Use and Development of Direct Quantum Dynamical Methods for Photochemistry

Léon L. E. Cigrang

2025



Department of Chemistry

A thesis submitted for the examination of
Doctor of Philosophy
at
University College London

I, Léon L. E. Cigrang confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Computational chemistry and molecular physics come together in a field known as Quantum Dynamics, which is about studying the time evolution of physical systems at the quantum mechanical level. Even though quantum theory provides an equation that completely describes this evolution, i.e. the Schrödinger equation, there does not exist a tractable way to solve it for the vast majority of molecular systems. This thesis explores that problem by using and developing computational methods which can provide approximate solutions and thereby contribute to a detailed mechanistic understanding of photoinduced reactions. The methods described in this work are based on the propagation of coupled nuclear wavepackets, ideally suited to account for the effects of non-adiabatic phenomena between excited electronic states. They are typically more computationally demanding than many alternatives, but higher accuracy is achieved in most cases. This is demonstrated in a comprehensive study of the photodissociation reactions of methanol, showing how high-level electronic structure calculations are used in combination with direct dynamics simulations to provide a comprehensive description of the fundamental phenomena that govern the observed reactions of photo-excited molecules. This is followed by the outline of a newly developed procedure that aims to incorporate the effects of explicit atomistic environments into accurate non-adiabatic quantum dynamics calculations. The approach is based on combining the efficiency of classical molecular mechanics calculations, with the accuracy of coupled nuclear wavepacket propagations. Results are shown which demonstrate the potential applicability of this method. Finally, a critical analysis is given of the various steps required to successfully use direct wavepacket methods. The aim here is to remove practical bottlenecks by suggesting possible automations, as well as consider best practices. It is argued that such evaluations will be invaluable in the context of community-wide efforts to progress the field as a whole.

Impact Statement

The immediate impact of the work presented in this thesis is mostly academic, but it has the potential to extend its reach into many areas of industrial and/or commercial interest. The topics covered are directly relevant to the community of users and developers of quantum dynamics simulations, and thus benefits those wanting to understand, in detail, the many challenges and opportunities related to the accurate modelling of molecular motion. In terms of applications for these simulations, they help with the interpretation of complicated spectroscopic experiments, which together deepen our understanding of the nature of light-matter interactions. It is not difficult to imagine how this might translate to the development of new materials designed to interact with light. This category includes sunscreens, sensors used in medical imaging, solar cells, and semiconductors just to name a few.

To be more specific, a chapter is dedicated to the study of the photo-dissociation reactions of methanol. Apart from offering insights into the behaviour of methanol itself, the methods discussed in this work should find general applicability in the study of small organic molecules and their reactions to incoming radiation. Such data is invaluable in atmospheric models used to predict the evolution of Earth's climate, and can be extended to outer space where accurate simulations of molecules observed in the interstellar medium contribute to working theories of how and where life might have originated, and how to spot its signatures on extraterrestrial bodies.

Efforts are also made towards the description of environment interactions on the photochemistry of organic molecules. Increasing our understanding of solvated systems for example is important in the study of biologically relevant processes, such as the degradation of the genetic code.

Finally, the thorough evaluation of various computational procedures carried out in the penultimate chapter should help the field of non-adiabatic dynamics progress towards maturity by establishing common standards and widening participation. Additionally, well defined workflows combined with the automation of certain tasks should contribute to fewer "wasteful" calculations and thus a more efficient use of computational resources. This is in line with the recent Green DiSC initiative which aims to minimise the environmental impacts of a research group's computing activities.

List of Publications

1. