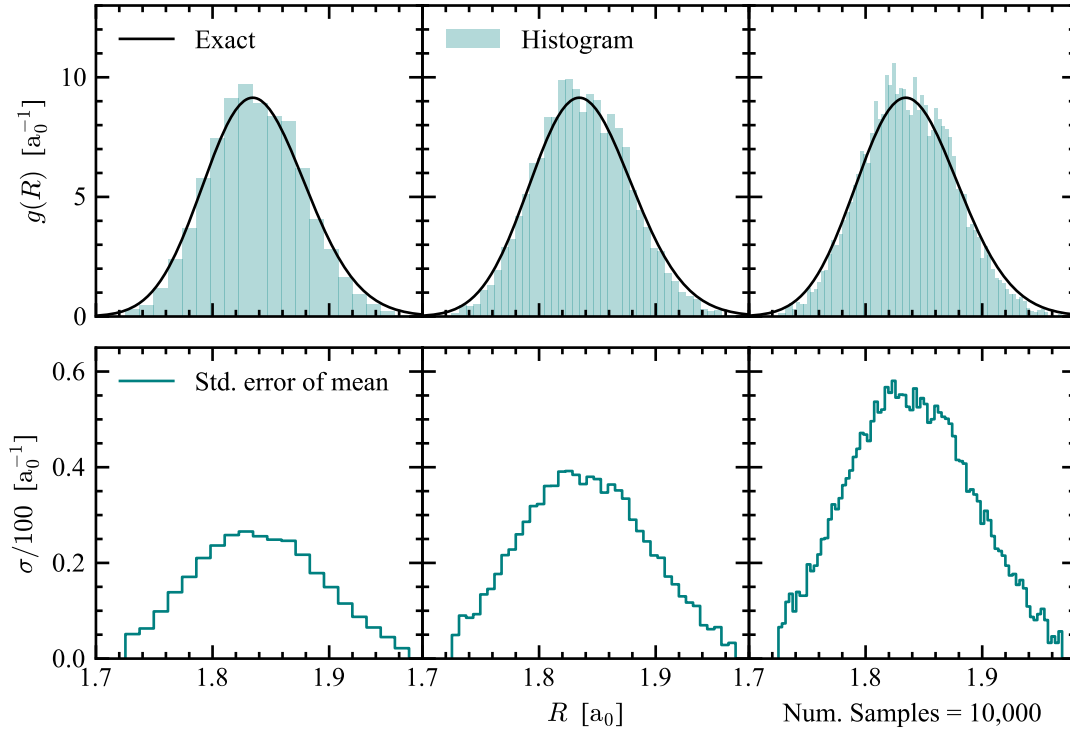
If  $P(\mathbf{x})$  is a normalised probability distribution, the distribution function for a general variable  $\xi(\mathbf{x}) \in [\Xi_{\min}, \Xi_{\max}]$  is

$$g(\Xi) = \langle \delta(\xi(\mathbf{x}) - \Xi) \rangle, \quad (5.1)$$

with

$$\langle \dots \rangle = \int d\mathbf{x} P(\mathbf{x}) (\dots). \quad (5.2)$$

The naive way to calculate  $g(\Xi)$  is to discretise the  $\Xi$  coordinate into a set of bins. A histogram is then obtained by repeatedly sampling from  $P(\mathbf{x})$ , and counting the number of times  $\xi(\mathbf{x})$  is within each of the bins. The heights of the normalised histogram correspond to  $g(\Xi)$ , evaluated at the centre of each bin.



**Figure 5.1:** Classical radial distribution function for the 3D version of the model from chapter 2, calculated using histograms of varying bin width.

**top** Distribution function calculated analytically and with histograms.

**bottom** Standard error of the mean for the histogram calculations.

The problem with this approach is that the resolution of the resulting distribution is tied to the width of the histogram bins. Narrower bins give a higher resolution, but contain a smaller fraction of the samples, thus requiring more total samples to converge the distribution. More quantitatively, the variance of the distribution diverges as the bin width decreases.<sup>103</sup>

A better approach comes from transforming to a new coordinate system where one of the variables is  $\xi$ ,

$$g(\Xi) = \int d\mathbf{x}' \int d\xi \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, x'_{N-1}, \xi)} \right\| P(\mathbf{x}) \delta(\xi - \Xi). \quad (5.3)$$

We now differentiate with respect to  $\Xi$  and integrate by parts with respect to  $\xi$ ,

$$\frac{dg}{d\Xi} = \int d\mathbf{x}' \int d\xi \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) \frac{\partial}{\partial \Xi} \delta(\xi - \Xi), \quad (5.4a)$$

$$= - \int d\mathbf{x}' \int d\xi \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) \frac{\partial}{\partial \xi} \delta(\xi - \Xi), \quad (5.4b)$$

$$= \int d\mathbf{x}' \int d\xi \delta(\xi - \Xi) \frac{\partial}{\partial \xi} \left( \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) \right), \quad (5.4c)$$

$$= \int d\mathbf{x}' \int d\xi \delta(\xi - \Xi) \times \left( \frac{\partial}{\partial \xi} \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) + \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| \frac{\partial P}{\partial \xi} \right). \quad (5.4d)$$

We now return to the original coordinate system

$$\begin{aligned} \frac{dg}{d\Xi} &= \int d\mathbf{x} \delta(\xi(\mathbf{x}) - \Xi) \\ &\quad \times \left( \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\|^{-1} \frac{\partial}{\partial \xi} \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) + \frac{\partial P}{\partial \xi} \right), \end{aligned} \quad (5.5a)$$

$$= \int d\mathbf{x} \delta(\xi(\mathbf{x}) - \Xi) \left( \frac{\partial}{\partial \xi} \ln \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| P(\mathbf{x}) + \frac{\partial P}{\partial \xi} \right). \quad (5.5b)$$

Finally, we can take out a factor of  $P(\mathbf{x})$

$$\frac{dg}{d\Xi} = \int d\mathbf{x} P(\mathbf{x}) \delta(\xi(\mathbf{x}) - \Xi) \left( \frac{\partial}{\partial \xi} \ln \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| + \frac{1}{P(\mathbf{x})} \frac{\partial P}{\partial \xi} \right), \quad (5.6a)$$

$$= \int d\mathbf{x} P(\mathbf{x}) \delta(\xi(\mathbf{x}) - \Xi) \left( \frac{\partial}{\partial \xi} \ln \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, \xi)} \right\| + \frac{\partial \ln P}{\partial \xi} \right). \quad (5.6b)$$

This gives an expression for the derivative as a constrained average

$$\frac{dg}{d\Xi} = \left\langle \delta(\xi(\mathbf{x}) - \Xi) W(\mathbf{x}) \right\rangle, \quad (5.7)$$

where the new estimator  $W(\mathbf{x})$  is

$$W(\mathbf{x}) = \frac{\partial}{\partial \xi} \ln \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, x'_{N-1}, \xi)} \right\| + \frac{\partial \ln P}{\partial \xi}. \quad (5.8)$$

The benefit of rewriting the derivative this way is that we can now integrate, either from  $\Xi_{\min}$  to  $\Xi$ , or from  $\Xi$  to  $\Xi_{\max}$  to give new expressions for  $g(\Xi)$

$$g(\Xi; 0) = g(\Xi_{\min}) + \left\langle h(\Xi - \xi(\mathbf{x})) W(\mathbf{x}) \right\rangle, \quad (5.9a)$$

$$g(\Xi; 1) = g(\Xi_{\max}) - \left\langle h(\xi(\mathbf{x}) - \Xi) W(\mathbf{x}) \right\rangle. \quad (5.9b)$$

The key change here is that instead of a delta function, the average now contains a Heaviside function,

$$h(\Xi - \xi) = \begin{cases} 1 & \Xi > \xi, \\ 0 & \Xi \leq \xi. \end{cases} \quad (5.10)$$

This means that an individual sample does not just contribute to where  $\Xi = \xi(\mathbf{x})$ , but to all values of  $\Xi$  greater than (or less than)  $\xi(\mathbf{x})$ , reducing the variance in the estimation of the distribution. This also removes the need for histogram bins; the distribution can be calculated with arbitrary resolution.<sup>103</sup>

A major caveat to this formulation is that samples also contribute to where  $\Xi$  is beyond the region of interest, when the distribution should have decayed to

zero (or some other fixed value). As a result, these regions of the distributions can display spuriously high variance and non-zero mean values. On the other hand, the opposite side of the distribution contains very few samples and accurately predicts the limiting value. The direction of this effect in Eq. (5.9a) is opposite to that of Eq. (5.9b) as the direction of the Heaviside function is reversed. It follows that we can combine the two equations to get the best of both,<sup>104</sup>

$$g(\Xi; \lambda) = (1 - \lambda(\Xi)) g(\Xi; 0) + \lambda(\Xi) g(\Xi; 1). \quad (5.11)$$

The variance in  $g(\Xi; \lambda)$  is simply a quadratic function of  $\lambda(\Xi)$

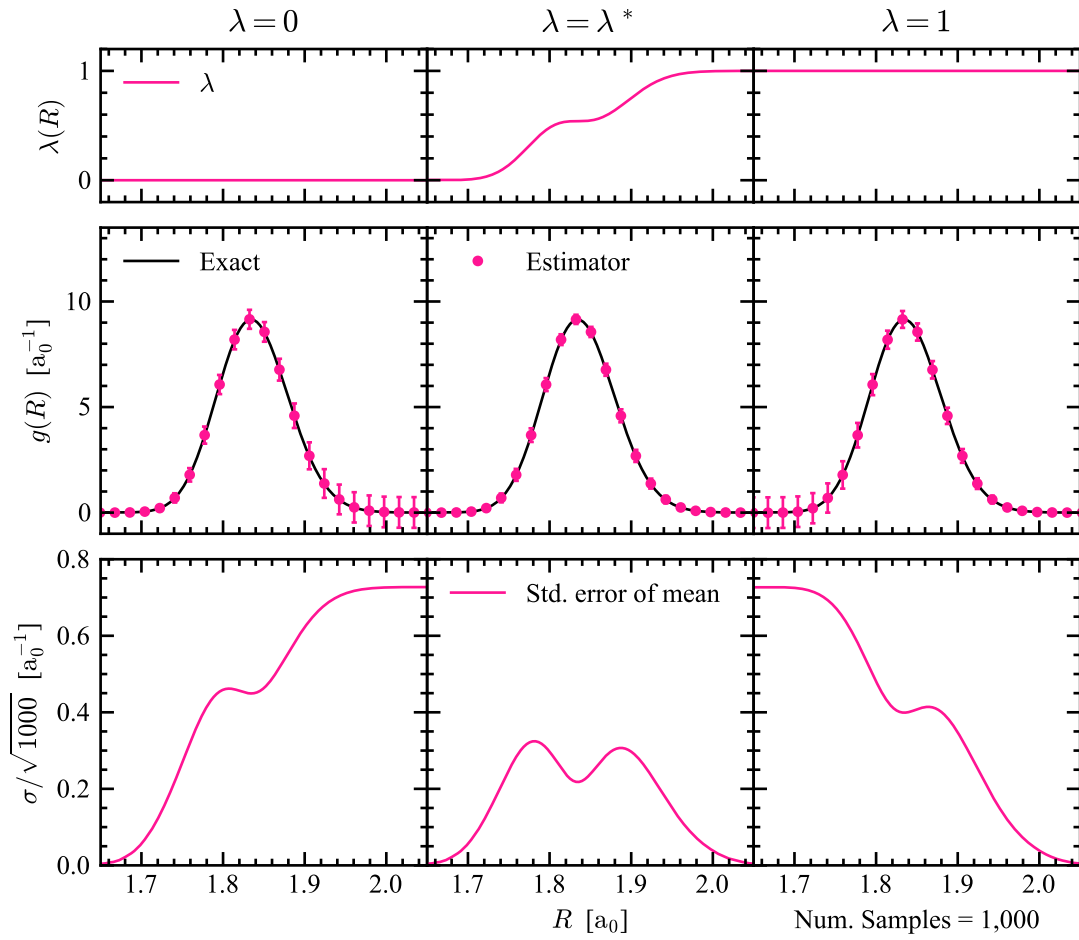
$$\sigma^2 = \left\langle \left( h(\Xi - \xi)W - \langle h(\Xi - \xi)W \rangle - \lambda(\Xi)(W - \langle W \rangle) \right)^2 \right\rangle, \quad (5.12)$$

which is minimised when  $\lambda$  is equal to

$$\lambda^*(\Xi) = \frac{\left\langle (h(\Xi - \xi)W - \langle h(\Xi - \xi)W \rangle)(W - \langle W \rangle) \right\rangle}{\left\langle (W - \langle W \rangle)^2 \right\rangle}. \quad (5.13)$$

Thus Eq. (5.11) with Eq. (5.13) gives optimal variance across the entire distribution.<sup>104</sup>

As a simple demonstration, histograms and reduced variance estimators were applied to the 3D version of the model problem from chapter 2, shown in figs. 5.1 and 5.2. In this case, reduced variance estimators give a standard error of the mean that is comparable to histograms, with only a tenth of the samples and at arbitrary resolution.



**Figure 5.2:** Classical radial distribution function for the 3D version of the model from chapter 2, calculated using Eq. (5.11) with different  $\lambda$ .

**top**  $\lambda(R)$  for each of the columns.

**middle** Distribution function calculated analytically and with Eq. (5.11). Error bars correspond to the bottom panels.

**bottom** Standard error of the mean for the reduced variance estimator calculations, if only 1,000 samples were used.

## 5.2 CLASSICAL ESTIMATORS

Now that the general framework of reduced variance estimators is established, we turn to the specific case of averages in the classical NVT ensemble, defined by

$$\langle \dots \rangle = \frac{1}{Z_1} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p}, \mathbf{q})} (\dots). \quad (5.14)$$

In the language of section 5.1, the normalised probability distribution  $P(\mathbf{q})$  is

$$P(\mathbf{q}) = \frac{1}{Z_1} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} e^{-\beta H(\mathbf{p}, \mathbf{q})}, \quad (5.15)$$

where, as defined in chapter 1,  $H(\mathbf{p}, \mathbf{q})$  is the classical Hamiltonian and  $Z_1$  is

$$Z_1 = \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p}, \mathbf{q})}.$$

A general classical distribution function is defined as

$$g(\Xi) = \left\langle \delta(\xi(\mathbf{q}) - \Xi) \right\rangle. \quad (5.16)$$

Repeating the steps of section 5.1 gives the corresponding  $W(\mathbf{q})$  estimator as

$$W(\mathbf{q}) = \frac{\partial \ln J}{\partial \xi} + \beta \frac{\partial \mathbf{q}}{\partial \xi} \cdot \mathbf{f}, \quad (5.17)$$

where  $J(\mathbf{q})$  is the Jacobian for the  $\mathbf{q} \rightarrow (\mathbf{q}', \xi)$  transformation

$$J(\mathbf{q}) = \left\| \frac{\partial(q_1, \dots, q_N)}{\partial(q'_1, \dots, q'_{N-1}, \xi)} \right\|, \quad (5.18)$$

and  $\mathbf{f}$  is the classical force on the system,

$$f_i = -\frac{\partial V}{\partial q_i}. \quad (5.19)$$

A key aspect of Eq. (5.17) is that the Hamiltonian enters only through the forces,  $\mathbf{f}$ . These are already calculated as part of a molecular dynamics simulation, minimising the computational cost of calculating  $W(\mathbf{q})$ . Another consequence is that Eq. (5.17) is remarkably general — the only extra information required is the transformation  $\mathbf{q} \rightarrow (\mathbf{q}', \xi)$ .

### 5.3 PATH INTEGRAL ESTIMATORS

Expanding on the classical distributions of section 5.2, reduced variance estimators are also applicable to path integral averages,

$$\langle \dots \rangle = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})} (\dots). \quad (5.20)$$

In this case,  $P(\underline{\mathbf{q}})$  is defined as

$$P(\underline{\mathbf{q}}) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})}, \quad (5.21)$$

where, as defined in chapter 1,  $H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})$  is the Hamiltonian for an  $n$ -bead ring polymer,  $\beta_n = \beta/n$ , and  $Z_n$  is

$$Z_n = \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\underline{\mathbf{p}}, \underline{\mathbf{q}})}.$$

Both derivations in this section also make use of the following identity for a derivative with respect to a mean

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x^{(j)}, \quad \frac{\partial}{\partial \bar{x}} = \sum_{j=1}^n \frac{\partial}{\partial x^{(j)}}. \quad (5.22)$$

### 5.3.1 Quasi-Centroid Distribution Functions

The distribution function for the quasi-centroid of a variable is

$$g(\Xi) = \left\langle \delta(\bar{\xi} - \Xi) \right\rangle, \quad (5.23)$$

where  $\bar{\xi}$  is the mean value of  $\xi$  around the ring polymer,

$$\bar{\xi}(\underline{\mathbf{q}}) = \frac{1}{n} \sum_{j=1}^n \xi^{(j)}, \quad \xi^{(j)} = \xi(\mathbf{q}^{(j)}).$$

As before, the aim is to transform into a coordinate system where  $\bar{\xi}$  is one of the variables. To this end, we first transform the coordinates of each of the ring polymer replicas,

$$\int d^n \mathbf{q} (\dots) = \int d^n \mathbf{q}' \int d^n \xi \left[ \prod_{j=1}^n J^{(j)} \right] (\dots). \quad (5.24a)$$

We now apply the ring polymer normal mode transform to  $\{\xi^{(j)}\}$ , and note that  $\tilde{\xi}_0 = \sqrt{n} \bar{\xi}$  to give

$$\int d^n \mathbf{q} (\dots) = \int d^n \mathbf{q}' \int d^n \tilde{\xi} \left[ \prod_{j=1}^n J^{(j)} \right] (\dots), \quad (5.24b)$$

$$= \int d^n \mathbf{q}' \int d^{n-1} \tilde{\xi} \int d\bar{\xi} \sqrt{n} \left[ \prod_{j=1}^n J^{(j)} \right] (\dots). \quad (5.24c)$$

Now  $\bar{\xi}$  is one of the integration variables, with the associated Jacobian

$$\left\| \frac{\partial \underline{\mathbf{q}}}{\partial (\underline{\mathbf{q}}', \dots, \bar{\xi})} \right\| = \sqrt{n} \prod_{j=1}^n J^{(j)}. \quad (5.25)$$

Repeating the steps of section 5.1 thus gives  $W(\underline{\mathbf{q}})$  as

$$\begin{aligned} W(\underline{\mathbf{q}}) &= \frac{\partial}{\partial \bar{\xi}} \ln \left[ \sqrt{n} \prod_{j=1}^n J^{(j)} \right] - \beta_n \frac{\partial H_n}{\partial \bar{\xi}}, \\ &= \sum_{j=1}^n \left[ \frac{\partial \ln J^{(j)}}{\partial \xi^{(j)}} - \beta_n \frac{\partial H_n}{\partial \xi^{(j)}} \right]. \end{aligned} \quad (5.26)$$

Finally, the derivative of  $H_n$  can be expanded to give  $W(\underline{\mathbf{q}})$  in more physical terms

$$W(\underline{\mathbf{q}}) = \sum_{j=1}^n \left[ \frac{\partial \ln J^{(j)}}{\partial \xi^{(j)}} + \beta_n \frac{\partial \mathbf{q}^{(j)}}{\partial \xi^{(j)}} \cdot \mathbf{f}^{(j)} \right], \quad (5.27)$$

where  $\mathbf{f}^{(j)}$  is the total force acting on the  $j^{\text{th}}$  ring polymer replica

$$f_i^{(j)} = -\frac{\partial V^{(j)}}{\partial q_i^{(j)}} - m_i \omega_n^2 \left( 2q_i^{(j)} - q_i^{(j-1)} - q_i^{(j+1)} \right). \quad (5.28)$$

This estimator shares many similarities with the classical estimator, Eq. (5.17). In fact, it is just the sum of Eq. (5.17) evaluated at each of the ring polymer replicas, except with the effective temperature  $\beta_n$ , and including forces from the ring polymer springs.

### 5.3.2 Centroid Distribution Functions

A general centroid distribution function is given by

$$g(\Xi) = \left\langle \delta(\xi(\bar{\mathbf{q}}) - \Xi) \right\rangle, \quad (5.29)$$

where the centroid is

$$\bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}^{(j)}.$$

To make  $\xi(\bar{\mathbf{q}})$  one of the integration variables, we first use the normal mode

transformation, and note that  $\tilde{\mathbf{q}}_0 = \sqrt{n} \bar{\mathbf{q}}$  to give

$$\int d^n \mathbf{q} (\dots) = \int d^n \tilde{\mathbf{q}} (\dots), \quad (5.30a)$$

$$= \int d^{n-1} \tilde{\mathbf{q}} \int d\bar{\mathbf{q}} n^{N/2} (\dots). \quad (5.30b)$$

The previous transformation can then be applied to  $\bar{\mathbf{q}}$  to get

$$\int d^n \mathbf{q} (\dots) = \int d^{n-1} \tilde{\mathbf{q}} \int d\bar{\mathbf{q}}' \int d\xi(\bar{\mathbf{q}}) n^{N/2} J(\bar{\mathbf{q}}) (\dots). \quad (5.30c)$$

Now  $\xi(\bar{\mathbf{q}})$  is one of the integration variables, with the associated Jacobian

$$\left\| \frac{\partial \mathbf{q}}{\partial(\underline{\mathbf{q}}', \dots, \xi(\bar{\mathbf{q}}))} \right\| = n^{N/2} J(\bar{\mathbf{q}}). \quad (5.31)$$

Repeating the steps of section 5.1 then gives the estimator  $W(\underline{\mathbf{q}})$  as

$$\begin{aligned} W(\underline{\mathbf{q}}) &= \frac{\partial}{\partial \xi(\bar{\mathbf{q}})} \ln [n^{N/2} J(\bar{\mathbf{q}})] - \beta_n \frac{\partial H_n}{\partial \xi(\bar{\mathbf{q}})}, \\ &= \frac{\partial \ln J(\bar{\mathbf{q}})}{\partial \xi(\bar{\mathbf{q}})} - \beta_n \frac{\partial H_n}{\partial \xi(\bar{\mathbf{q}})}. \end{aligned} \quad (5.32)$$

Finally, expanding the derivative of  $H_n$  gives

$$W(\underline{\mathbf{q}}) = \frac{\partial \ln J(\bar{\mathbf{q}})}{\partial \xi(\bar{\mathbf{q}})} + \beta \frac{\partial \bar{\mathbf{q}}}{\partial \xi(\bar{\mathbf{q}})} \cdot \bar{\mathbf{f}}, \quad (5.33)$$

where  $\bar{\mathbf{f}}$  is the mean force on each of the replicas

$$\bar{f}_i = \frac{1}{n} \sum_{j=1}^n f_i^{(j)} = -\frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial q_i^{(j)}}. \quad (5.34)$$

Once again, the centroid estimator is very similar to the classical estimator in Eq. (5.17), except that it is evaluated at the centroid and uses the mean force of the ring polymer.

## 5.4 DISTRIBUTION FUNCTIONS FOR MOLECULES

In this section, reduced variance estimators are derived for the most common distribution functions relating to the geometry of molecules. As a demonstration, these are used to produce various distribution functions of  $\text{H}_2\text{O}_2$ . Details of this system are given in section 4.6. All results were generated using the PILE integration scheme<sup>14</sup> in its original OBABO ordering, with a time step of  $\Delta t = 1$  atomic time unit.

Notation used in this section	
$\mathbf{q}_i$	Position vector of atom $i$
$\mathbf{f}_i$	Force acting on atom $i$
$\mathbf{r}_{ij}$	$\mathbf{q}_i - \mathbf{q}_j$

The more general framework that was developed in sections 5.2 and 5.3 demonstrates the similarity between centroid and quasi-centroid estimators and their classical counterparts. Consequently, this section only explicitly deals with classical estimators, as it is straightforward to deduce what the path integral extensions should be. By considering Eq. (5.17) and the analogous path integral estimators, this simply amounts to providing expressions for  $(\partial \ln J / \partial \xi)$  and  $(\partial \mathbf{q} / \partial \xi) \cdot \mathbf{f}$ .

As explained previously, the main aim is to provide a transformation from the standard Cartesian position vectors to a coordinate system where  $\xi$  is one of the variables. To this end, a key step used throughout this section is to express an interatomic vector in spherical coordinates,

$$\mathbf{r}_{ij} = r \cos \phi \sin \theta \hat{\mathbf{u}} + r \sin \phi \sin \theta \hat{\mathbf{v}} + r \cos \theta \hat{\mathbf{w}},$$

where  $r \in [0, \infty]$ ,  $\phi \in [0, 2\pi]$ ,  $\theta \in [0, \pi]$  and  $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$  are orthonormal basis vectors. This allows for one of the atoms to be redefined in spherical coordinates relative to

the other one, e.g.:

$$\mathbf{q}_1(\mathbf{q}_2, r, \phi, \theta) \equiv \mathbf{q}_2 + \mathbf{r}_{12}.$$

The Jacobian for this change of variables,  $(\mathbf{q}_1, \mathbf{q}_2, \dots) \rightarrow (r, \phi, \theta, \mathbf{q}_2, \dots)$ , is

$$\left\| \frac{\partial(\mathbf{q}_1, \mathbf{q}_2, \dots)}{\partial(r, \phi, \theta, \mathbf{q}_2, \dots)} \right\| = r^2 \sin \theta. \quad (5.35)$$

This result is true regardless of  $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$ , including when the basis vectors depend on the positions of the other atoms in the system. This result is derived fully in section C.1.

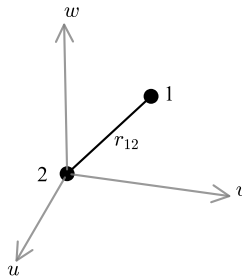
### 5.4.1 Bond Length Distributions

Let  $\xi(\mathbf{q}) \equiv r_{12}$  be the bond length between atoms 1 and 2. The change of variables is achieved by redefining  $\mathbf{q}_1$  relative to  $\mathbf{q}_2$ ,

$$\mathbf{q}_1 = \mathbf{q}_2 + \mathbf{r}_{12}, \quad (5.36a)$$

$$\mathbf{r}_{12} = r \cos \phi \sin \theta \hat{\mathbf{u}} + r \sin \phi \sin \theta \hat{\mathbf{v}} + r \cos \theta \hat{\mathbf{w}}. \quad (5.36b)$$

The choice of orthonormal basis vectors is completely arbitrary. Clearly,  $r$  corresponds to  $r_{12}$ .



Changing variables from  $(\mathbf{q}_1, \mathbf{q}_2, \dots) \rightarrow (r_{12}, \phi, \theta, \mathbf{q}_2, \dots)$  and using Eq. (5.35) gives

$$\frac{\partial \ln J}{\partial r_{12}} = \frac{2}{r_{12}}. \quad (5.37)$$

Only  $\mathbf{q}_1$  explicitly depends on  $r_{12}$ , thus giving

$$\frac{\partial \mathbf{q}}{\partial r_{12}} \cdot \mathbf{f} = \frac{\partial \mathbf{q}_1}{\partial r_{12}} \cdot \mathbf{f}_1, \quad (5.38a)$$

$$= \frac{\mathbf{r}_{12}}{r_{12}} \cdot \mathbf{f}_1. \quad (5.38b)$$

The choice to redefine  $\mathbf{q}_1$  was arbitrary, and instead we could have redefined  $\mathbf{q}_2$  relative to  $\mathbf{q}_1$ , which would simply swap the 1 and 2 indices,

$$\frac{\partial \mathbf{q}}{\partial r_{12}} \cdot \mathbf{f} = \frac{\mathbf{r}_{21}}{r_{12}} \cdot \mathbf{f}_2. \quad (5.38c)$$

Noting that  $\mathbf{r}_{12} = -\mathbf{r}_{21}$ , taking the mean of the two expressions gives the final result as

$$\frac{\partial \mathbf{q}}{\partial r_{12}} \cdot \mathbf{f} = \frac{1}{2} \frac{\mathbf{r}_{12}}{r_{12}} \cdot (\mathbf{f}_1 - \mathbf{f}_2). \quad (5.38d)$$

This result has been validated numerically for the classical, centroid and quasi-centroid cases, and fig. 5.3 shows that reduced variance estimators give the correct OH bond length distribution for  $\text{H}_2\text{O}_2$ .

A final point of consideration is that, because the Jacobian term depends only on  $r_{12}$ , it is possible to take this term out of the estimator

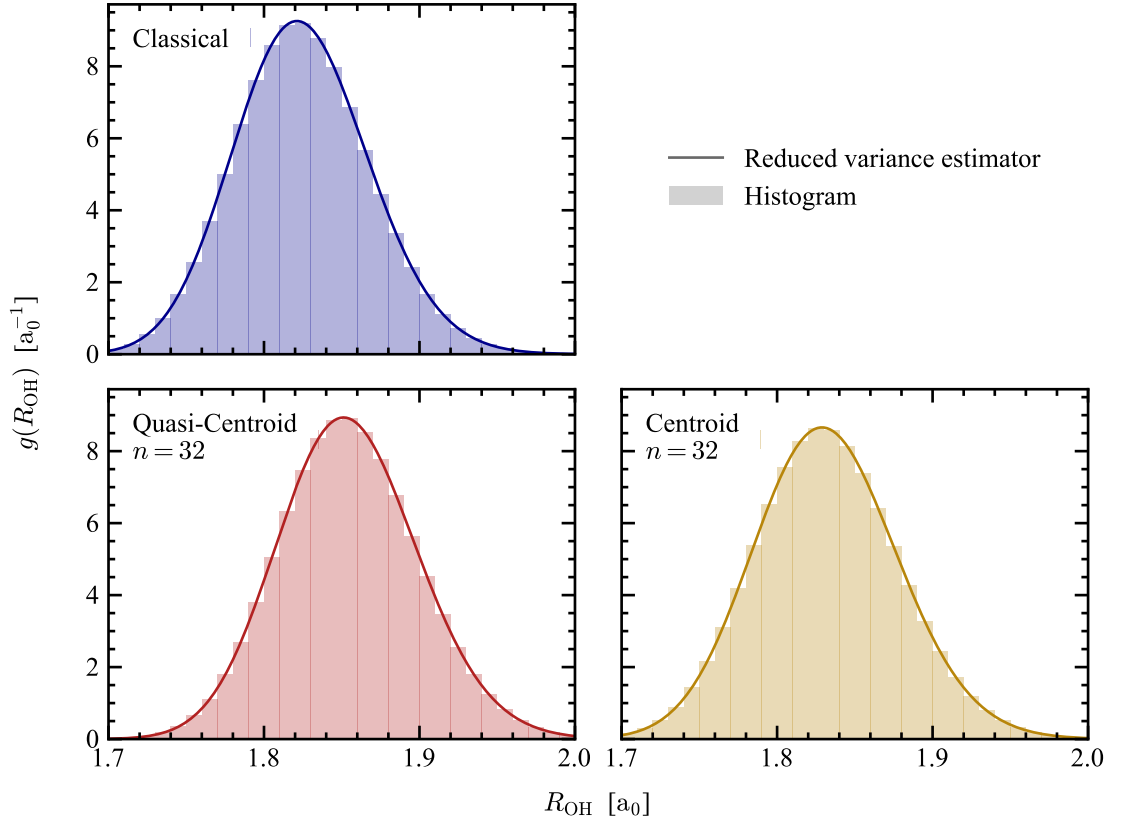
$$\begin{aligned} \frac{dg}{dR} &= \left\langle \delta(r_{12} - R) \left( \frac{2}{r_{12}} + \frac{\beta}{2} \frac{\mathbf{r}_{12}}{r_{12}} \cdot (\mathbf{f}_1 - \mathbf{f}_2) \right) \right\rangle, \\ &= \frac{2}{R} g(R) + \frac{\beta}{2} \left\langle \delta(r_{12} - R) \frac{\mathbf{r}_{12}}{r_{12}} \cdot (\mathbf{f}_1 - \mathbf{f}_2) \right\rangle. \end{aligned} \quad (5.39)$$

Rearranging and dividing by  $R^2$  then gives a new reduced variance estimator for the slightly different function  $g(R)/R^2$ , which is typically used instead of  $g(R)$  when

studying the condensed phase,<sup>103</sup>

$$\begin{aligned} \frac{1}{R^2} \frac{dg}{dR} - \frac{2}{R^3} g(R) &= \frac{\beta}{2R^2} \left\langle \delta(r_{12} - R) \frac{\mathbf{r}_{12}}{r_{12}} \cdot (\mathbf{f}_1 - \mathbf{f}_2) \right\rangle, \\ \frac{d}{dR} \left( \frac{g(R)}{R^2} \right) &= \frac{\beta}{2} \left\langle \delta(r_{12} - R) \frac{\mathbf{r}_{12}}{r_{12}^3} \cdot (\mathbf{f}_1 - \mathbf{f}_2) \right\rangle. \end{aligned} \quad (5.40)$$

It should be noted that, due to the different constraint on  $R$ , this reformulation is possible for the classical and centroid distributions, but *not* the equivalent quasi-centroid distribution.



**Figure 5.3:** Fully converged OH bond length distributions for  $\text{H}_2\text{O}_2$  at 300 K. Wide histogram bins are used to achieve complete convergence.

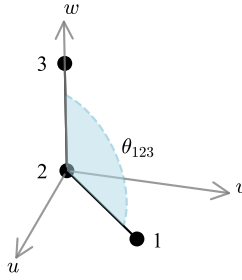
### 5.4.2 Bond Angle Distribution

Let  $\xi(\mathbf{q}) = \theta_{123}$  be the angle formed by the atoms 1, 2 and 3. The change of variables is achieved by redefining  $\mathbf{q}_1$  relative to the central atom  $\mathbf{q}_2$ , using  $\mathbf{r}_{32}$  as the  $\hat{\mathbf{w}}$  axis,

$$\mathbf{q}_1 = \mathbf{q}_2 + \mathbf{r}_{12}, \quad (5.41a)$$

$$\mathbf{r}_{12} = r \cos \phi \sin \theta \hat{\mathbf{u}} + r \sin \phi \sin \theta \hat{\mathbf{v}} + r \cos \theta \frac{\mathbf{r}_{32}}{r_{32}}. \quad (5.41b)$$

The remaining vectors  $\hat{\mathbf{u}}$  and  $\hat{\mathbf{v}}$  are unimportant. With this definition,  $r = r_{12}$ , and the polar angle  $\theta$  corresponds to the bond angle between the three atoms,  $\theta_{123}$ .



Changing variables from  $(\mathbf{q}_1, \mathbf{q}_2, \dots) \rightarrow (r, \phi, \theta_{123}, \mathbf{q}_2, \dots)$  and using Eq. (5.35) gives

$$\frac{\partial \ln J}{\partial \theta_{123}} = \cot \theta_{123}. \quad (5.42)$$

Only  $\mathbf{q}_1$  explicitly depends on  $\theta_{123}$ , thus giving

$$\frac{\partial \mathbf{q}}{\partial \theta_{123}} \cdot \mathbf{f} = \frac{\partial \mathbf{q}_1}{\partial \theta_{123}} \cdot \mathbf{f}_1, \quad (5.43a)$$

$$= \frac{r_{12}}{\sin \theta_{123}} \left( \cos \theta_{123} \frac{\mathbf{r}_{12}}{r_{12}} - \frac{\mathbf{r}_{32}}{r_{32}} \right) \cdot \mathbf{f}_1. \quad (5.43b)$$

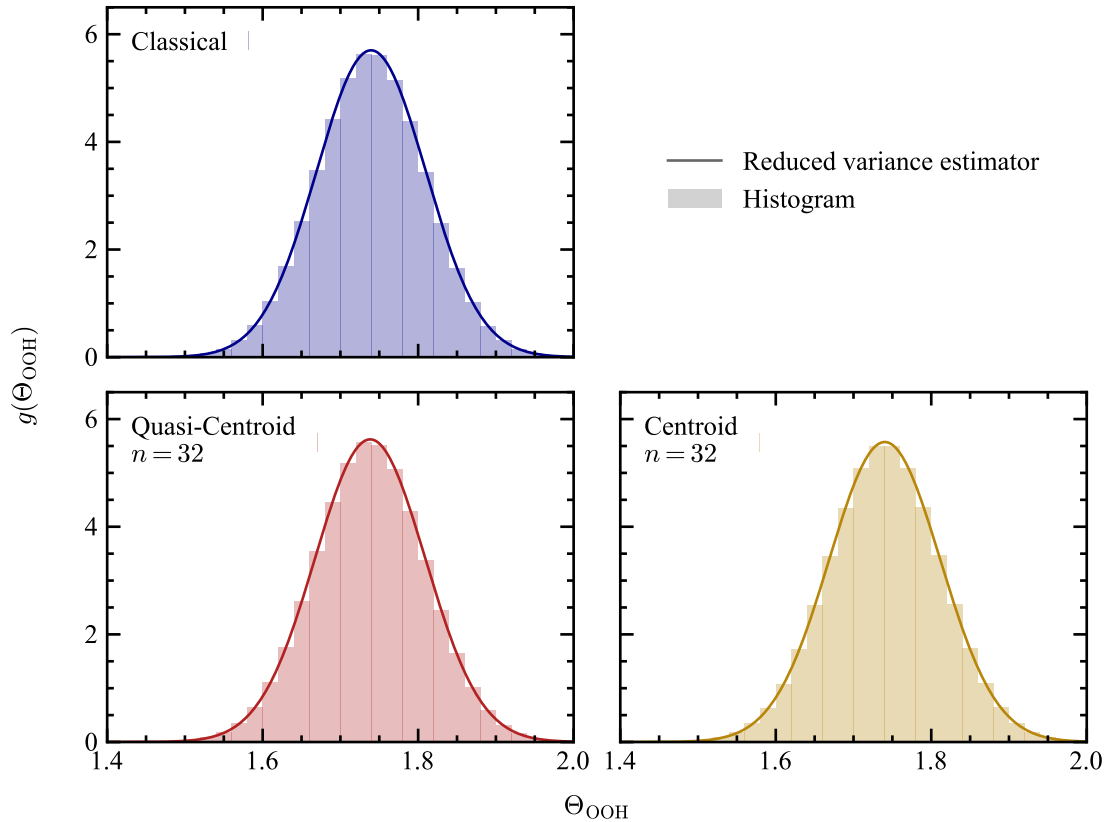
The choice to redefine  $\mathbf{q}_1$  was arbitrary, and we could have instead chosen to redefine  $\mathbf{q}_3$  relative to  $\mathbf{q}_2$ . This would have the effect of swapping the 1 and 3 indices,

$$\frac{\partial \mathbf{q}}{\partial \theta_{123}} \cdot \mathbf{f} = \frac{r_{32}}{\sin \theta_{123}} \left( \cos \theta_{123} \frac{\mathbf{r}_{32}}{r_{32}} - \frac{\mathbf{r}_{12}}{r_{12}} \right) \cdot \mathbf{f}_3. \quad (5.43c)$$

Taking the average of the two gives the final result as

$$\frac{\partial \mathbf{q}}{\partial \theta_{123}} \cdot \mathbf{f} = \frac{1}{2 \sin \theta_{123}} \left[ r_{12} \left( \cos \theta_{123} \frac{\mathbf{r}_{12}}{r_{12}} - \frac{\mathbf{r}_{32}}{r_{32}} \right) \cdot \mathbf{f}_1 + r_{32} \left( \cos \theta_{123} \frac{\mathbf{r}_{32}}{r_{32}} - \frac{\mathbf{r}_{12}}{r_{12}} \right) \cdot \mathbf{f}_3 \right]. \quad (5.43d)$$

This result has been validated numerically for the classical, centroid and quasi-centroid cases, and fig. 5.4 shows that reduced variance estimators give the correct OOH bond angle distribution for  $\text{H}_2\text{O}_2$ .



**Figure 5.4:** Fully converged OOH bond angle distributions for  $\text{H}_2\text{O}_2$  at 300 K. Wide histogram bins are used to achieve complete convergence.

A final point of consideration is that, because the Jacobian term depends only on  $\theta_{123}$ , it is possible to take this term out of the estimator,

$$\begin{aligned} \frac{dg}{d\Theta} &= \left\langle \delta(\theta_{123} - \Theta) \left( \cot \theta_{123} + \frac{\beta}{2} \left( \frac{\partial \mathbf{q}_1}{\partial \theta_{123}} \cdot \mathbf{f}_1 + \frac{\partial \mathbf{q}_3}{\partial \theta_{123}} \cdot \mathbf{f}_3 \right) \right) \right\rangle, \\ &= \cot \Theta g(\Theta) + \frac{\beta}{2} \left\langle \delta(\theta_{123} - \Theta) \left( \frac{\partial \mathbf{q}_1}{\partial \theta_{123}} \cdot \mathbf{f}_1 + \frac{\partial \mathbf{q}_3}{\partial \theta_{123}} \cdot \mathbf{f}_3 \right) \right\rangle. \end{aligned} \quad (5.44)$$

Rearranging and dividing by  $\sin \Theta$  then gives a new reduced variance estimator for the slightly different function  $g(\Theta)/\sin \Theta$ :

$$\begin{aligned} \frac{1}{\sin \Theta} \frac{dg}{d\Theta} - \frac{\cot \Theta}{\sin \Theta} g(\Theta) &= \frac{\beta}{2 \sin \Theta} \left\langle \delta(\theta_{123} - \Theta) \left( \frac{\partial \mathbf{q}_1}{\partial \theta_{123}} \cdot \mathbf{f}_1 + \frac{\partial \mathbf{q}_3}{\partial \theta_{123}} \cdot \mathbf{f}_3 \right) \right\rangle, \\ \frac{d}{d\Theta} \left( \frac{g(\Theta)}{\sin \Theta} \right) &= \frac{\beta}{2} \left\langle \frac{\delta(\theta_{123} - \Theta)}{\sin \theta_{123}} \left( \frac{\partial \mathbf{q}_1}{\partial \theta_{123}} \cdot \mathbf{f}_1 + \frac{\partial \mathbf{q}_3}{\partial \theta_{123}} \cdot \mathbf{f}_3 \right) \right\rangle. \end{aligned} \quad (5.45)$$

It should be noted that, due to the different constraint on  $\Theta$ , this reformulation is possible for the classical and centroid distributions, but *not* the equivalent quasi-centroid distribution.

### 5.4.3 Torsion Angle Distribution

Let  $\xi(\mathbf{q}) = \phi_{1234}$  be the torsion angle of atoms 1-4, which is the angle between the planes containing atoms 1,2,3 and 2,3,4. The change of variables is achieved by redefining  $\mathbf{q}_1$  relative to  $\mathbf{q}_2$

$$\mathbf{q}_1 = \mathbf{q}_2 + \mathbf{r}_{12}, \quad (5.46a)$$

$$\mathbf{r}_{12} = r \cos \phi \sin \theta \hat{\mathbf{u}} + r \sin \phi \sin \theta \hat{\mathbf{v}} + r \cos \theta \hat{\mathbf{w}}, \quad (5.46b)$$

where the basis vectors are defined by the three other atoms. The  $w$ -axis is defined to be parallel to  $\mathbf{r}_{23}$ ,

$$\hat{\mathbf{w}} = \frac{\mathbf{r}_{23}}{r_{23}}. \quad (5.46c)$$

The  $u$ -axis is defined such that  $\mathbf{r}_{43}$  lies in the  $uw$  plane,

$$\mathbf{r}_{43} = (\mathbf{r}_{43} \cdot \hat{\mathbf{u}})\hat{\mathbf{u}} + (\mathbf{r}_{43} \cdot \hat{\mathbf{w}})\hat{\mathbf{w}},$$

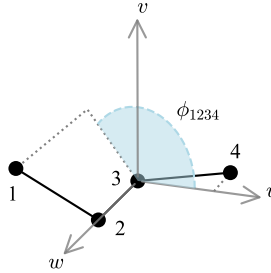
which rearranges to give

$$\hat{\mathbf{u}} = \frac{\mathbf{r}_{43} - (\mathbf{r}_{43} \cdot \hat{\mathbf{w}})\hat{\mathbf{w}}}{|\mathbf{r}_{43} - (\mathbf{r}_{43} \cdot \hat{\mathbf{w}})\hat{\mathbf{w}}|}. \quad (5.46d)$$

The remaining vector is defined to maintain orthonormality,

$$\hat{\mathbf{v}} = \hat{\mathbf{w}} \times \hat{\mathbf{u}}. \quad (5.46e)$$

By construction, atoms 2,3,4 all lie in the  $uw$  plane, meaning  $\phi$  corresponds to the torsion angle. This definition also matches the IUPAC convention for the direction of measuring torsion angles.<sup>98</sup>



Changing variables from  $(\mathbf{q}_1, \mathbf{q}_2, \dots) \rightarrow (r, \phi_{1234}, \theta, \mathbf{q}_2, \dots)$ , and using Eq. (5.35) gives

$$\frac{\partial \ln J}{\partial \phi_{1234}} = 0. \quad (5.47)$$

Only  $\mathbf{q}_1$  explicitly depends on  $\phi_{1234}$ , thus giving

$$\frac{\partial \mathbf{q}}{\partial \phi_{1234}} \cdot \mathbf{f} = \frac{\partial \mathbf{q}_1}{\partial \phi_{1234}} \cdot \mathbf{f}_1, \quad (5.48a)$$

$$= \frac{(\mathbf{r}_{23} \times \mathbf{r}_{12})}{r_{23}} \cdot \mathbf{f}_1. \quad (5.48b)$$

Torsion angles are identical if measured in reverse, i.e.:  $\phi_{1234} = \phi_{4321}$ . This is equivalent to if we had redefined  $\mathbf{q}_4$  instead of  $\mathbf{q}_1$ , which would give

$$\frac{\partial \mathbf{q}}{\partial \phi_{1234}} \cdot \mathbf{f} = \frac{(\mathbf{r}_{32} \times \mathbf{r}_{43})}{r_{32}} \cdot \mathbf{f}_4. \quad (5.48c)$$

Taking the average of the two expressions, and using  $\mathbf{r}_{ij} = -\mathbf{r}_{ji}$  gives the final result as

$$\frac{\partial \mathbf{q}}{\partial \phi_{1234}} \cdot \mathbf{f} = \frac{1}{2r_{23}} \left( (\mathbf{r}_{23} \times \mathbf{r}_{12}) \cdot \mathbf{f}_1 + (\mathbf{r}_{23} \times \mathbf{r}_{34}) \cdot \mathbf{f}_4 \right). \quad (5.48d)$$

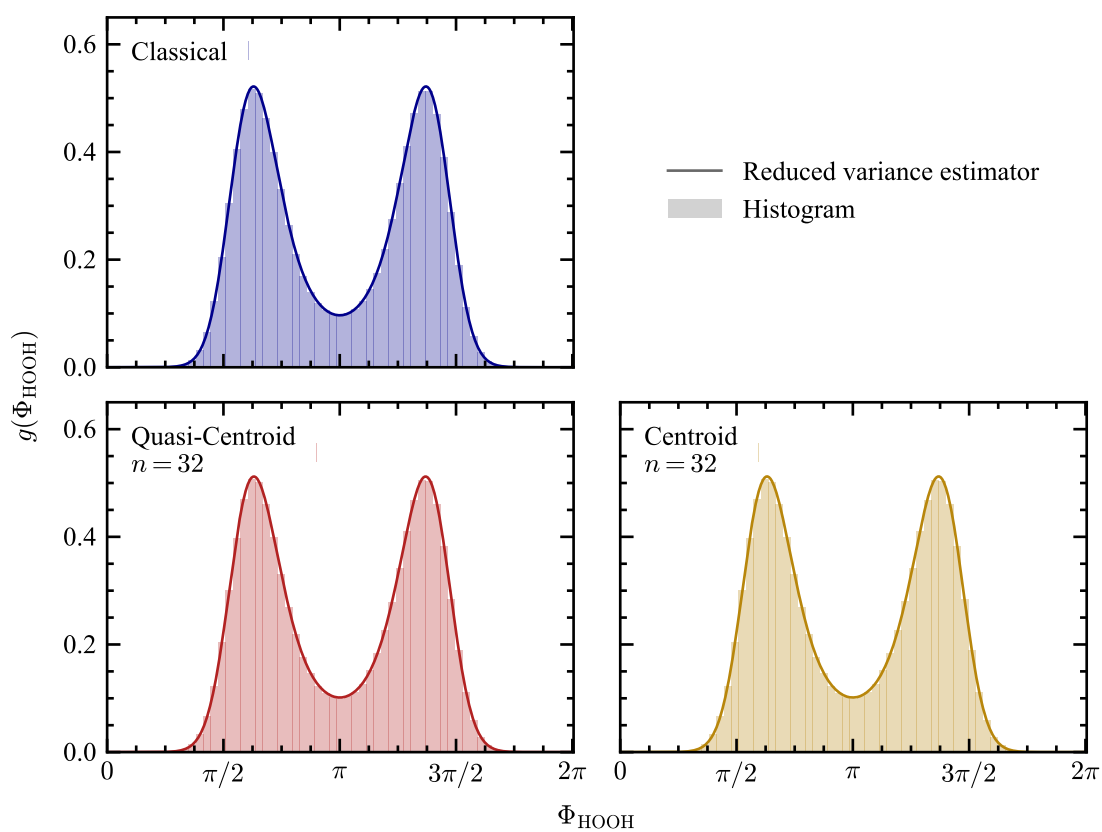
This result has been validated numerically for the classical, centroid and quasi-centroid cases, and fig. 5.5 shows that reduced variance estimators give the correct torsion angle distribution for  $\text{H}_2\text{O}_2$ .

In this case, the constrained rotation about the O–O bond avoids the synperiplanar ( $\Phi = 0, 2\pi$ ) conformer entirely, which gives predictable limits of  $g(0) = g(2\pi) = 0$  to use with Eq. (5.9). This will not be the case generally, as free rotation in both directions could result in a distribution that never reaches zero. However this issue is easily avoided because the torsion angle is periodic, meaning even if the limiting values are non-zero, they must be the same,  $g(0) = g(2\pi) = g_{\text{ref}}$ . For torsion angles, Eq. (5.11) now reads as

$$g(\Phi; \lambda) = g_{\text{ref}} + \left\langle h(\Phi - \phi(\mathbf{q}))W(\mathbf{q}) \right\rangle - \lambda(\Phi) \left\langle W(\mathbf{q}) \right\rangle. \quad (5.49a)$$

Because Eq. (5.13) is completely independent of any reference values of  $g(\Phi)$ , one can first calculate  $g(\Phi; \lambda^*) - g_{\text{ref}}$  directly. As  $g(\Phi)$  should be normalised,  $g_{\text{ref}}$  is simply the scalar value that must be added to give a normalised distribution

$$g_{\text{ref}} = \frac{1}{2\pi} - \frac{1}{2\pi} \int_0^{2\pi} d\Phi \left( \left\langle h(\Phi - \phi(\mathbf{q}))W(\mathbf{q}) \right\rangle - \lambda(\Phi) \left\langle W(\mathbf{q}) \right\rangle \right). \quad (5.49b)$$



**Figure 5.5:** Fully converged torsion angle distributions for  $\text{H}_2\text{O}_2$  at 300 K. Wide histogram bins are used to achieve complete convergence.

## 5.5 SUMMARY OF EQUATIONS

### General Theory

Let  $P(\mathbf{x})$  be a normalised probability distribution. The distribution function for a general variable  $\xi(\mathbf{x}) \in [\Xi_{\min}, \Xi_{\max}]$  is

$$g(\Xi) = \left\langle \delta(\xi(\mathbf{x}) - \Xi) \right\rangle = \int d\mathbf{x} P(\mathbf{x}) \delta(\xi(\mathbf{x}) - \Xi).$$

Reduced variance estimator for  $g(\Xi)$ :

$$\begin{aligned} g(\Xi; 0) &= g(\Xi_{\min}) + \left\langle h(\Xi - \xi(\mathbf{x})) W(\mathbf{x}) \right\rangle, \\ g(\Xi; 1) &= g(\Xi_{\max}) - \left\langle h(\xi(\mathbf{x}) - \Xi) W(\mathbf{x}) \right\rangle, \end{aligned}$$

with

$$W(\mathbf{x}) = \frac{\partial}{\partial \xi} \ln \left\| \frac{\partial(x_1, \dots, x_N)}{\partial(x'_1, \dots, x'_{N-1}, \xi)} \right\| + \frac{\partial \ln P}{\partial \xi}.$$

The best estimation of the entire distribution is

$$g(\Xi) = (1 - \lambda^*(\Xi)) g(\Xi; 0) + \lambda^*(\Xi) g(\Xi; 1),$$

where  $\lambda^*(\Xi)$  is given by

$$\lambda^*(\Xi) = \frac{\left\langle (h(\Xi - \xi)W - \langle h(\Xi - \xi)W \rangle)(W - \langle W \rangle) \right\rangle}{\left\langle (W - \langle W \rangle)^2 \right\rangle}.$$

## Classical Distribution Functions

$$g(\Xi) = \frac{1}{Z_1} \frac{1}{(2\pi\hbar)^N} \int d\mathbf{p} \int d\mathbf{q} e^{-\beta H(\mathbf{p}, \mathbf{q})} \delta(\xi(\mathbf{q}) - \Xi)$$

$$W(\mathbf{q}) = \frac{\partial \ln J}{\partial \xi} + \beta \frac{\partial \mathbf{q}}{\partial \xi} \cdot \mathbf{f}$$

$$J(\mathbf{q}) = \left\| \frac{\partial(q_1, \dots, q_N)}{\partial(q'_1, \dots, q'_{N-1}, \xi)} \right\|$$

$$f_i = -\frac{\partial V}{\partial q_i}$$

## Quasi-Centroid Distribution Functions

$$g(\Xi) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\bar{\xi} - \Xi)$$

$$\bar{\xi} = \frac{1}{n} \sum_{j=1}^n \xi^{(j)} \quad \xi^{(j)} = \xi(\mathbf{q}^{(j)})$$

$$W(\mathbf{q}) = \sum_{j=1}^n \left[ \frac{\partial \ln J(\mathbf{q}^{(j)})}{\partial \xi^{(j)}} + \beta_n \frac{\partial \mathbf{q}^{(j)}}{\partial \xi^{(j)}} \cdot \mathbf{f}^{(j)} \right]$$

$$f_i^{(j)} = -\frac{\partial V^{(j)}}{\partial q_i^{(j)}} - m_i \omega_n^2 (2q_i^{(j)} - q_i^{(j-1)} - q_i^{(j+1)})$$

## Centroid Distribution Functions

$$g(\Xi) = \frac{1}{Z_n} \frac{1}{(2\pi\hbar)^{Nn}} \int d^n \mathbf{p} \int d^n \mathbf{q} e^{-\beta_n H_n(\mathbf{p}, \mathbf{q})} \delta(\xi(\bar{\mathbf{q}}) - \Xi)$$

$$\bar{\mathbf{q}} = \frac{1}{n} \sum_{j=1}^n \mathbf{q}^{(j)}$$

$$W(\mathbf{q}) = \frac{\partial \ln J(\bar{\mathbf{q}})}{\partial \xi(\bar{\mathbf{q}})} + \beta \frac{\partial \bar{\mathbf{q}}}{\partial \xi(\bar{\mathbf{q}})} \cdot \bar{\mathbf{f}}$$

$$\bar{f}_i = \frac{1}{n} \sum_{j=1}^n f_i^{(j)} = -\frac{1}{n} \sum_{j=1}^n \frac{\partial V^{(j)}}{\partial q_i^{(j)}}$$

## Bond Length Distribution

$$\frac{\partial \ln J}{\partial r_{12}} = \frac{2}{r_{12}}$$

$$\frac{\partial \mathbf{q}}{\partial r_{12}} \cdot \mathbf{f} = \frac{1}{2} \frac{\mathbf{r}_{12}}{r_{12}} \cdot (\mathbf{f}_1 - \mathbf{f}_2)$$

## Bond Angle Distribution

$$\frac{\partial \ln J}{\partial \theta_{123}} = \cot \theta_{123}$$

$$\frac{\partial \mathbf{q}}{\partial \theta_{123}} \cdot \mathbf{f} = \frac{1}{2 \sin \theta_{123}} \left[ r_{12} \left( \cos \theta_{123} \frac{\mathbf{r}_{12}}{r_{12}} - \frac{\mathbf{r}_{32}}{r_{32}} \right) \cdot \mathbf{f}_1 + r_{32} \left( \cos \theta_{123} \frac{\mathbf{r}_{32}}{r_{32}} - \frac{\mathbf{r}_{12}}{r_{12}} \right) \cdot \mathbf{f}_3 \right]$$

## Torsion Angle Distribution

$$\frac{\partial \ln J}{\partial \phi_{1234}} = 0$$

$$\frac{\partial \mathbf{q}}{\partial \phi_{1234}} \cdot \mathbf{f} = \frac{1}{2r_{23}} \left( (\mathbf{r}_{23} \times \mathbf{r}_{12}) \cdot \mathbf{f}_1 + (\mathbf{r}_{23} \times \mathbf{r}_{34}) \cdot \mathbf{f}_4 \right)$$

## CONCLUSIONS

The vibrational dynamics of molecules is significantly and unavoidably influenced by nuclear quantum effects. In other fields of chemistry, imaginary time path integrals have been an efficient tool to tackle this issue. However, these established methods have not fared so well in this particular application.

One of the most promising attempts to address this issue has been quasi-centroid molecular dynamics (QCMD),<sup>26</sup> which is essentially a curvilinear variation on the popular centroid molecular dynamics (CMD) method.<sup>25,39</sup> While QCMD has shown great potential, it still faces several challenges. In fact, QCMD has only been applied successfully to a handful of examples and has limitations that have so far prevented its use for molecules larger than  $\text{NH}_3$ .

As it turns out, CMD is actually the specific case of QCMD with Cartesian coordinates. It also happens to be that this coordinate choice is the most convenient; the forces acting on the quasi-centroid are usually affected by the path integral springs, but these terms cancel out in Cartesian coordinates. Compared to CMD, the quasi-centroid forces are inherently harder to calculate, and the various CMD algorithms built up over the last three decades are not always easily transferred. The outcome is a slow QCMD implementation that requires approximations to the forces on the quasi-centroid. Also, this all assumes that a quasi-centroid can even be found, which might not always be possible.

Despite these concerns, QCMD is still a promising idea because when it does work, it performs consistently better than any other path integral method.<sup>9,24,26</sup> Therefore, this thesis has presented a framework that corrects these issues, and makes the quasi-centroid approach more practical.

Chapter 2 demonstrated that QCMD is not just an *ad hoc* fix to CMD; the underlying theory is in fact a more general version of the well-established path integral centroid formalism.<sup>25</sup> This leads to several interesting observations. Firstly, the quasi-centroid formalism is theoretically equivalent to quantum mechanics, for static equilibrium properties of distinguishable particles. It is also actually an example of coarse-graining, because the  $N$ -dimensional effective quasi-centroid potential is defined such that it reproduces the statistics of an  $(N \times n)$ -dimensional path integral. By exploiting this connection, the ideas presented here show what happens when established techniques from the coarse-graining literature are applied to the quasi-centroid problem.

Fast quasi-centroid molecular dynamics (f-QCMD) is successful because it produces the effective potential energy surface directly. This avoids all the problems associated with calculating the quasi-centroid forces from a path integral simulation, because they never are calculated. There is not even any need to locate a quasi-centroid. Instead, f-QCMD just aims to reproduce the path integral statistics that are most important to vibrational dynamics; as much or as little as that might be.

In the examples from chapter 4, QCMD and f-QCMD produced practically identical spectra, which suggests that both methods make valid approximations in order to calculate the quasi-centroid forces. For f-QCMD, this is particularly impressive because it means a complete potential energy surface is being reproduced with only one-dimensional corrections. The reason this works is that the effective potential is built on top of the classical surface, which is already the natural high-temperature limit of the system, and should already capture most of the correlation. Any corrections are just small perturbations that can be uncorrelated. The use of corrections based on valence coordinates (bond lengths, bond angles) also aids this process, as they are generally uncoupled anyway.<sup>57–59</sup>

It is worth mentioning that, in principle, f-QCMD has the option to use higher-order corrections, but based on the results presented, it seems that such an approach

is unnecessary. This same line of reasoning applies to iterative Boltzmann inversion<sup>80,81</sup> (IBI), which was successful because the starting classical system is already close to the desired outcome. Again, more sophisticated alternatives<sup>84,86,87</sup> could have been used instead, but IBI was completely sufficient and there was no problem here that needed more than 4 iterations to converge. Beyond this thesis, research in the group has run into bigger challenges for IBI when trying to apply f-QCMD to liquid water. However it can still converge rapidly with some minor improvements to the algorithm.<sup>105</sup>

In the gas phase, QCMD has only so far been applied to  $\text{H}_2\text{O}$  and  $\text{NH}_3$ , but the work here has gone well beyond this. The next logical step,  $\text{CH}_4$ , presented a significant challenge for QCMD because the number of bond lengths and angles ( $4+6 = 10$ ) is greater than the number of vibrational degrees of freedom ( $3 \times 5 - 6 = 9$ ), which means the quasi-centroid is overdetermined.<sup>8</sup>  $\text{H}_2\text{O}_2$  allowed an investigation into how well torsional motion is described. HDO demonstrated that modelling deuteration can be as simple as changing the mass of one of the atoms. Across all molecules and temperatures, f-QCMD gave consistently accurate estimates of fundamental band frequencies and band shapes, making it the best method considered in this study.

Frequency prediction was noticeably worse for combination bands, and even more so for overtones. The reason is that f-QCMD is ultimately just classical dynamics, but with the fundamental frequencies of the potential slightly modified. Consequently, combination/overtone frequencies are merely the sum of the fundamental frequencies, which means the fundamental errors are compounded. Overtones are even less accurate because this also ignores anharmonic effects that cause non-uniform spacing between quantum energy levels. This is why CMD occasionally gives surprisingly accurate overtone frequencies, when the red-shift of the curvature problem partially compensates for this anharmonicity.

Combination and overtone bands present even more problems as all the approximate methods here tended to get their intensities wrong. An interesting suggestion

for fixing this is a simple scaling factor which depends only on temperature and the fundamental frequencies, and can be applied as a final step to the spectrum.<sup>31,32</sup> Chapter 4 provided a comprehensive test of Eq. (1.54) for gas-phase molecules, and it worked well for H<sub>2</sub>O and HDO where there are clearly defined bands for each combination. The issue is that spectra are rarely this simple; there was usually considerable overlap between different bands and applying corrections properly was very difficult. In fact, correction was often impossible, as Eq. (1.54) is only applicable to binary combinations and overtones. Furthermore, IR inactive modes have no visible fundamental frequency to use as input for the correction, but they can still produce visible combination bands.\* On reflection, this type of intensity correction is an elegant idea and can sometimes be used to give impressive results. However it is just not practical enough in many circumstances, and so does not provide a reliable solution to the problem of overtone/combination intensity.

Another area where f-QCMD struggles is when the dynamics involves some kind of barrier crossing or large-amplitude motion, such as the inversion of NH<sub>3</sub> or the torsional motion of H<sub>2</sub>O<sub>2</sub>. These two examples failed for slightly different reasons, but both highlight points that are worth focusing on.

For NH<sub>3</sub>, the problem arose because the effective potential reproduced the distributions of the mean bond angles,  $\bar{\theta}_{jk}$ , but not the inversion coordinate  $\bar{\rho}$ . As  $\rho$  is a non-linear function of the bond angles, a path integral draped over the inversion barrier did not have its quasi-centroid at the transition state.<sup>24</sup> The core of the issue is that the quasi-centroid is defined with coordinates appropriate for small displacements from equilibrium, because that is exactly what most vibrations are. However, they do not always give a good description of large-amplitude motion.

The solution, it seems, would be to explicitly account for the large-amplitude motion as well, but it is still unclear how well small- and large-amplitude vibrations can be described simultaneously. For example, the problem here occurs because the

---

\*For f-QCMD, combination frequencies are just the sum of the fundamental frequencies, so it might be possible to deduce missing fundamental frequency with some extra work.

inversion coordinate,  $\rho$ , and the bond angles,  $\theta_{jk}$ , are related non-linearly, and the mean of one does not correspond to the mean of the other.

The study of  $\text{H}_2\text{O}_2$  did explicitly account for the large-amplitude torsional motion, and as a result had a fairly reasonable effective potential. The issue here was that the barrier is low, which leads to significant tunnelling splitting between the quantum torsional eigenstates that f-QCMD fails to capture.

Large-amplitude motion does seem to be a problematic area for f-QCMD, and it is definitely something that should be considered in future developments. However, relatively speaking, it does not fail that badly. At the very least, f-QCMD actually allows large-amplitude motion to occur; the adiabatic QCMD algorithm cannot handle it. For  $\text{NH}_3$ , CMD clearly gives the best description of the symmetric bend, but f-QCMD is out by less than  $100\text{ cm}^{-1}$  and performs better in all other regards. For  $\text{H}_2\text{O}_2$ , no path integral method can correctly capture the tunnelling splitting, and f-QCMD still gave the best line shape at 300 K.

Although f-QCMD itself has not worked flawlessly, the technique for calculating distribution functions outlined in chapter 5 has been an unmitigated success. This was used throughout the thesis to rapidly and precisely calculate all kinds of distribution functions, for classical and path integral systems. Thinking more generally, one possible downside is that it requires limiting values of the target distribution function, which might not be known beforehand. It also requires the forces, which might be hard to calculate for some systems, but will already be known for any kind of molecular dynamics simulation. If both these criteria are met, which they usually are, there is no obvious reason why any distribution function should not be calculated this way.

The research presented here gives a fairly thorough evaluation of f-QCMD for the vibrational spectroscopy of gas-phase molecules. Nevertheless, there is more that could be done in the gas-phase. For one, this thesis only addresses infrared spectroscopy, but with access to a polarisability surface, Raman spectroscopy is equally

possible with minimal alteration. A tougher and more interesting gas-phase study would be a fluxional system where atoms can exchange between distinct environments. Work on this has already started within the group, with ongoing research into the  $\text{CH}_5^+$  ion. Another point worth considering is that all the f-QCMD calculations here used an approximation to the dipole moment,  $\bar{\mu}_n \approx \mu$ . No real investigation was made into the validity of this approximation, and it may be worthwhile to do so.

The elevated temperature modification to CMD<sup>51</sup> has been mentioned frequently; the only reason it was not tested properly is because it was published too late in the writing process. At 300 K, CMD tended to perform quite well for the molecules considered here, which suggests that this method should also perform quite well. A more thorough investigation would definitely be worthwhile.<sup>105</sup>

While the focus here has been on a fast implementation of *quasi*-centroid molecular dynamics, the method described for generating the effective potential energy surface is completely applicable to the centroid case too. Indeed, this is exactly how all the CMD results were generated. Over the years, a substantial amount of work has gone into improving the implementation of CMD — in part, f-QCMD was necessary because these algorithms do not always translate well to the quasi-centroid. It would be interesting to see how this approach compares to the current cutting-edge of CMD.<sup>51,72</sup>

With this thesis, f-QCMD is now properly verified against gas-phase molecules where exact quantum results are obtainable. Now looking to possible future work, the ideal application and the true test of f-QCMD is in the condensed phase. This comes with its own set of challenges, for example: how exactly does the curvature problem work with two interacting molecules? However, these questions are beyond the scope of this thesis.

Applying QCMD to the condensed phase is completely possible, and it has already been used to study liquid water and ice.<sup>9,26,65</sup> Research in the group contin-

ues this trend, and has compared (adiabatic)<sup>65</sup> QCMD, f-QCMD and temperature elevated CMD<sup>51</sup> to liquid water and hexagonal ice.<sup>105</sup> As it turns out, the three algorithms do not always agree with each other. Also, the condensed phase turns out to be a harder challenge for IBI, as the input distribution functions are now multimodal. This motivates a new improvement to the standard IBI algorithm.

On a similar note, the group has also used f-QCMD to study the dynamics of liquid water in an optical cavity, which has a profound effect on the spectroscopy.<sup>106</sup> This is helping to shed some light on the topic, particularly with regards to previous studies that used the less reliable TRPMD approach.

To conclude, it should now be clear that accurately describing the dynamics of molecules is indeed a very difficult challenge. The long history of methods that have tried and failed to do so should be testament to this. This thesis does not claim to have all the answers, but it does take a big step in the right direction, towards a definitive method of simulating vibrational dynamics.



## REFERENCES

1. Zwanzig, R. *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford ; New York, 2001).
2. Habershon, S., Fanourgakis, G. S. & Manolopoulos, D. E. Comparison of path integral molecular dynamics methods for the infrared absorption spectrum of liquid water. *J. Chem. Phys.* **129**, 74501 (2008).
3. Hilborn, R. C. Einstein coefficients, cross sections, f values, dipole moments, and all that. *Am. J. Phys.* **50**, 982–986 (1982).
4. Gordon, I. E. *et al.* The HITRAN2020 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **277**, 107949 (2022).
5. Tennyson, J. & Yurchenko, S. N. The ExoMol project: Software for computing large molecular line lists. *Int. J. Quantum Chem.* **117**, 92–103 (2017).
6. Yurchenko, S. N., Al-Refaie, A. F. & Tennyson, J. EXOCROSS: a general program for generating spectra from molecular line lists. *A&A* **614** (2018).
7. Hollas, J. M. *Modern Spectroscopy* 4th. eng (J. Wiley, Chichester, 2004).
8. Fletcher, T., Zhu, A., Lawrence, J. E. & Manolopoulos, D. E. Fast quasi-centroid molecular dynamics. *J. Chem. Phys.* **155**, 231101 (2021).
9. Benson, R. L., Trenins, G. & Althorpe, S. C. Which quantum statistics–classical dynamics method is best for water? *Faraday Discuss.* **221**, 350–366 (2020).
10. Reddy, S. K., Moberg, D. R., Straight, S. C. & Paesani, F. Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations with the MB-pol potential energy function. *J. Chem. Phys.* **147**, 244504 (2017).
11. Ramírez, R., López-Ciudad, T., Kumar P, P. & Marx, D. Quantum corrections to classical time-correlation functions: Hydrogen bonding and anharmonic floppy modes. *J. Chem. Phys.* **121**, 3973–3983 (2004).
12. Hele, T. J. H. Thermal quantum time-correlation functions from classical-like dynamics. *Mol. Phys.* **115**, 1435–1462 (2017).
13. Tuckerman, M. E. *Statistical Mechanics: theory and molecular simulation* (Oxford University Press, New York, 2010).
14. Ceriotti, M., Parrinello, M., Markland, T. E. & Manolopoulos, D. E. Efficient stochastic thermostating of path integral molecular dynamics. *J. Chem. Phys.* **133** (2010).

15. Craig, I. R. & Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* **121**, 3368–3373 (2004).
16. Miller, T. F. & Manolopoulos, D. E. Quantum diffusion in liquid para-hydrogen from ring-polymer molecular dynamics. *J. Chem. Phys.* **122**, 184503 (2005).
17. Pérez de Tudela, R., Aoiz, F. J., Suleimanov, Y. V. & Manolopoulos, D. E. Chemical Reaction Rates from Ring Polymer Molecular Dynamics: Zero Point Energy Conservation in  $\text{Mu} + \text{H}_2 \rightarrow \text{MuH} + \text{H}$ . *J. Phys. Chem. Lett.* **3**, 493–497 (2012).
18. Collepardo-Guevara, R., Suleimanov, Y. V. & Manolopoulos, D. E. Bimolecular reaction rates from ring polymer molecular dynamics. *J. Chem. Phys.* **130**, 174713 (2009).
19. Suleimanov, Y. V., Collepardo-Guevara, R. & Manolopoulos, D. E. Bimolecular reaction rates from ring polymer molecular dynamics: Application to  $\text{H} + \text{CH}_4 \rightarrow \text{H}_2 + \text{CH}_3$ . *J. Chem. Phys.* **134**, 44131 (2011).
20. Lawrence, J. E. & Manolopoulos, D. E. Path integral methods for reaction rates in complex systems. *Faraday Discuss.* **221**, 9–29 (2020).
21. Witt, A. *et al.* On the applicability of centroid and ring polymer path integral molecular dynamics for vibrational spectroscopy. *J. Chem. Phys.* **130**, 194510 (2009).
22. Rossi, M., Ceriotti, M. & Manolopoulos, D. E. How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* **140**, 234116 (2014).
23. Rossi, M. *et al.* Communication: On the consistency of approximate quantum dynamics simulation methods for vibrational spectra in the condensed phase. *J. Chem. Phys.* **141**, 181101 (2014).
24. Haggard, C., Sadhasivam, V. G., Trenins, G. & Althorpe, S. C. Testing the quas centroid molecular dynamics method on gas-phase ammonia. *J. Chem. Phys.* **155**, 174120 (2021).
25. Jang, S. & Voth, G. A. Path integral centroid variables and the formulation of their exact real time dynamics. *J. Chem. Phys.* **111**, 2357–2370 (1999).
26. Trenins, G., Willatt, M. J. & Althorpe, S. C. Path-integral dynamics of water using curvilinear centroids. *J. Chem. Phys.* **151**, 54109 (2019).
27. Trenins, G. *Quas centroid Molecular Dynamics* PhD Thesis (University of Cambridge, 2020), 119.
28. Hele, T. J. H., Willatt, M. J., Muolo, A. & Althorpe, S. C. Boltzmann-conserving classical dynamics in quantum time-correlation functions: “Matsubara dynamics”. *J. Chem. Phys.* **142**, 134103 (2015).
29. Shi, Q. & Geva, E. A relationship between semiclassical and centroid correlation functions. *J. Chem. Phys.* **118**, 8173–8184 (2003).

30. Habershon, S. & Manolopoulos, D. E. Zero point energy leakage in condensed phase dynamics: An assessment of quantum simulation methods for liquid water. *J. Chem. Phys.* **131**, 244518 (2009).
31. Plé, T. *et al.* Anharmonic spectral features via trajectory-based quantum dynamics: A perturbative analysis of the interplay between dynamics and sampling. *J. Chem. Phys.* **155**, 104108 (2021).
32. Benson, R. L. & Althorpe, S. C. On the “Matsubara heating” of overtone intensities and Fermi splittings. *J. Chem. Phys.* **155**, 104107 (2021).
33. Yurchenko, S. N. *et al.* A Variationally Computed T = 300 K Line List for NH<sub>3</sub>. *J. Phys. Chem. A* **113**, 11845–11855 (2009).
34. Partridge, H. & Schwenke, D. W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *J. Chem. Phys.* **106**, 4618–4639 (1997).
35. Schwenke, D. W. & Partridge, H. Convergence testing of the analytic representation of an ab initio dipole moment function for water: Improved fitting yields improved intensities. *J. Chem. Phys.* **113**, 6592–6597 (2000).
36. Feynman, R. P., Styer, D. F. & Hibbs, A. R. *Quantum Mechanics and Path Integrals* Emended. eng (Dover, Mineola, N.Y., 2010).
37. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. I. Equilibrium properties. *J. Chem. Phys.* **100**, 5093–5105 (1994).
38. Ramírez, R. & López-Ciudad, T. The Schrödinger formulation of the Feynman path centroid density. *J. Chem. Phys.* **111**, 3339–3348 (1999).
39. Jang, S. & Voth, G. A. A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables. *J. Chem. Phys.* **111**, 2371–2384 (1999).
40. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. II. Dynamical properties. *J. Chem. Phys.* **100**, 5106–5117 (1994).
41. Sutherland, B. J., Moore, W. H. D. & Manolopoulos, D. E. Nuclear quantum effects in thermal conductivity from centroid molecular dynamics. *J. Chem. Phys.* **154**, 174104 (2021).
42. Geva, E., Shi, Q. & Voth, G. A. Quantum-mechanical reaction rate constants from centroid molecular dynamics simulations. *J. Chem. Phys.* **115**, 9209–9222 (2001).
43. Hone, T. D. & Voth, G. A. A centroid molecular dynamics study of liquid para-hydrogen and ortho-deuterium. *J. Chem. Phys.* **121**, 6412–6422 (2004).
44. Paesani, F. & Voth, G. A. A quantitative assessment of the accuracy of centroid molecular dynamics for the calculation of the infrared spectrum of liquid water. *J. Chem. Phys.* **132**, 14105 (2010).

45. Ivanov, S. D., Witt, A., Shiga, M. & Marx, D. Communications: On artificial frequency shifts in infrared spectra obtained from centroid molecular dynamics: Quantum liquid water. *J. Chem. Phys.* **132**, 31101 (2010).
46. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. III. Phase space formalism and analysis of centroid molecular dynamics. *J. Chem. Phys.* **101**, 6157–6167 (1994).
47. Reichman, D. R., Roy, P.-N., Jang, S. & Voth, G. A. A Feynman path centroid dynamics approach for the computation of time correlation functions involving nonlinear operators. *J. Chem. Phys.* **113**, 919–929 (2000).
48. Medders, G. R. & Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through “First-Principles” Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **11**, 1145–1154 (2015).
49. Ramírez, R. & López-Ciudad, T. Low lying vibrational excitation energies from equilibrium path integral simulations. *J. Chem. Phys.* **115**, 103–114 (2001).
50. Paesani, F., Xantheas, S. S. & Voth, G. A. Infrared Spectroscopy and Hydrogen-Bond Dynamics of Liquid Water from Centroid Molecular Dynamics with an Ab Initio-Based Force Field. *J. Phys. Chem. B* **113**, 13118–13130 (2009).
51. Musil, F. *et al.* Quantum dynamics using path integral coarse-graining. *J. Chem. Phys.* **157**, 181102 (2022).
52. Marx, D. & Muser, M. H. Path integral simulations of rotors: theory and applications. *J. Phys. Condens. Matter* **11**, R117 (1999).
53. Bramley, M. J., Green, W. H. & Handy, N. C. Vibration-rotation coordinates and kinetic energy operators for polyatomic molecules. *Mol. Phys.* **73**, 1183–1208 (1991).
54. Jackels, C. F., Gu, Z. & Truhlar, D. G. Reaction-path potential and vibrational frequencies in terms of curvilinear internal coordinates. *J. Chem. Phys.* **102**, 3188–3201 (1995).
55. Colbert, D. T. & Miller, W. H. A novel discrete variable representation for quantum mechanical reactive scattering via the S-matrix Kohn method. *J. Chem. Phys.* **96**, 1982–1991 (1992).
56. Wilson, E. B., Decius, J. C. & Cross, P. C. *Molecular Vibrations: the theory of infrared and Raman vibrational spectra* eng (Dover, New York, 1980).
57. Stauch, T. & Dreuw, A. On the use of different coordinate systems in mechanochemical force analyses. *J. Chem. Phys.* **143**, 74118 (2015).
58. Bakken, V. & Helgaker, T. The efficient optimization of molecular geometries using redundant internal coordinates. *J. Chem. Phys.* **117**, 9160–9174 (2002).
59. Baker, J., Kessi, A. & Delley, B. The generation and use of delocalized internal coordinates in geometry optimization. *J. Chem. Phys.* **105**, 192–212 (1996).
60. Pulay, P. & Fogarasi, G. Geometry optimization in redundant internal coordinates. *J. Chem. Phys.* **96**, 2856–2860 (1992).

61. Von Arnim, M. & Ahlrichs, R. Geometry optimization in generalized natural internal coordinates. *J. Chem. Phys.* **111**, 9183–9190 (1999).
62. Cao, J. & Voth, G. A. The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics. *J. Chem. Phys.* **101**, 6168–6183 (1994).
63. Ka, B. J. & Voth, G. A. An Efficient and Accurate Implementation of Centroid Molecular Dynamics Using a Gaussian Approximation. *J. Phys. Chem. A* **109**, 11609–11617 (2005).
64. Parrinello, M. & Rahman, A. Study of an F center in molten KCl. *J. Chem. Phys.* **80**, 860–867 (1984).
65. Trenins, G., Haggard, C. & Althorpe, S. C. Improved torque estimator for condensed-phase quasicentroid molecular dynamics. *J. Chem. Phys.* **157**, 174108 (2022).
66. Yuan, Y., Li, J., Li, X.-Z. & Wang, F. The strengths and limitations of effective centroid force models explored by studying isotopic effects in liquid water. *J. Chem. Phys.* **148**, 184102 (2018).
67. Hone, T. D., Rossky, P. J. & Voth, G. A. A comparative study of imaginary time path integral based methods for quantum dynamics. *J. Chem. Phys.* **124**, 154103 (2006).
68. Pavese, M. & Voth, G. A. Pseudopotentials for centroid molecular dynamics: Application to self-diffusion in liquid para-hydrogen. *Chem. Phys. Lett.* **249**, 231–236 (1996).
69. Hone, T. D., Izvekov, S. & Voth, G. A. Fast centroid molecular dynamics: A force-matching approach for the predetermination of the effective centroid forces. *J. Chem. Phys.* **122**, 54105 (2005).
70. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **128**, 244114 (2008).
71. Paesani, F. *et al.* An accurate and simple quantum model for liquid water. *J. Chem. Phys.* **125**, 184507 (2006).
72. Loose, T. D., Sahrman, P. G. & Voth, G. A. Centroid Molecular Dynamics Can Be Greatly Accelerated through Neural Network Learned Centroid Forces Derived from Path Integral Molecular Dynamics. *J. Chem. Theory Comput.* **18**, 5856–5863 (2022).
73. Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **49**, 197–198 (1974).
74. Li, Y., Abberton, B. C., Kröger, M. & Liu, W. K. *Challenges in Multiscale Modeling of Polymer Dynamics* 2013.
75. Joshi, S. Y. & Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **47**, 786–803 (2021).

76. Sinitskiy, A. V. & Voth, G. A. A reductionist perspective on quantum statistical mechanics: Coarse-graining of path integrals. *J. Chem. Phys.* **143**, 94104 (2015).
77. Ryu, W. H. & Voth, G. A. Coarse-Graining of Imaginary Time Feynman Path Integrals: Inclusion of Intramolecular Interactions and Bottom-up Force-Matching. *J. Phys. Chem. A* **126**, 6004–6019 (2022).
78. Ercolessi, F. & Adams, J. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **26**, 583 (1994).
79. Izvekov, S. & Voth, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys.* **123**, 134105 (2005).
80. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **24**, 1624–1636 (2003).
81. Soper, A. K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **202**, 295–306 (1996).
82. Agrawal, V., Arya, G. & Oswald, J. Simultaneous Iterative Boltzmann Inversion for Coarse-Graining of Polyurea. *Macromolecules* **47**, 3378–3389 (2014).
83. Jain, S., Garde, S. & Kumar, S. K. Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials? *Ind. Eng. Chem. Res.* **45**, 5614–5618 (2006).
84. Lyubartsev, A. P. & Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **52**, 3730–3737 (1995).
85. Rosenberger, D., Hanke, M. & van der Vegt, N. F. A. Comparison of iterative inverse coarse-graining methods. *Eur. Phys. J. Spec. Top.* **225**, 1323–1345 (2016).
86. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **140**, 224104 (2014).
87. Heinen, M. Calculating particle pair potentials from fluid-state pair correlations: Iterative ornstein–zernike inversion. *J. Comput. Chem.* **39**, 1531–1543 (2018).
88. Zhu, A. *Quantum effects in vibrational spectroscopy from quasi-centroid molecular dynamics* Part II Thesis (University of Oxford, 2021).
89. Yurchenko, S. N. *et al.* Potential energy surface of HDO up to 25000cm<sup>-1</sup>. *J. Chem. Phys.* **128**, 44312 (2008).
90. Voronin, B. A. *et al.* A high accuracy computed line list for the HDO molecule. *Mon. Not. R. Astron. Soc.* **402**, 492–496 (2010).
91. Lodi, L. *et al.* A new ab initio ground-state dipole moment surface for the water molecule. *J. Chem. Phys.* **128**, 44304 (2008).
92. Nguyen, H. V. L., Gulaczyk, I., Kr glewski, M. & Kleiner, I. Large amplitude inversion tunneling motion in ammonia, methylamine, hydrazine, and

- secondary amines: From structure determination to coordination chemistry. *Coord. Chem. Rev.* **436**, 213797 (2021).
93. Yurchenko, S. N., Tennyson, J., Barber, R. J. & Thiel, W. Vibrational transition moments of CH<sub>4</sub> from first principles. *J. Mol. Spectrosc.* **291**, 69–76 (2013).
94. Yurchenko, S. N. & Tennyson, J. ExoMol line lists – IV. The rotation–vibration spectrum of methane up to 1500 K. *Mon. Not. R. Astron. Soc.* **440**, 1649–1661 (2014).
95. Blunt, V., Cedeno, D. & Manzanares, C. Vibrational overtone spectroscopy of methane in liquid argon solutions. *Mol. Phys.* **91**, 3–18 (1997).
96. Małyszczek, P. & Koput, J. Accurate ab initio potential energy surface and vibration-rotation energy levels of hydrogen peroxide. *J. Comput. Chem.* **34**, 337–345 (2013).
97. Al-Refaie, A. F. *et al.* A variationally calculated room temperature line-list for H<sub>2</sub>O<sub>2</sub>. *J. Mol. Spectrosc.* **318**, 84–90 (2015).
98. Moss, G. P. Basic terminology of stereochemistry (IUPAC Recommendations 1996). *Pure and Applied Chemistry* **68**, 2193–2222 (1996).
99. Hunt, R. H., Leacock, R. A., Peters, C. W. & Hecht, K. T. Internal-Rotation in Hydrogen Peroxide: The Far-Infrared Spectrum and the Determination of the Hindering Potential. *J. Chem. Phys.* **42**, 1931–1946 (1965).
100. Flaud, J.-M., Camy-Peyret, C., Johns, J. W. C. & Carli, B. The far infrared spectrum of H<sub>2</sub>O<sub>2</sub>. First observation of the staggering of the levels and determination of the cis barrier. *J. Chem. Phys.* **91**, 1504–1510 (1989).
101. Giguère, P. A. & Srinivasan, T. K. K. A Raman study of H<sub>2</sub>O<sub>2</sub> and D<sub>2</sub>O<sub>2</sub> vapor. *J. Raman Spectrosc.* **2**, 125–132 (1974).
102. Borgis, D., Assaraf, R., Rotenberg, B. & Vuilleumier, R. Computation of pair distribution functions and three-dimensional densities with a reduced variance principle. *Mol. Phys.* **111**, 3486–3492 (2013).
103. Rotenberg, B. Use the force! Reduced variance estimators for densities, radial distribution functions, and local mobilities in molecular simulations. *J. Chem. Phys.* **153**, 150902 (2020).
104. Coles, S. W., Mangaud, E., Frenkel, D. & Rotenberg, B. Reduced variance analysis of molecular dynamics simulations by linear combination of estimators. *J. Chem. Phys.* **154**, 191101 (2021).
105. Lawrence, J. E., Lieberherr, A. Z., Fletcher, T. & Manolopoulos, D. E. Fast quasi-centroid molecular dynamics for water and ice. *Under review at J. Phys. Chem. B* (2023).
106. Lieberherr, A. Z., Furniss, S. T. E., Lawrence, J. E. & Manolopoulos, D. E. Vibrational strong coupling in liquid water from cavity molecular dynamics. *J. Chem. Phys.* **158**, 234106 (2023).



## APPENDICES



## APPENDIX A

### INTENSITY CORRECTION FACTOR

For this system, the Hamiltonian is

$$\hat{H} = \hat{H}_0 + \epsilon \hat{V}, \quad (\text{A.1a})$$

$$\hat{H}_0 = \sum_i^N \left( \frac{\hat{p}_i^2}{2m_i} + \frac{1}{2} m_i \Omega_i^2 \hat{q}_i^2 \right), \quad (\text{A.1b})$$

$$\hat{V} = \frac{1}{6} \sum_{ijk}^N \eta_{ijk} \hat{q}_i \hat{q}_j \hat{q}_k, \quad (\text{A.1c})$$

where the coefficients  $\eta_{ijk}$  are invariant under permutation of indices. The dipole moment operator is

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0 + \sum_i^N \boldsymbol{\mu}_i \hat{q}_i, \quad (\text{A.2})$$

where  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\mu}_i$  are real (as is required for  $\hat{\boldsymbol{\mu}}$  to be Hermitian).

The eigenstates of  $\hat{H}$  are expanded using standard perturbation theory,

$$|\mathbf{n}\rangle = |\mathbf{n}^{(0)}\rangle + \epsilon |\mathbf{n}^{(1)}\rangle + \mathcal{O}(\epsilon^2).$$

The unperturbed energies of the system are the eigenvalues of  $\hat{H}_0$ ,

$$E_{\mathbf{n}}^{(0)} = \langle \mathbf{n}^{(0)} | \hat{H}_0 | \mathbf{n}^{(0)} \rangle = \hbar \sum_i^N \Omega_i \left( n_i + \frac{1}{2} \right), \quad (\text{A.3a})$$

giving

$$\omega_{\mathbf{n}\mathbf{n}'}^{(0)} = \sum_i^N \Omega_i (n_i - n'_i). \quad (\text{A.3b})$$

### A.1 DIPOLE MOMENT MATRIX ELEMENTS

This section derives the matrix elements of the dipole moment operator,

$$\langle \mathbf{n}' | \hat{\boldsymbol{\mu}} | \mathbf{n} \rangle = \langle \mathbf{n}'^{(0)} | \hat{\boldsymbol{\mu}} | \mathbf{n}^{(0)} \rangle + \epsilon \langle \mathbf{n}'^{(0)} | \hat{\boldsymbol{\mu}} | \mathbf{n}^{(1)} \rangle + \epsilon \langle \mathbf{n}'^{(1)} | \hat{\boldsymbol{\mu}} | \mathbf{n}^{(0)} \rangle + \mathcal{O}(\epsilon^2). \quad (\text{A.4})$$

To avoid unnecessary notation, superscripts are dropped from unperturbed eigenstates for the rest of this section, i.e.:  $|\mathbf{n}\rangle \equiv |n_1^{(0)} \dots n_N^{(0)}\rangle$  and  $\omega_{\mathbf{n}\mathbf{n}'} \equiv \omega_{\mathbf{n}\mathbf{n}'}^{(0)}$ .

#### Zeroth-Order

The zeroth-order matrix element is given by

$$\langle \mathbf{n}' | \hat{\boldsymbol{\mu}} | \mathbf{n} \rangle = \boldsymbol{\mu}_0 \langle \mathbf{n}' | \mathbf{n} \rangle + \sum_i^N \boldsymbol{\mu}_i \langle \mathbf{n}' | \hat{q}_i | \mathbf{n} \rangle. \quad (\text{A.5})$$

#### First-Order

The first-order perturbative correction to  $|\mathbf{n}\rangle$  is

$$|\mathbf{n}^{(1)}\rangle = \sum_{j \neq \mathbf{n}} |\mathbf{j}\rangle \frac{\langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\hbar \omega_{\mathbf{n}j}}.$$

This gives one of the first-order matrix elements as

$$\begin{aligned} \langle \mathbf{n}' | \hat{\boldsymbol{\mu}} | \mathbf{n}^{(1)} \rangle &= \sum_{j \neq \mathbf{n}} \boldsymbol{\mu}_0 \frac{\langle \mathbf{n}' | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\hbar \omega_{\mathbf{n}j}} + \sum_{j \neq \mathbf{n}} \sum_i^N \boldsymbol{\mu}_i \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\hbar \omega_{\mathbf{n}j}}, \\ &= \frac{\boldsymbol{\mu}_0}{\hbar} \frac{\langle \mathbf{n}' | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'}} + \sum_i^N \frac{\boldsymbol{\mu}_i}{\hbar} \sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}j}}. \end{aligned} \quad (\text{A.6})$$

We now introduce the harmonic oscillator ladder operators

$$\begin{aligned}\hat{a}_i^\dagger |\mathbf{n}\rangle &= \sqrt{n_i + 1} |n_1, \dots, n_i + 1, \dots, n_N\rangle, \\ \hat{a} |\mathbf{n}\rangle &= \sqrt{n_i} |n_1, \dots, n_i - 1, \dots, n_N\rangle,\end{aligned}$$

and expand the position operator

$$\hat{q}_i = \sqrt{\frac{\hbar}{2m_i\Omega_i}} (\hat{a}_i^\dagger + \hat{a}_i),$$

to give

$$\sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{j}}} = \sqrt{\frac{\hbar}{2m_i\Omega_i}} \sum_{j \neq \mathbf{n}} \left( \frac{\langle \mathbf{n}' | \hat{a}_i^\dagger | \mathbf{j} \rangle}{\omega_{\mathbf{n}\mathbf{j}}} + \frac{\langle \mathbf{n}' | \hat{a}_i | \mathbf{j} \rangle}{\omega_{\mathbf{n}\mathbf{j}}} \right) \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle. \quad (\text{A.7})$$

The ladder operators restrict the possible values of  $|\mathbf{j}\rangle$  to  $|n'_1, \dots, n'_i \pm 1, \dots, n'_N\rangle$ , which means the frequencies  $\omega_{\mathbf{n}\mathbf{j}}$  can be deduced

$$\sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{j}}} = \sqrt{\frac{\hbar}{2m_i\Omega_i}} \sum_{j \neq \mathbf{n}} \left( \frac{\langle \mathbf{n}' | \hat{a}_i^\dagger | \mathbf{j} \rangle}{\omega_{\mathbf{n}\mathbf{n}'} + \Omega_i} + \frac{\langle \mathbf{n}' | \hat{a}_i | \mathbf{j} \rangle}{\omega_{\mathbf{n}\mathbf{n}'} - \Omega_i} \right) \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle. \quad (\text{A.8})$$

Because  $\langle \mathbf{n} | \hat{V} | \mathbf{n} \rangle = 0$ , the sum over  $\mathbf{j}$  can now include  $\mathbf{n}$ , which then gives the identity operator

$$\sum_{j \neq \mathbf{n}} |\mathbf{j}\rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle = \sum_{\mathbf{j}} |\mathbf{j}\rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle = \hat{V} | \mathbf{n} \rangle.$$

Using this result,

$$\sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{j}}} = \sqrt{\frac{\hbar}{2m_i\Omega_i}} \left( \frac{\langle \mathbf{n}' | \hat{a}_i^\dagger \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'} + \Omega_i} + \frac{\langle \mathbf{n}' | \hat{a}_i \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'} - \Omega_i} \right), \quad (\text{A.9})$$

and rearranging gives

$$\sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}j}} = \sqrt{\frac{\hbar}{2m_i\Omega_i}} \frac{\omega_{\mathbf{n}\mathbf{n}'} \langle \mathbf{n}' | (\hat{a}_i^\dagger + \hat{a}_i) \hat{V} | \mathbf{n} \rangle - \Omega_i \langle \mathbf{n}' | (\hat{a}_i^\dagger - \hat{a}_i) \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \quad (\text{A.10})$$

We now return to the position and momentum operators,

$$\hat{p}_i = i \sqrt{\frac{\hbar m_i \Omega_i}{2}} (\hat{a}_i^\dagger - \hat{a}_i),$$

to give

$$\sum_{j \neq \mathbf{n}} \frac{\langle \mathbf{n}' | \hat{q}_i | \mathbf{j} \rangle \langle \mathbf{j} | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}j}} = \frac{\omega_{\mathbf{n}\mathbf{n}'} \langle \mathbf{n}' | \hat{q}_i \hat{V} | \mathbf{n} \rangle + i \langle \mathbf{n}' | \hat{p}_i \hat{V} | \mathbf{n} \rangle / m_i}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \quad (\text{A.11})$$

Using this result, Eq. (A.6) is now

$$\langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle = \frac{\mu_0}{\hbar} \frac{\langle \mathbf{n}' | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'}} + \sum_i^N \frac{\mu_i}{\hbar} \frac{\omega_{\mathbf{n}\mathbf{n}'} \langle \mathbf{n}' | \hat{q}_i \hat{V} | \mathbf{n} \rangle + i \langle \mathbf{n}' | \hat{p}_i \hat{V} | \mathbf{n} \rangle / m_i}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \quad (\text{A.12})$$

The other first-order matrix element is easily obtained by simply swapping the indices and taking the complex conjugate

$$\begin{aligned} \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle &= \langle \mathbf{n} | \hat{\mu} | \mathbf{n}'^{(1)} \rangle^*, \\ &= \frac{\mu_0}{\hbar} \frac{\langle \mathbf{n}' | \hat{V} | \mathbf{n} \rangle}{\omega_{\mathbf{n}'\mathbf{n}}} + \sum_i^N \frac{\mu_i}{\hbar} \frac{\omega_{\mathbf{n}'\mathbf{n}} \langle \mathbf{n}' | \hat{V} \hat{q}_i | \mathbf{n} \rangle - i \langle \mathbf{n}' | \hat{V} \hat{p}_i | \mathbf{n} \rangle / m_i}{\omega_{\mathbf{n}'\mathbf{n}}^2 - \Omega_i^2}. \end{aligned} \quad (\text{A.13})$$

Noting that  $\omega_{\mathbf{n}'\mathbf{n}} = -\omega_{\mathbf{n}\mathbf{n}'}$ , the sum of the two matrix elements is therefore

$$\langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle + \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle = \sum_i^N \frac{\mu_i}{\hbar} \frac{\omega_{\mathbf{n}\mathbf{n}'} \langle \mathbf{n}' | [\hat{q}_i, \hat{V}] | \mathbf{n} \rangle + i \langle \mathbf{n}' | [\hat{p}_i, \hat{V}] | \mathbf{n} \rangle / m_i}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \quad (\text{A.14})$$

Because the perturbation  $\hat{V}$  depends only on position,  $[\hat{q}_i, \hat{V}] = 0$ . By inserting the definition of  $\hat{V}$ , the other commutator is

$$\begin{aligned} [\hat{p}_i, \hat{V}] &= \frac{1}{6} \sum_{jkl}^N \eta_{jkl} [\hat{p}_i, \hat{q}_j \hat{q}_k \hat{q}_l], \\ &= \frac{1}{6} \sum_{jkl}^N \eta_{jkl} \left( [\hat{p}_i, \hat{q}_j] \hat{q}_k \hat{q}_l + \hat{q}_j [\hat{p}_i, \hat{q}_k] \hat{q}_l + \hat{q}_j \hat{q}_k [\hat{p}_i, \hat{q}_l] \right). \end{aligned} \quad (\text{A.15})$$

Using the commutator  $[\hat{p}_i, \hat{q}_j] = -i\hbar\delta_{ij}$  gives

$$\begin{aligned} [\hat{p}_i, \hat{V}] &= -\frac{i\hbar}{6} \left( \sum_{kl}^N \eta_{ikl} \hat{q}_k \hat{q}_l + \sum_{jl}^N \eta_{jil} \hat{q}_j \hat{q}_l + \sum_{jk}^N \eta_{jki} \hat{q}_j \hat{q}_k \right), \\ &= -\frac{i\hbar}{6} \sum_{jk}^N \left( \eta_{ijk} + \eta_{jik} + \eta_{jki} \right) \hat{q}_j \hat{q}_k. \end{aligned} \quad (\text{A.16})$$

Finally, we use the fact that  $\eta_{ijk}$  is invariant to permutation of indices to give

$$[\hat{p}_i, \hat{V}] = -\frac{i\hbar}{2} \sum_{jk}^N \eta_{ijk} \hat{q}_j \hat{q}_k. \quad (\text{A.17})$$

Substituting this result back into Eq. (A.14) gives

$$\langle \mathbf{n}' | \hat{\boldsymbol{\mu}} | \mathbf{n}^{(1)} \rangle + \langle \mathbf{n}'^{(1)} | \hat{\boldsymbol{\mu}} | \mathbf{n} \rangle = \sum_{ijk}^N \frac{\boldsymbol{\mu}_i \eta_{ijk}}{2m_i} \frac{\langle \mathbf{n}' | \hat{q}_j \hat{q}_k | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \quad (\text{A.18})$$

## A.2 ALLOWED TRANSITIONS

To avoid unnecessary notation, superscripts are dropped from unperturbed eigenstates for this section, i.e.:  $|\mathbf{n}\rangle \equiv |n_1^{(0)} \dots n_N^{(0)}\rangle$  and  $\omega_{\mathbf{n}\mathbf{n}'} \equiv \omega_{\mathbf{n}\mathbf{n}'}^{(0)}$ . For convenience, we also define

$$\tilde{\mu}_i = \frac{\mu_i}{\sqrt{m_i \Omega_i}} \quad (\text{A.19})$$

and

$$\tilde{\eta}_{ijk} = \frac{\eta_{ijk}}{\sqrt{m_i m_j m_k \Omega_i \Omega_j \Omega_k}}. \quad (\text{A.20})$$

Using the results derived in Eq. (A.5) and Eq. (A.18), the total dipole moment to first-order in perturbation theory is

$$\begin{aligned} \langle \mathbf{n}' | \hat{\mu} | \mathbf{n} \rangle + \epsilon \langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle + \epsilon \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle \\ = \mu_0 \langle \mathbf{n}' | \mathbf{n} \rangle + \sum_i^N \mu_i \langle \mathbf{n}' | \hat{q}_i | \mathbf{n} \rangle + \epsilon \sum_{ijk}^N \frac{\mu_i \eta_{ijk}}{2m_i} \frac{\langle \mathbf{n}' | \hat{q}_j \hat{q}_k | \mathbf{n} \rangle}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2}. \end{aligned} \quad (\text{A.21})$$

To better rationalise this in terms of transitions, the position operators can be expanded in terms of the ladder operators,

$$\begin{aligned} \langle \mathbf{n}' | \hat{\mu} | \mathbf{n} \rangle + \epsilon \langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle + \epsilon \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle \\ = \mu_0 \langle \mathbf{n}' | \mathbf{n} \rangle + \sqrt{\frac{\hbar}{2}} \sum_i^N \tilde{\mu}_i \langle \mathbf{n}' | \hat{a}_i^\dagger + \hat{a}_i | \mathbf{n} \rangle \\ + \epsilon \frac{\hbar}{4} \sum_{ijk}^N \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2} \langle \mathbf{n}' | (\hat{a}_j^\dagger + \hat{a}_j)(\hat{a}_k^\dagger + \hat{a}_k) | \mathbf{n} \rangle. \end{aligned} \quad (\text{A.22})$$

Using properties of the ladder operators

$$\begin{aligned} \hat{a}_i^\dagger \hat{a}_i | \mathbf{n} \rangle &= n_i | \mathbf{n} \rangle, \\ \hat{a}_i \hat{a}_i^\dagger | \mathbf{n} \rangle &= (n_i + 1) | \mathbf{n} \rangle, \end{aligned}$$

this can be expanded even further

$$\begin{aligned}
& \langle \mathbf{n}' | \hat{\mu} | \mathbf{n} \rangle + \epsilon \langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle + \epsilon \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle \\
&= \left( \mu_0 + \epsilon \frac{\hbar}{4} \sum_{ij} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijj}}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2} (2n_j + 1) \right) \langle \mathbf{n}' | \mathbf{n} \rangle \\
&+ \sqrt{\frac{\hbar}{2}} \sum_i \tilde{\mu}_i \left( \langle \mathbf{n}' | \hat{a}_i^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_i | \mathbf{n} \rangle \right) \\
&+ \epsilon \frac{\hbar}{4} \sum_{ij} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijj}}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_j^\dagger \hat{a}_j^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_j | \mathbf{n} \rangle \right) \\
&+ \epsilon \frac{\hbar}{4} \sum_{ij} \sum_{k \neq j} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_j^\dagger \hat{a}_k^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_k | \mathbf{n} \rangle \right) \\
&+ \epsilon \frac{\hbar}{4} \sum_{ij} \sum_{k \neq j} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{\omega_{\mathbf{n}\mathbf{n}'}^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_k^\dagger \hat{a}_j | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_k^\dagger | \mathbf{n} \rangle \right). \quad (\text{A.23})
\end{aligned}$$

With the ladder operators, the allowed values of  $\mathbf{n}'$  can be deduced, along with the corresponding transition frequencies,  $\omega_{\mathbf{n}\mathbf{n}'}$ ,

$$\begin{aligned}
& \langle \mathbf{n}' | \hat{\mu} | \mathbf{n} \rangle + \epsilon \langle \mathbf{n}' | \hat{\mu} | \mathbf{n}^{(1)} \rangle + \epsilon \langle \mathbf{n}'^{(1)} | \hat{\mu} | \mathbf{n} \rangle \\
&= \left( \mu_0 - \epsilon \frac{\hbar}{4} \sum_{ij} \frac{\tilde{\mu}_i \tilde{\eta}_{ijj}}{\Omega_i} (2n_j + 1) \right) \langle \mathbf{n}' | \mathbf{n} \rangle \quad (\text{A.24a})
\end{aligned}$$

$$+ \sqrt{\frac{\hbar}{2}} \sum_i \tilde{\mu}_i \left( \langle \mathbf{n}' | \hat{a}_i^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_i | \mathbf{n} \rangle \right) \quad (\text{A.24b})$$

$$+ \epsilon \frac{\hbar}{4} \sum_{ij} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijj}}{4\Omega_j^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_j^\dagger \hat{a}_j^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_j | \mathbf{n} \rangle \right) \quad (\text{A.24c})$$

$$+ \epsilon \frac{\hbar}{4} \sum_{ij} \sum_{k \neq j} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{(\Omega_j + \Omega_k)^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_j^\dagger \hat{a}_k^\dagger | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_k | \mathbf{n} \rangle \right) \quad (\text{A.24d})$$

$$+ \epsilon \frac{\hbar}{4} \sum_{ij} \sum_{k \neq j} \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{(\Omega_j - \Omega_k)^2 - \Omega_i^2} \left( \langle \mathbf{n}' | \hat{a}_k^\dagger \hat{a}_j | \mathbf{n} \rangle + \langle \mathbf{n}' | \hat{a}_j \hat{a}_k^\dagger | \mathbf{n} \rangle \right). \quad (\text{A.24e})$$

This fully expanded form of the dipole moment can be easily interpreted in terms of spectral transitions:

A.24a No transition,  $\omega_{nn'} = 0$

- the diagonal elements of  $\hat{\boldsymbol{\mu}}$

A.24b Fundamental transitions,  $\omega_{nn'} = \Omega_i$

- a single excitation

A.24c First overtones,  $\omega_{nn'} = 2\Omega_j$

- two excitations occur in the same mode

A.24d Binary combinations,  $\omega_{nn'} = \Omega_j + \Omega_k$

- two excitations occur in different modes

A.24e Binary differences,  $\omega_{nn'} = \Omega_j - \Omega_k$

- one mode is excited and another is de-excited.

### A.3 INTEGRATED INTENSITIES

Using Eq. (1.6), the integrated absorption intensity at a frequency  $\omega$  is

$$\begin{aligned} I(\omega) &= \lim_{\delta\omega \rightarrow 0} \int_{\omega-\delta\omega}^{\omega+\delta\omega} d\omega' n(\omega') \alpha(\omega'), \\ &= \frac{\pi}{3c\epsilon_0 v} \frac{\omega (1 - e^{-\beta\hbar\omega})}{\hbar Z} \sum_{\substack{j,k: \\ \omega_{kj}=\omega}} e^{-\beta E_j} |\langle j | \hat{\boldsymbol{\mu}} | k \rangle|^2, \end{aligned} \quad (\text{A.25})$$

where the sum is over all  $j, k$  such that  $\omega_{kj} = \omega$ .

The energy levels of this particular system are unchanged to first-order in perturbation theory

$$E_{\mathbf{n}} = \langle \mathbf{n} | \hat{H} | \mathbf{n} \rangle = \hbar \sum_i^N \Omega_i \left( n_i + \frac{1}{2} \right) + \mathcal{O}(\epsilon^2). \quad (\text{A.26})$$

Because the levels are evenly spaced, the transition frequencies are simply integer combinations of the harmonic frequencies,

$$\omega = \sum_i^N \Delta n_i \Omega_i + \mathcal{O}(\epsilon^2). \quad (\text{A.27})$$

This also means that every state  $|\mathbf{n}\rangle$  will have a state  $|\mathbf{n} + \Delta\mathbf{n}\rangle$  corresponding to  $\omega$ . Therefore, for this system  $I(\omega)$  is

$$I(\omega) = \frac{\pi}{3c\epsilon_0 v} \frac{\omega (1 - e^{-\beta\hbar\omega})}{\hbar Z} \sum_{\mathbf{n}} e^{-\beta E_{\mathbf{n}}} |\langle \mathbf{n} | \hat{\boldsymbol{\mu}} | \mathbf{n} + \Delta\mathbf{n} \rangle|^2. \quad (\text{A.28})$$

It is worth noting that this assumes a unique  $\Delta\mathbf{n}$  for each value of  $\omega$ , which is not necessarily the case. An obvious violation of this assumption is degenerate harmonic frequencies, e.g.:  $\Omega_1 = \Omega_2$ . However, the only effect this has on  $I(\omega)$  is that the final result is summed over all the degenerate frequencies that could contribute to  $\omega$ , and the final classical intensity correction factor is entirely unaffected.

A more challenging case would be when non-degenerate harmonic frequencies combine to give degenerate combinations. For example,  $2\Omega_1 = \Omega_2 + \Omega_3$  where  $\Omega_1 \neq \Omega_2 \neq \Omega_3$ . In practice, this is never exactly true, but rotational structure and line broadening effects can cause significant overlap of the resulting spectral bands. This can make it unclear what intensity correction to apply, and as of yet there is no general prescription for what to do in these cases.

A further consequence of Eq. (A.26) is that it leads directly to several useful identities. The first is shown by differentiating the energy with respect to  $\Omega_i$

$$\frac{\partial}{\partial \Omega_i} e^{-\beta E_{\mathbf{n}}} = -\beta \hbar \left( n_i + \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2),$$

which rearranges to give

$$n_i e^{-\beta E_{\mathbf{n}}} = \left( -\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_i} - \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2). \quad (\text{A.29})$$

Using Eq. (A.26), the partition function can be expressed in closed-form,

$$Z = \sum_{\mathbf{n}} e^{-\beta E_{\mathbf{n}}} = \prod_i^N \frac{1}{2 \sinh(\beta \hbar \Omega_i / 2)} + \mathcal{O}(\epsilon^2). \quad (\text{A.30})$$

Again, we can differentiate with respect to  $\Omega_i$

$$\frac{\partial Z}{\partial \Omega_i} = -\frac{\beta \hbar}{2} \coth\left(\frac{\beta \hbar \Omega_i}{2}\right) Z + \mathcal{O}(\epsilon^2),$$

and rearrange to get the second identity

$$\begin{aligned} \left(-\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_i} \pm \frac{1}{2}\right) Z &= \left(\frac{1}{2} \coth\left(\frac{\beta \hbar \Omega_i}{2}\right) \pm \frac{1}{2}\right) Z + \mathcal{O}(\epsilon^2), \\ &= \pm \frac{1}{2} \left(\pm \coth\left(\frac{\beta \hbar \Omega_i}{2}\right) + 1\right) Z + \mathcal{O}(\epsilon^2), \\ &= \pm \frac{1}{2} \left(\coth\left(\pm \frac{\beta \hbar \Omega_i}{2}\right) + 1\right) Z + \mathcal{O}(\epsilon^2), \\ &= \pm \frac{1}{2} \left(\frac{1 + e^{\mp \beta \hbar \Omega_i}}{1 - e^{\mp \beta \hbar \Omega_i}} + 1\right) Z + \mathcal{O}(\epsilon^2), \\ &= \pm \frac{1}{1 - e^{\mp \beta \hbar \Omega_i}} Z + \mathcal{O}(\epsilon^2). \end{aligned} \quad (\text{A.31})$$

### A.3.1 Single Excitations

A fundamental band of frequency  $\omega = \Omega_j$  is caused by single excitations,

$$|\mathbf{n} + \Delta\mathbf{n}\rangle = \frac{\hat{a}_j^\dagger |\mathbf{n}\rangle}{\sqrt{n_j + 1}}. \quad (\text{A.32})$$

Using Eq. (A.24b), the corresponding transition dipole moment is

$$\langle \mathbf{n} | \hat{\boldsymbol{\mu}} | \mathbf{n} + \Delta\mathbf{n} \rangle = \tilde{\boldsymbol{\mu}}_j \sqrt{\frac{\hbar(n_j + 1)}{2}} + \mathcal{O}(\epsilon^2), \quad (\text{A.33})$$

and Eq. (A.28) is

$$I(\Omega_j) = \frac{\pi}{3c\epsilon_0 v} \frac{\Omega_j |\tilde{\boldsymbol{\mu}}_j|^2}{2} \frac{(1 - e^{-\beta\hbar\Omega_j})}{Z} \sum_{\mathbf{n}} (n_j + 1) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2). \quad (\text{A.34})$$

Using Eq. (A.29) and Eq. (A.31), the sum can now be evaluated

$$\begin{aligned} \sum_{\mathbf{n}} (n_j + 1) e^{-\beta E_{\mathbf{n}}} &= \sum_{\mathbf{n}} \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2), \\ &= \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) Z + \mathcal{O}(\epsilon^2), \\ &= \frac{Z}{1 - e^{-\beta\hbar\Omega_j}} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (\text{A.35})$$

Substituting this result back then gives the final integrated intensity for a fundamental transition

$$I(\Omega_j) = \frac{\pi}{3c\epsilon_0 v} \frac{\Omega_j |\tilde{\boldsymbol{\mu}}_j|^2}{2} + \mathcal{O}(\epsilon^2). \quad (\text{A.36})$$

### A.3.2 Double Excitations

When the total state changes by two quanta, Eq. (A.28) can be written concisely as

$$I(\Omega_j \pm \Omega_k) = \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j \pm \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{(\Omega_j \pm \Omega_k)^2 - \Omega_i^2} \right|^2 \\ \times \pm \frac{2\hbar}{2 - \delta_{jk}} \frac{(1 - e^{-\beta\hbar(\Omega_j \pm \Omega_k)})}{(1 - e^{-\beta\hbar\Omega_j})(1 - e^{\mp\beta\hbar\Omega_k})} + \mathcal{O}(\epsilon^3). \quad (\text{A.37})$$

Here,  $\omega = \Omega_j + \Omega_j$  are first overtones,  $\omega = \Omega_j + \Omega_k$  are binary combination bands, and  $\omega = \Omega_j - \Omega_k$  are binary difference bands. The individual results for each transition type are derived at the end of this section.

This can be simplified further, using the following expansion

$$1 - e^{x+y} = \frac{(1 + e^x)(1 - e^y) + (1 - e^x)(1 + e^y)}{2}.$$

It follows that

$$\frac{(1 - e^{-\beta\hbar(\Omega_j \pm \Omega_k)})}{(1 - e^{-\beta\hbar\Omega_j})(1 - e^{\mp\beta\hbar\Omega_k})} = \frac{1}{2} \left( \frac{1 + e^{-\beta\hbar\Omega_j}}{1 - e^{-\beta\hbar\Omega_j}} + \frac{1 + e^{\mp\beta\hbar\Omega_k}}{1 - e^{\mp\beta\hbar\Omega_k}} \right), \\ = \frac{1}{2} \left( \coth\left(\frac{\beta\hbar\Omega_j}{2}\right) + \coth\left(\pm\frac{\beta\hbar\Omega_k}{2}\right) \right), \\ = \frac{1}{2} \left( \coth\left(\frac{\beta\hbar\Omega_j}{2}\right) \pm \coth\left(\frac{\beta\hbar\Omega_k}{2}\right) \right), \quad (\text{A.38})$$

and

$$I(\Omega_j \pm \Omega_k) = \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j \pm \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\mu}_i \tilde{\eta}_{ijk}}{(\Omega_j \pm \Omega_k)^2 - \Omega_i^2} \right|^2 \\ \times \frac{\hbar}{2 - \delta_{jk}} \left( \coth\left(\frac{\beta\hbar\Omega_k}{2}\right) \pm \coth\left(\frac{\beta\hbar\Omega_j}{2}\right) \right) + \mathcal{O}(\epsilon^3). \quad (\text{A.39})$$

### Overtones

First overtones ( $\omega = \Omega_j + \Omega_j$ ) arise from two excitations in the same mode,

$$|\mathbf{n} + \Delta\mathbf{n}\rangle = \frac{\hat{a}_j^\dagger \hat{a}_j^\dagger}{\sqrt{(n_j + 2)(n_j + 1)}} |\mathbf{n}\rangle. \quad (\text{A.40})$$

Using Eq. (A.24c), the corresponding transition dipole moment is

$$\langle \mathbf{n} | \hat{\boldsymbol{\mu}} | \mathbf{n} + \Delta\mathbf{n} \rangle = \epsilon \frac{\hbar}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijj}}{(\Omega_j + \Omega_j)^2 - \Omega_i^2} \sqrt{(n_j + 2)(n_j + 1)} + \mathcal{O}(\epsilon^2), \quad (\text{A.41})$$

and Eq. (A.28) is

$$\begin{aligned} I(\Omega_j + \Omega_j) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j + \Omega_j) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijj}}{(\Omega_j + \Omega_j)^2 - \Omega_i^2} \right|^2 \\ &\quad \times \hbar \frac{(1 - e^{-\beta\hbar(\Omega_j + \Omega_j)})}{Z} \sum_{\mathbf{n}} (n_j + 2)(n_j + 1) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.42})$$

Using Eq. (A.29) and Eq. (A.31), the above sum can then be evaluated

$$\begin{aligned} \sum_{\mathbf{n}} (n_j + 2)(n_j + 1) e^{-\beta E_{\mathbf{n}}} &= \sum_{\mathbf{n}} \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{3}{2} \right) \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2), \\ &= \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{3}{2} \right) \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) Z + \mathcal{O}(\epsilon^2), \\ &= \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{3}{2} \right) \frac{Z}{1 - e^{-\beta\hbar\Omega_j}} + \mathcal{O}(\epsilon^2), \\ &= \frac{2Z}{(1 - e^{-\beta\hbar\Omega_j})^2} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (\text{A.43})$$

Substituting this back gives the integrated intensity for first overtones as

$$\begin{aligned} I(\Omega_j + \Omega_j) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j + \Omega_j) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijj}}{(\Omega_j + \Omega_j)^2 - \Omega_i^2} \right|^2 \\ &\quad \times 2\hbar \frac{(1 - e^{-\beta\hbar(\Omega_j + \Omega_j)})}{(1 - e^{-\beta\hbar\Omega_j})^2} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.44})$$

### Combinations

Binary combinations ( $\omega = \Omega_j + \Omega_k$ ) arise from two excitations in different modes,

$$|\mathbf{n} + \Delta \mathbf{n}\rangle = \frac{\hat{a}_j^\dagger \hat{a}_k^\dagger}{\sqrt{(n_j + 1)(n_k + 1)}} |\mathbf{n}\rangle. \quad (\text{A.45})$$

Using Eq. (A.24d), the corresponding transition dipole moment is

$$\langle \mathbf{n} | \hat{\boldsymbol{\mu}} | \mathbf{n} + \Delta \mathbf{n} \rangle = \epsilon \frac{\hbar}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j + \Omega_k)^2 - \Omega_i^2} \sqrt{(n_j + 1)(n_k + 1)} + \mathcal{O}(\epsilon^2), \quad (\text{A.46})$$

and Eq. (A.28) is

$$\begin{aligned} I(\Omega_j + \Omega_k) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j + \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j + \Omega_k)^2 - \Omega_i^2} \right|^2 \\ &\quad \times \hbar \frac{(1 - e^{-\beta \hbar (\Omega_j + \Omega_k)})}{Z} \sum_{\mathbf{n}} (n_j + 1)(n_k + 1) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.47})$$

Using Eq. (A.29) and Eq. (A.31), the above sum can then be evaluated

$$\begin{aligned} \sum_{\mathbf{n}} (n_j + 1)(n_k + 1) e^{-\beta E_{\mathbf{n}}} &= \sum_{\mathbf{n}} \left( -\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) \left( -\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_k} + \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2), \\ &= \left( -\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) \left( -\frac{1}{\beta \hbar} \frac{\partial}{\partial \Omega_k} + \frac{1}{2} \right) Z + \mathcal{O}(\epsilon^2), \\ &= \frac{Z}{(1 - e^{-\beta \hbar \Omega_j})(1 - e^{-\beta \hbar \Omega_k})} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (\text{A.48})$$

Substituting this back gives the integrated intensity for binary combinations as

$$\begin{aligned} I(\Omega_j + \Omega_k) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j + \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j + \Omega_k)^2 - \Omega_i^2} \right|^2 \\ &\quad \times \hbar \frac{(1 - e^{-\beta \hbar (\Omega_j + \Omega_k)})}{(1 - e^{-\beta \hbar \Omega_j})(1 - e^{-\beta \hbar \Omega_k})} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.49})$$

### Differences

Binary differences ( $\omega = \Omega_j - \Omega_k$ ) arise from an excitation in one mode and a de-excitation in another,

$$|\mathbf{n} + \Delta\mathbf{n}\rangle = \frac{\hat{a}_j^\dagger \hat{a}_k}{\sqrt{(n_j + 1)n_k}} |\mathbf{n}\rangle. \quad (\text{A.50})$$

Using Eq. (A.24e), the corresponding transition dipole moment is

$$\langle \mathbf{n} | \hat{\boldsymbol{\mu}} | \mathbf{n} + \Delta\mathbf{n} \rangle = \epsilon \frac{\hbar}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j - \Omega_k)^2 - \Omega_i^2} \sqrt{(n_j + 1)n_k} + \mathcal{O}(\epsilon^2), \quad (\text{A.51})$$

and Eq. (A.28) is

$$\begin{aligned} I(\Omega_j - \Omega_k) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j - \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j - \Omega_k)^2 - \Omega_i^2} \right|^2 \\ &\quad \times \hbar \frac{(1 - e^{-\beta\hbar(\Omega_j - \Omega_k)})}{Z} \sum_{\mathbf{n}} (n_j + 1)n_k e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.52})$$

Using Eq. (A.29) and Eq. (A.31), the above sum can then be evaluated

$$\begin{aligned} \sum_{\mathbf{n}} (n_j + 1)n_k e^{-\beta E_{\mathbf{n}}} &= \sum_{\mathbf{n}} \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_k} - \frac{1}{2} \right) e^{-\beta E_{\mathbf{n}}} + \mathcal{O}(\epsilon^2), \\ &= \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_j} + \frac{1}{2} \right) \left( -\frac{1}{\beta\hbar} \frac{\partial}{\partial \Omega_k} - \frac{1}{2} \right) Z + \mathcal{O}(\epsilon^2), \\ &= \frac{-Z}{(1 - e^{-\beta\hbar\Omega_j})(1 - e^{\beta\hbar\Omega_k})} + \mathcal{O}(\epsilon^2). \end{aligned} \quad (\text{A.53})$$

Substituting this back gives the integrated intensity for binary differences as

$$\begin{aligned} I(\Omega_j - \Omega_k) &= \epsilon^2 \frac{\pi}{3c\epsilon_0 v} (\Omega_j - \Omega_k) \left| \frac{1}{4} \sum_i^N \frac{\Omega_i \tilde{\boldsymbol{\mu}}_i \tilde{\eta}_{ijk}}{(\Omega_j - \Omega_k)^2 - \Omega_i^2} \right|^2 \\ &\quad \times -\hbar \frac{(1 - e^{-\beta\hbar(\Omega_j - \Omega_k)})}{(1 - e^{-\beta\hbar\Omega_j})(1 - e^{\beta\hbar\Omega_k})} + \mathcal{O}(\epsilon^3). \end{aligned} \quad (\text{A.54})$$



## APPENDIX B

# ITERATIVE BOLTZMANN INVERSION CONVERGENCE

This appendix presents the complete convergence of iterative Boltzmann inversion (IBI) for the f-QCMD calculations in chapter 4. For more information on iterative Boltzmann inversion, see section 3.3.2 and section 4.1.

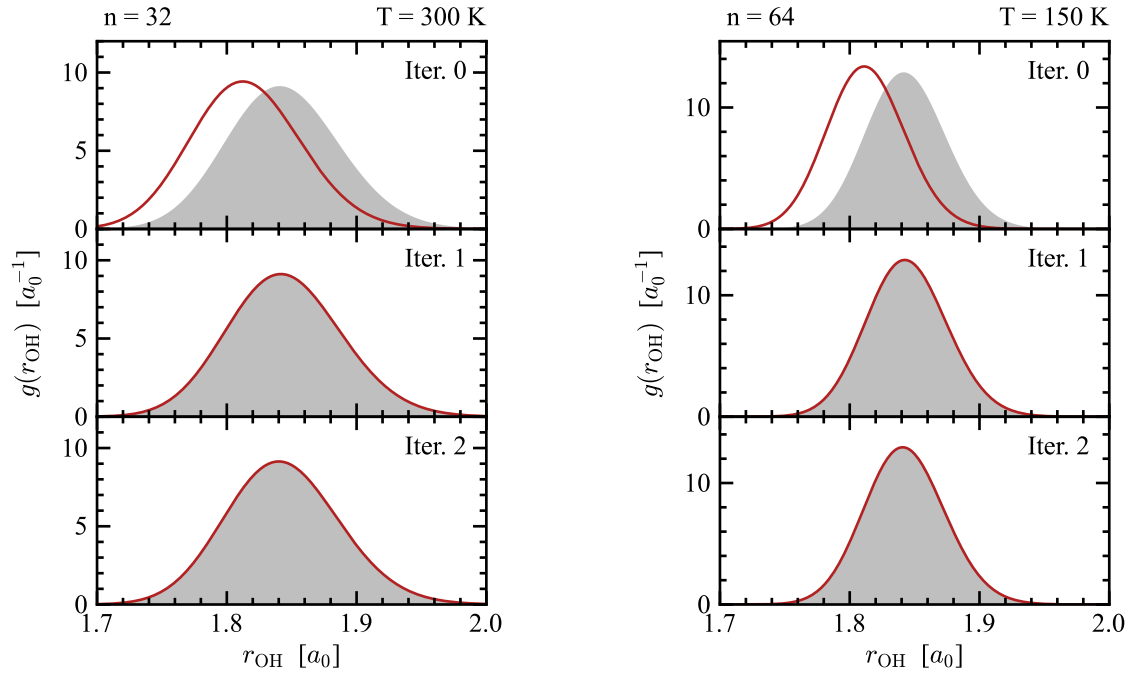
In the case of f-QCMD, the targets for IBI are quasi-centroid distribution functions from path integral calculations. These are shown along with the iterative distribution functions at every step until convergence was achieved.

The same process was also used to produce the CMD results in chapter 4. In every case the convergence for CMD was just as quick, if not quicker, than f-QCMD. Therefore, convergence is only shown for f-QCMD.

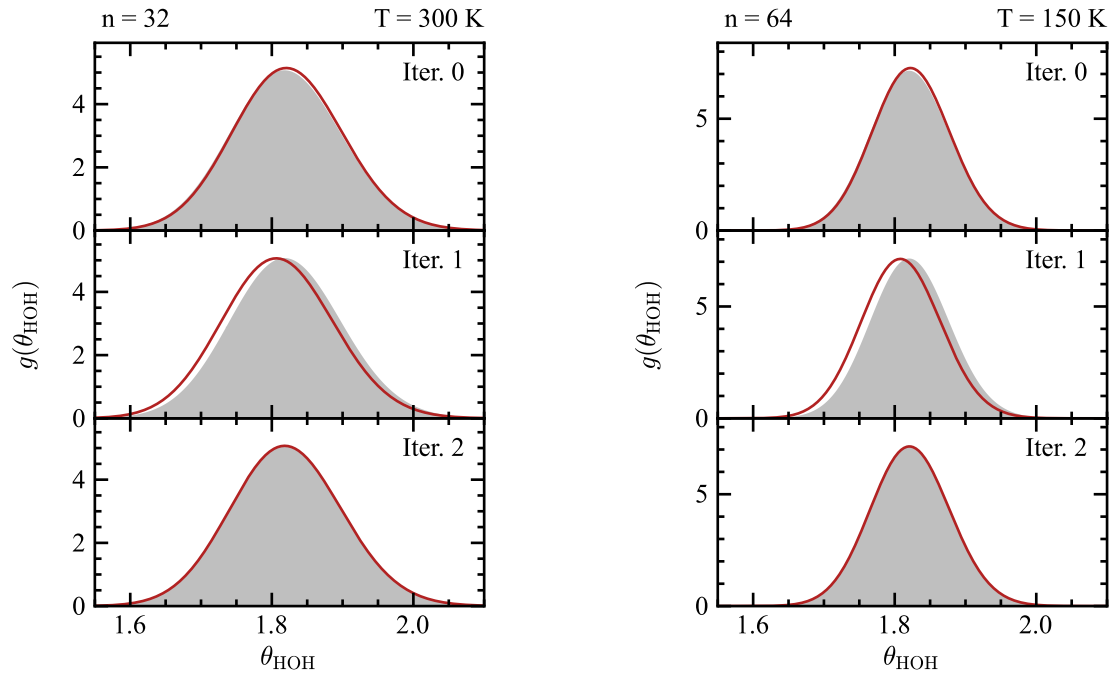
## LEGEND

-  Quasi-Centroid (Path Integral)
-  Iterative Boltzmann Inversion



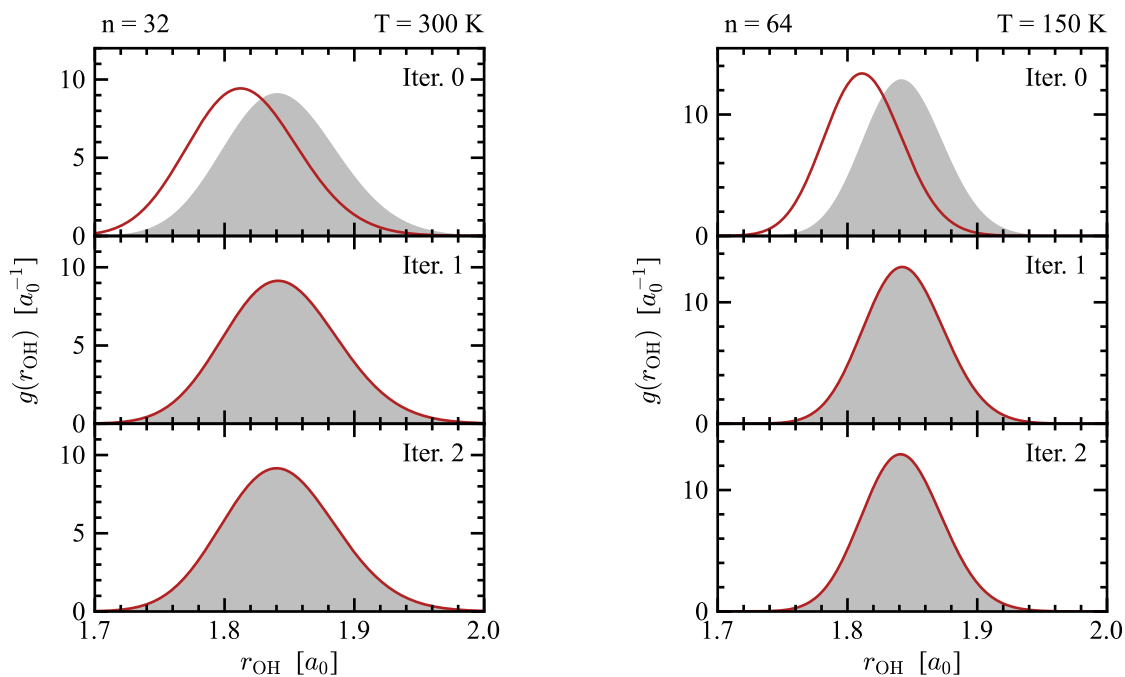
B.1 H<sub>2</sub>O

**Figure B.1.i:** IBI convergence of the O–H bond length in H<sub>2</sub>O.

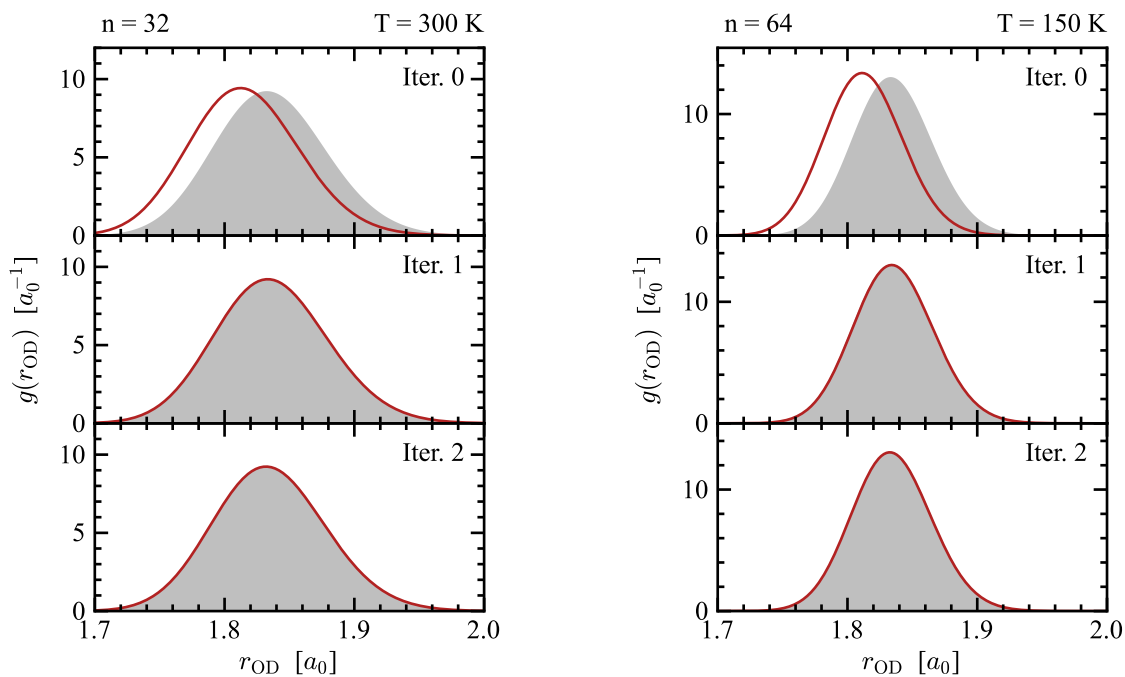


**Figure B.1.ii:** IBI convergence of the H–O–H bond angle in H<sub>2</sub>O.

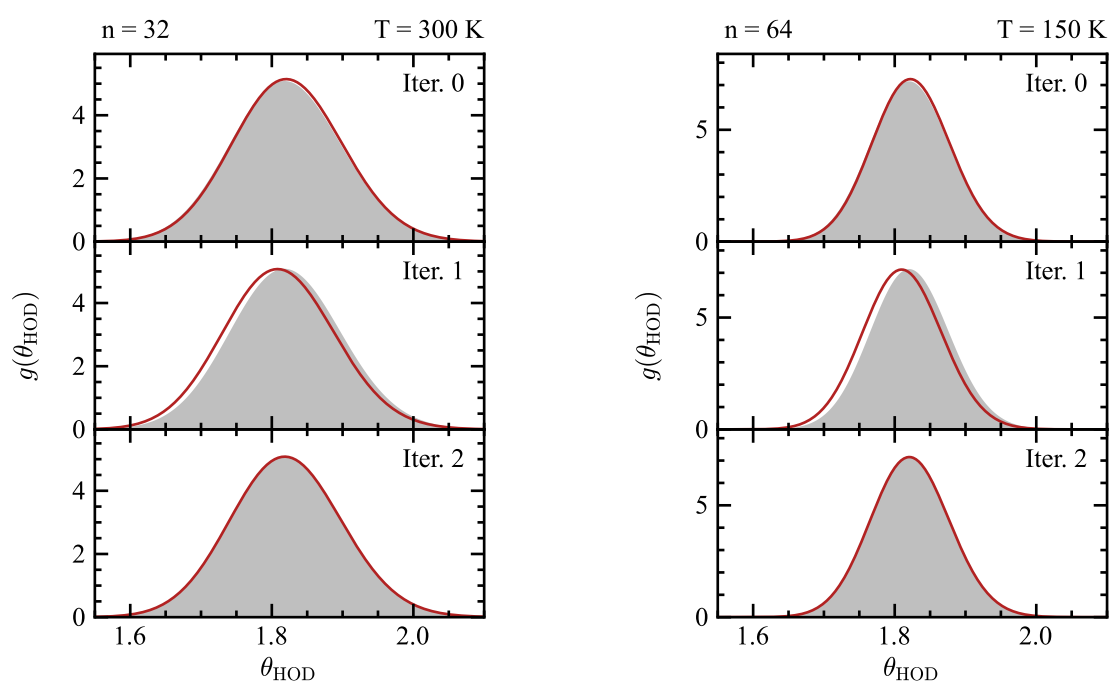
## B.2 HDO



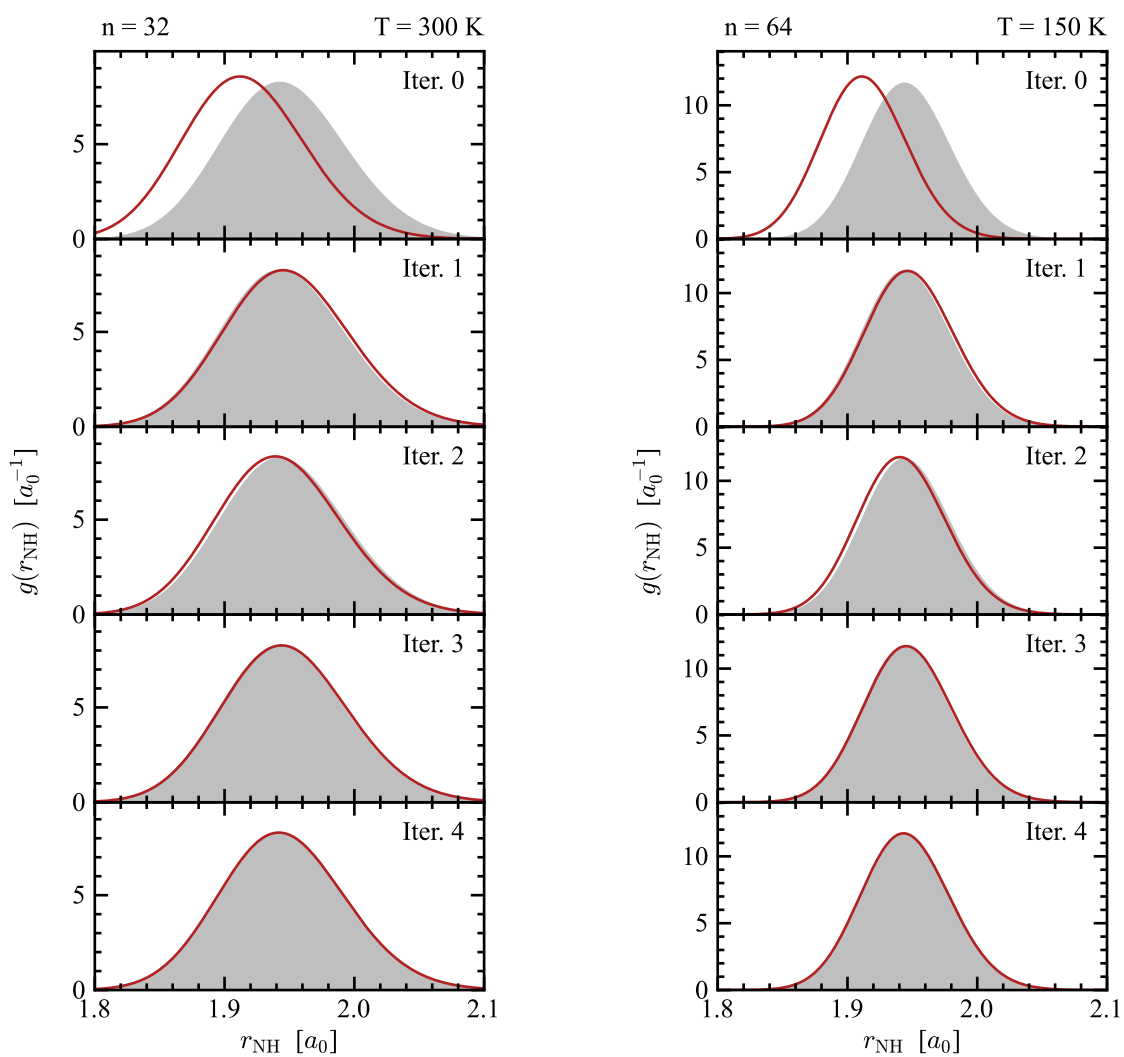
**Figure B.2.i:** IBI convergence of the O–H bond length in HDO.



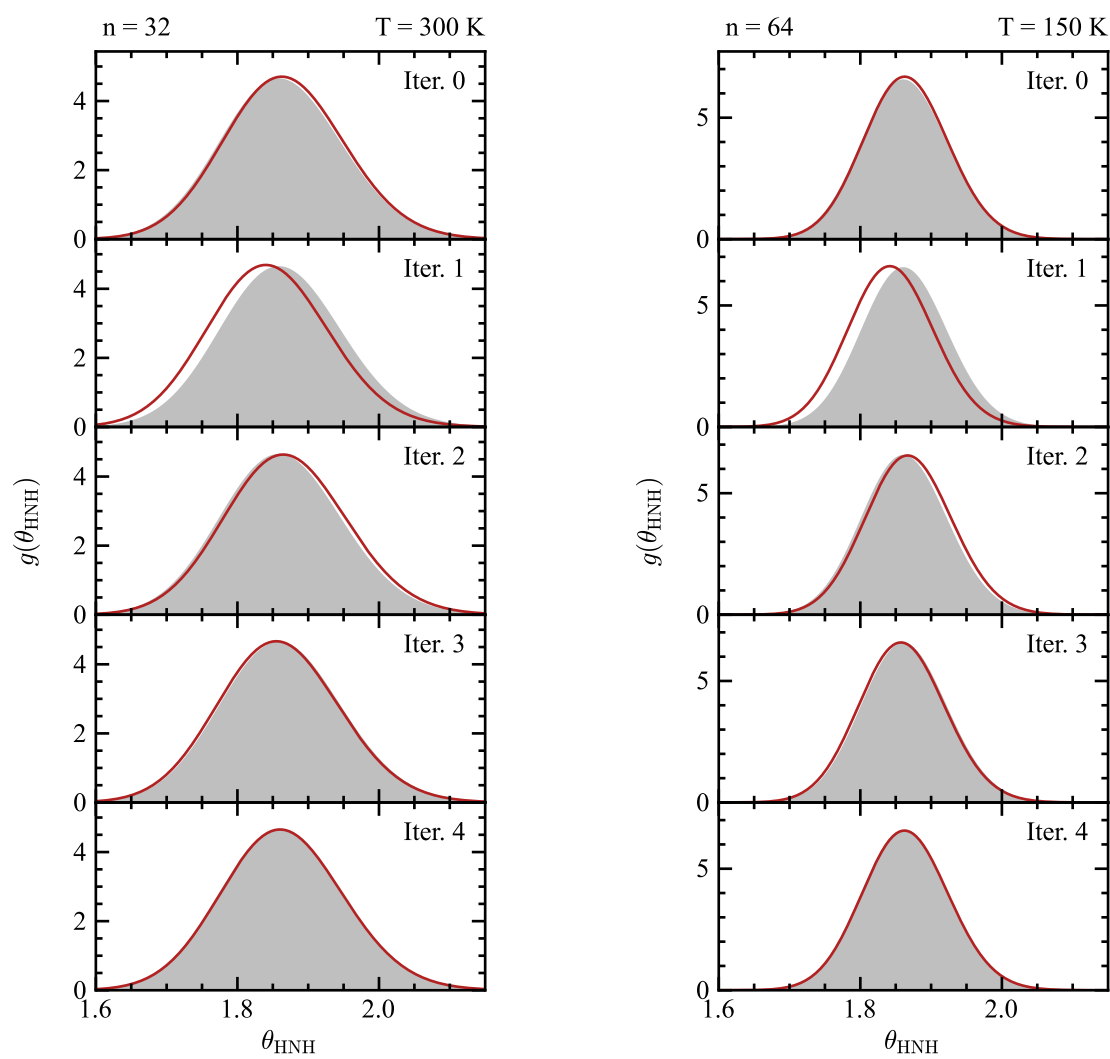
**Figure B.2.ii:** IBI convergence of the O–D bond length in HDO.



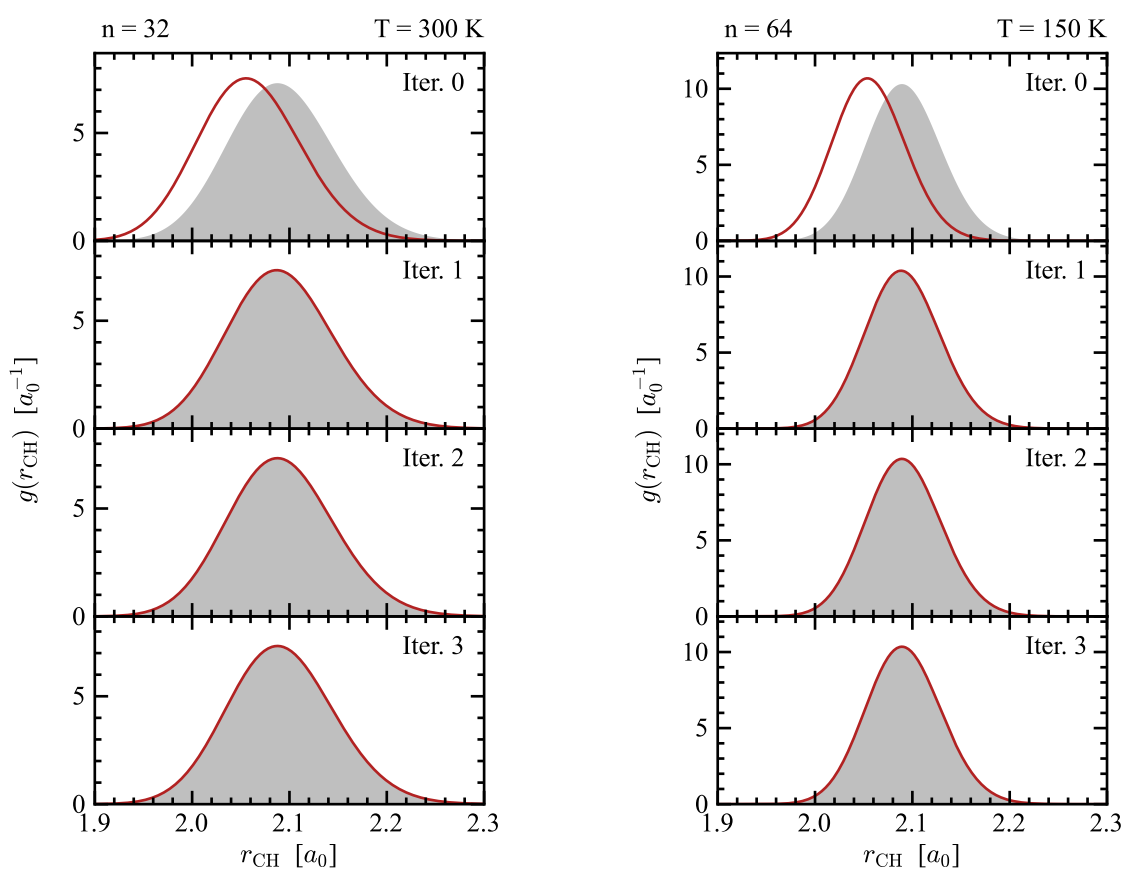
**Figure B.2.iii:** IBI convergence of the H–O–D bond angle in HDO.

B.3  $\text{NH}_3$ 

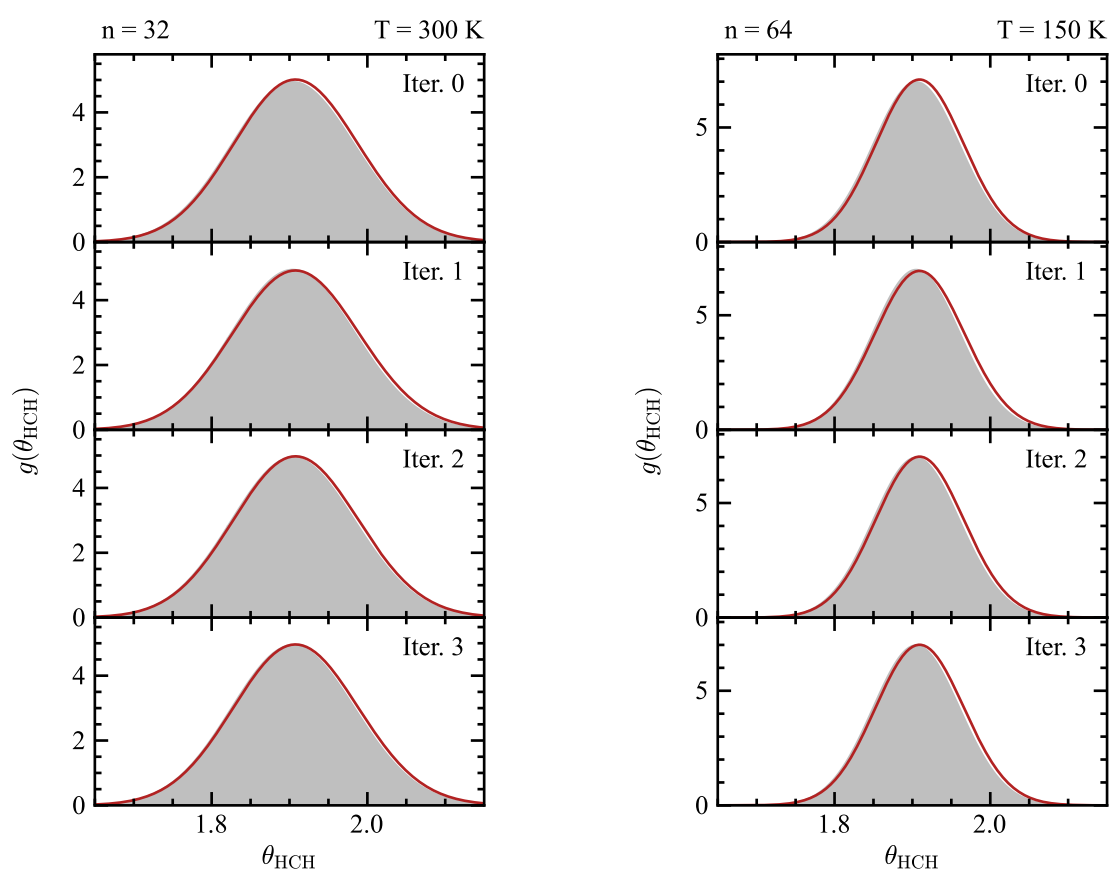
**Figure B.3.i:** IBI convergence of the N–H bond length in  $\text{NH}_3$ .



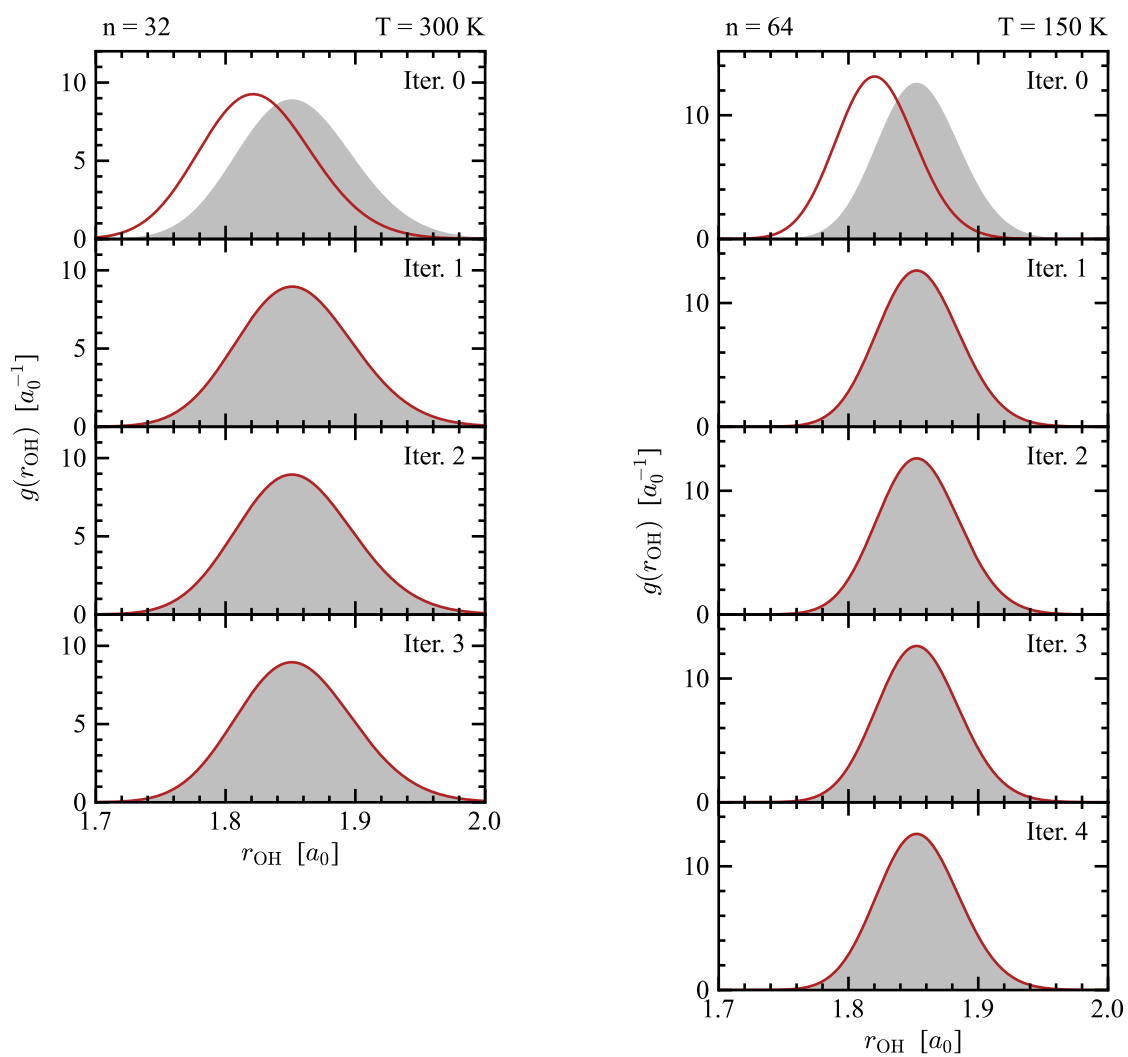
**Figure B.3.ii:** IBI convergence of the H–N–H bond angle in NH<sub>3</sub>.

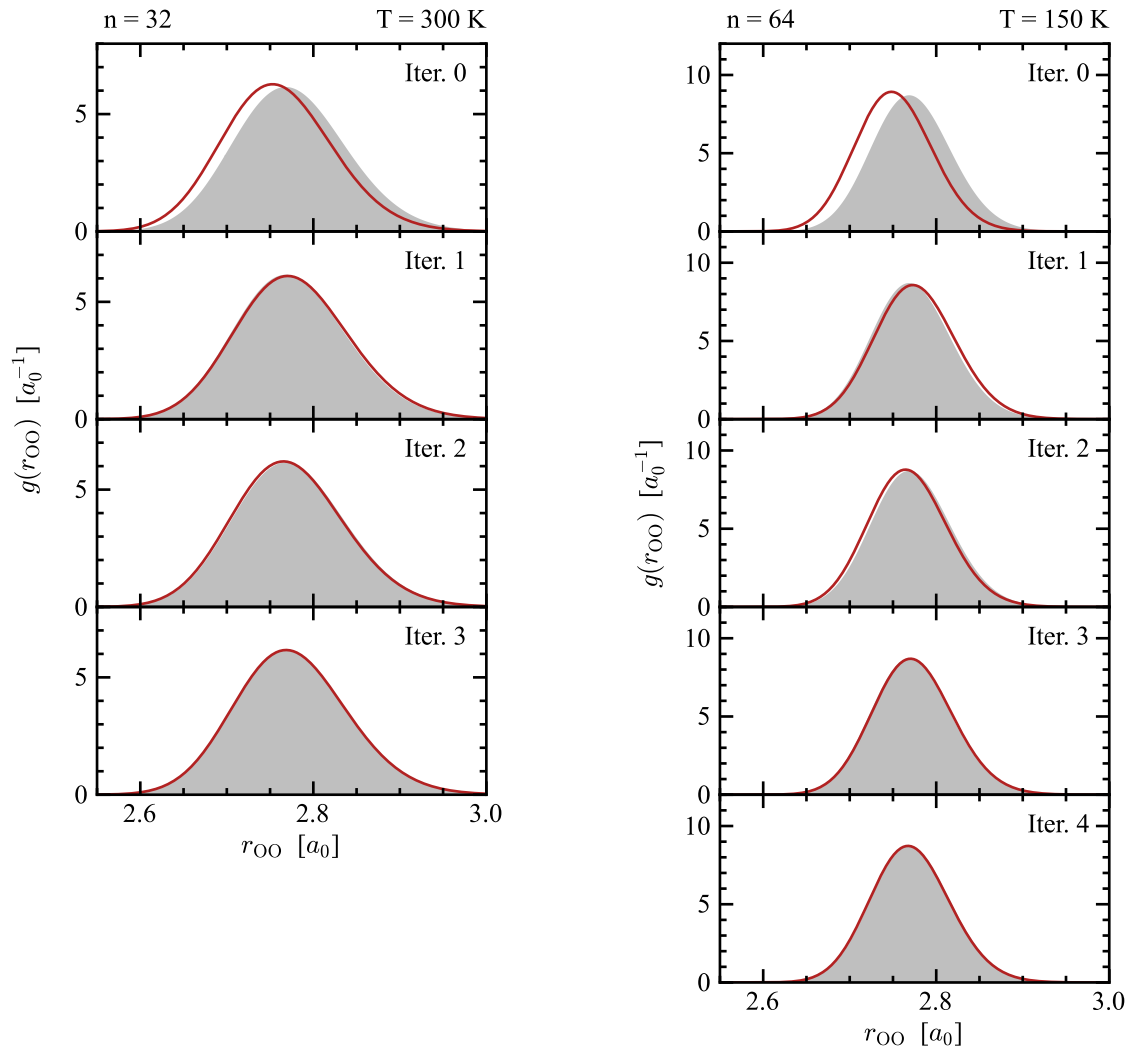
B.4 CH<sub>4</sub>

**Figure B.4.i:** IBI convergence of the C–H bond length in CH<sub>4</sub>.

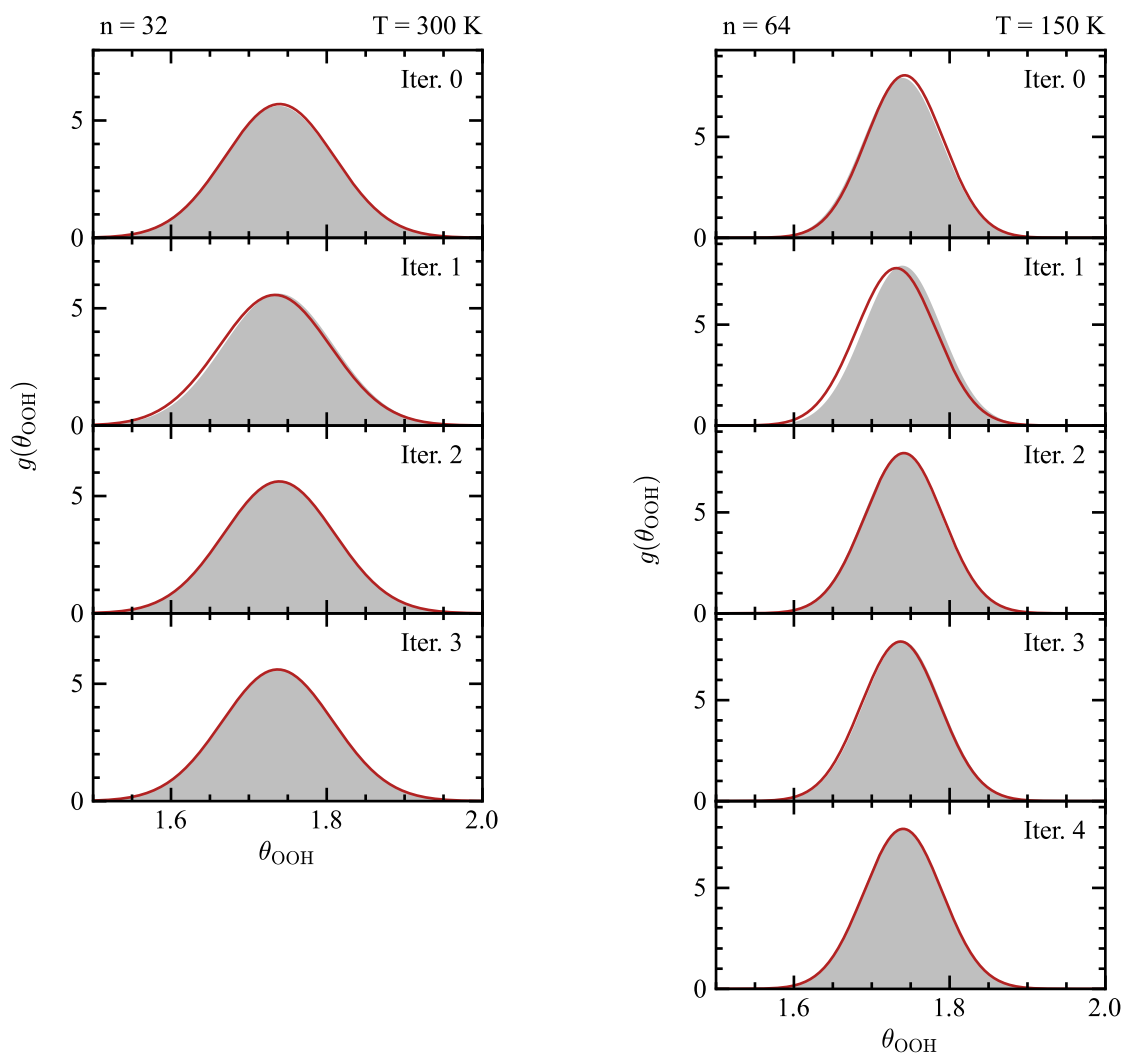


**Figure B.4.ii:** IBI convergence of the H–C–H bond angle in CH<sub>4</sub>.

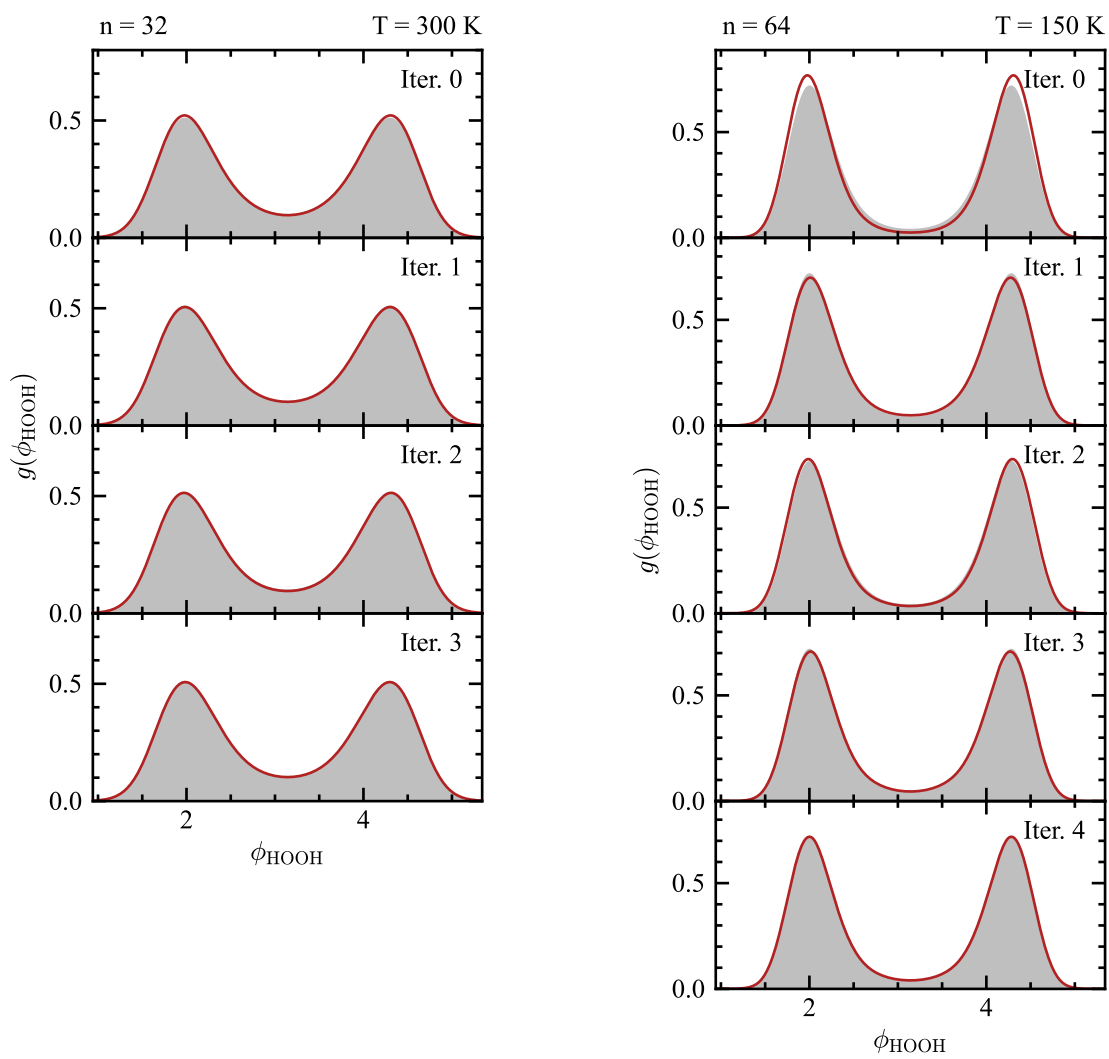
B.5  $\text{H}_2\text{O}_2$ **Figure B.5.i:** IBI convergence of the O-H bond length in  $\text{H}_2\text{O}_2$ .



**Figure B.5.ii:** IBI convergence of the O–O bond length in H<sub>2</sub>O<sub>2</sub>.



**Figure B.5.iii:** IBI convergence of the O–O–H bond angle in H<sub>2</sub>O<sub>2</sub>.



**Figure B.5.iv:** IBI convergence of the torsion angle in H<sub>2</sub>O<sub>2</sub>.



## APPENDIX C

### REDUCED VARIANCE ESTIMATORS

#### C.1 JACOBIAN DERIVATION

Let  $(\mathbf{q}_1, \mathbf{q}_2, \dots)$  be a set of 3D Cartesian position vectors. We define the vector  $\mathbf{q}_1$  relative to  $\mathbf{q}_2$  in spherical coordinates

$$\mathbf{q}_1 = \mathbf{q}_2 + \mathbf{r}_{12}, \quad (\text{C.1a})$$

$$\mathbf{r}_{12} = r \cos \phi \sin \theta \hat{\mathbf{u}} + r \sin \phi \sin \theta \hat{\mathbf{v}} + r \cos \theta \hat{\mathbf{w}}, \quad (\text{C.1b})$$

where  $r \in [0, \infty]$ ,  $\phi \in [0, 2\pi]$ ,  $\theta \in [0, \pi]$ . The basis vectors  $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{\mathbf{w}})$  are orthonormal and, in general, can depend on the other position vectors  $(\mathbf{q}_2, \mathbf{q}_3, \dots)$ .

We now change variables from  $(\mathbf{q}_1, \mathbf{q}_2, \dots) \rightarrow (r, \phi, \theta, \mathbf{q}_2, \dots)$ . The Jacobian associated with this transformation is

$$J = \left\| \frac{\partial(\mathbf{q}_1, \mathbf{q}_2, \dots)}{\partial(r, \phi, \theta, \mathbf{q}_2, \dots)} \right\| = \left\| \frac{\partial(x_1, y_1, z_1, x_2, y_2, \dots)}{\partial(r, \phi, \theta, x_2, y_2, \dots)} \right\| \quad (\text{C.2})$$

where

$$\frac{\partial(f_1, f_2, \dots)}{\partial(g_1, g_2, \dots)} = \begin{pmatrix} \frac{\partial f_1}{\partial g_1} & \frac{\partial f_1}{\partial g_2} & \dots \\ \frac{\partial f_2}{\partial g_1} & \frac{\partial f_2}{\partial g_2} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}. \quad (\text{C.3})$$

When expanded as a block matrix, the Jacobian is

$$J = \begin{vmatrix} \frac{\partial \mathbf{q}_1}{\partial(r, \phi, \theta)} & \frac{\partial \mathbf{q}_1}{\partial(\mathbf{q}_2, \mathbf{q}_3, \dots)} \\ \frac{\partial(\mathbf{q}_2, \mathbf{q}_3, \dots)}{\partial(r, \phi, \theta)} & \frac{\partial(\mathbf{q}_2, \mathbf{q}_3, \dots)}{\partial(\mathbf{q}_2, \mathbf{q}_3, \dots)} \end{vmatrix}. \quad (\text{C.4})$$

Clearly, the bottom-right sub-matrix is the identity matrix. In the bottom-left sub-matrix, every partial derivatives will keep  $\mathbf{q}_2, \mathbf{q}_3, \dots$  constant, and so every element will be 0,

$$J = \begin{vmatrix} \frac{\partial \mathbf{q}_1}{\partial(r, \phi, \theta)} & \frac{\partial \mathbf{q}_1}{\partial(\mathbf{q}_2, \mathbf{q}_3, \dots)} \\ \mathbf{0} & \mathbf{1} \end{vmatrix}. \quad (\text{C.5})$$

Therefore, all elements can be ignored except for the top-left sub-matrix,

$$J = \left\| \frac{\partial \mathbf{q}_1}{\partial(r, \phi, \theta)} \right\| = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \phi} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial y_1}{\partial r} & \frac{\partial y_1}{\partial \phi} & \frac{\partial y_1}{\partial \theta} \\ \frac{\partial z_1}{\partial r} & \frac{\partial z_1}{\partial \phi} & \frac{\partial z_1}{\partial \theta} \end{vmatrix}, \quad (\text{C.6})$$

which can also be expressed as a triple product

$$J = \left| \begin{pmatrix} \frac{\partial x_1}{\partial r} \\ \frac{\partial y_1}{\partial r} \\ \frac{\partial z_1}{\partial r} \end{pmatrix} \cdot \left[ \begin{pmatrix} \frac{\partial x_1}{\partial \phi} \\ \frac{\partial y_1}{\partial \phi} \\ \frac{\partial z_1}{\partial \phi} \end{pmatrix} \times \begin{pmatrix} \frac{\partial x_1}{\partial \theta} \\ \frac{\partial y_1}{\partial \theta} \\ \frac{\partial z_1}{\partial \theta} \end{pmatrix} \right] \right| = \left| \frac{\partial \mathbf{q}_1}{\partial r} \cdot \left( \frac{\partial \mathbf{q}_1}{\partial \phi} \times \frac{\partial \mathbf{q}_1}{\partial \theta} \right) \right|. \quad (\text{C.7})$$

The partial derivative of  $\mathbf{q}_1$  with respect to  $r$ ,  $\phi$  and  $\theta$  are as follows.

$$\begin{aligned} \frac{\partial \mathbf{q}_1}{\partial r} &= \cos \phi \sin \theta \hat{\mathbf{u}} + \sin \phi \sin \theta \hat{\mathbf{v}} + \cos \theta \hat{\mathbf{w}}, \\ &= \frac{\mathbf{r}_{12}}{r}. \end{aligned} \quad (\text{C.8})$$

$$\begin{aligned}
\frac{\partial \mathbf{q}_1}{\partial \phi} &= -r \sin \phi \sin \theta \hat{\mathbf{u}} + r \cos \phi \sin \theta \hat{\mathbf{v}}, \\
&= \hat{\mathbf{w}} \times \mathbf{r}_{12}.
\end{aligned} \tag{C.9}$$

$$\begin{aligned}
\frac{\partial \mathbf{q}_1}{\partial \theta} &= r \cos \phi \cos \theta \hat{\mathbf{u}} + r \sin \phi \cos \theta \hat{\mathbf{v}} - r \sin \theta \hat{\mathbf{w}}, \\
&= \frac{\cos \theta}{\sin \theta} \left( \mathbf{r}_{12} - \frac{r}{\cos \theta} \hat{\mathbf{w}} \right).
\end{aligned} \tag{C.10}$$

Substituting these results back into the expression for  $J$  gives

$$\begin{aligned}
J &= \left| \frac{\mathbf{r}_{12}}{r} \cdot \left( (\hat{\mathbf{w}} \times \mathbf{r}_{12}) \times \frac{\cos \theta}{\sin \theta} \left( \mathbf{r}_{12} - \frac{r}{\cos \theta} \hat{\mathbf{w}} \right) \right) \right|, \\
&= \left| \frac{\cos \theta}{r \sin \theta} \mathbf{r}_{12} \cdot \left( (\hat{\mathbf{w}} \times \mathbf{r}_{12}) \times \mathbf{r}_{12} - \frac{r}{\cos \theta} (\hat{\mathbf{w}} \times \mathbf{r}_{12}) \times \hat{\mathbf{w}} \right) \right|.
\end{aligned} \tag{C.11}$$

Using the identity

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}, \tag{C.12}$$

then gives

$$\begin{aligned}
J &= \left| \frac{\cos \theta}{r \sin \theta} \mathbf{r}_{12} \cdot \left( (\hat{\mathbf{w}} \cdot \mathbf{r}_{12})\mathbf{r}_{12} - (\mathbf{r}_{12} \cdot \mathbf{r}_{12})\hat{\mathbf{w}} - \frac{r}{\cos \theta} (\hat{\mathbf{w}} \cdot \hat{\mathbf{w}})\mathbf{r}_{12} + \frac{r}{\cos \theta} (\mathbf{r}_{12} \cdot \hat{\mathbf{w}})\hat{\mathbf{w}} \right) \right|, \\
&= \left| \frac{\cos \theta}{r \sin \theta} \mathbf{r}_{12} \cdot \left( r \cos \theta \mathbf{r}_{12} - r^2 \hat{\mathbf{w}} - \frac{r}{\cos \theta} \mathbf{r}_{12} + r^2 \hat{\mathbf{w}} \right) \right|, \\
&= \left| \frac{1}{\sin \theta} (\mathbf{r}_{12} \cdot \mathbf{r}_{12}) (\cos^2 \theta - 1) \right|, \\
&= |-r^2 \sin \theta|, \\
&= r^2 \sin \theta.
\end{aligned} \tag{C.13}$$

This confirms the intuitive result that the Jacobian is the same as the standard Cartesian to spherical coordinate transform.

## C.2 IMPLEMENTATION

This section presents a computational implementation of reduced variance estimators that is both efficient and minimises memory usage.

The most straightforward approach is to store  $\xi(\mathbf{x})$  and  $W(\mathbf{x})$  for each of the  $S$  samples of  $\mathbf{x}$ . Evaluating  $g(\Xi; \lambda^*)$  at a single value of  $\Xi$  then takes  $\mathcal{O}(S)$  operations, with potential optimisation if the samples are first sorted in order of  $\xi$ . Retaining each individual sample gives complete flexibility as to where the distribution is evaluated, but necessitates  $\mathcal{O}(S)$  space complexity. This could be problematic, as needing several thousand samples (or more) is not an unreasonable scenario.

An alternative, at the cost of complete flexibility in evaluating  $g(\Xi)$ , is to choose an array of  $N$  different  $\Xi$  values at which the distribution will be evaluated.

---

```
# Ordered array of  $\Xi$  values

# We can treat  $\Xi_{-1} = \Xi_{\min}$  and  $\Xi_N = \Xi_{\max}$ 

 $\Xi = \{\Xi_0, \dots, \Xi_{N-1}\}$ 
```

---

During the calculation, we accumulate a separate array where element  $j$  is the sum of all the samples of  $W(\mathbf{x})$  for which  $\Xi_{j-1} \leq \xi(\mathbf{x}) < \Xi_j$ ; the first and last elements contain any samples outside of the defined range. This requires  $\mathcal{O}(N)$  memory, which is likely to be better than  $\mathcal{O}(S)$  provided a reasonable grid of  $\Xi$  values is chosen.

---

```
 $\mathbf{W} = \{W_0, \dots, W_N\} \leftarrow \mathbf{0}$ 
 $\mathbf{Y} = \{Y_0, \dots, Y_N\} \leftarrow \mathbf{0}$ 
for  $i$  in  $\{0, \dots, S-1\}$ 
     $\mathbf{x} \leftarrow \text{sample}(\dots)$  # Sample  $\mathbf{x}$  from the probability distribution
     $j \leftarrow \text{search}(\Xi, \xi(\mathbf{x}))$  # Find  $j$  such that  $\Xi_{j-1} \leq \xi(\mathbf{x}) < \Xi_j$ 
     $W_j \leftarrow W_j + W(\mathbf{x})$ 
     $Y_j \leftarrow Y_j + W(\mathbf{x})^2$ 
end for
```

---

Because  $\Xi$  is sorted, searching for  $j$  requires  $\mathcal{O}(\log N)$  operations in general, and is trivial if the array is uniformly spaced.

The distribution function, evaluated from either end, can then be recovered in  $\mathcal{O}(N)$  operations by calculating the prefix/suffix sum of  $\mathbf{W}$ .

---

```
function DISTRIBUTION_FROM_MIN( $\mathbf{W}$ ,  $S$ ,  $g(\Xi_{\min})$ )
```

```
   $\mathbf{g}$  =  $\{g_0, \dots, g_{N-1}\}$ 
```

```
   $g_0 \leftarrow W_0$ 
```

```
  for  $j$  in  $\{1, \dots, N-1\}$ 
```

```
     $g_j \leftarrow g_{j-1} + W_j$ 
```

```
  end for
```

```
  return  $g(\Xi_{\min}) + \mathbf{g}/S$ 
```

```
end function
```

---

```
function DISTRIBUTION_FROM_MAX( $\mathbf{W}$ ,  $S$ ,  $g(\Xi_{\max})$ )
```

```
   $\mathbf{g} = \{g_0, \dots, g_{N-1}\}$ 
```

```
   $g_{N-1} \leftarrow W_N$ 
```

```
  for  $j$  in  $\{N-1, \dots, 1\}$ 
```

```
     $g_{j-1} \leftarrow g_j + W_j$ 
```

```
  end for
```

```
  return  $g(\Xi_{\max}) - \mathbf{g}/S$ 
```

```
end function
```

---

Calculating  $\lambda^*$  with  $\mathcal{O}(N)$  memory is also possible by rewriting Eq. (5.13) as

$$\lambda^*(\Xi) = \frac{\langle h(\Xi - \xi)W^2 \rangle - \langle h(\Xi - \xi)W \rangle \langle W \rangle}{\langle W^2 \rangle - \langle W \rangle^2}. \quad (\text{C.14})$$

This requires one further array of size  $N + 1$ , this time storing samples of  $W(\mathbf{x})^2$ , which is denoted here as  $\mathbf{Y}$ . Noting that the denominator of the above expression

is simply the  $\Xi \rightarrow \Xi_{\max}$  limit of the numerator,  $\lambda^*$  can be calculated in  $\mathcal{O}(N)$  operations as follows.

---

```

function MIN_VARIANCE_FRACTION(W, Y, S)
     $\langle W \rangle \leftarrow \text{sum}(\mathbf{W})/S$ 
     $\boldsymbol{\lambda} = \{\lambda_0, \dots, \lambda_N\}$ 
     $\lambda_0 \leftarrow Y_0 - W_0 \langle W \rangle$ 
    for j in  $\{1, \dots, N\}$ 
         $\lambda_j \leftarrow \lambda_{j-1} + Y_j - W_j \langle W \rangle$ 
    end for
    return  $\{\lambda_0, \dots, \lambda_{N-1}\}/\lambda_N$ 
end function

```

---

Combining all three algorithms into a single  $\mathcal{O}(N)$  function for  $g(\Xi; \lambda^*)$  is straightforward. However, it should be stressed that the computational cost of these algorithms is unlikely to be significant compared to the cost of actually sampling the distribution; the real benefit is the potential savings on memory compared to storing every sample.

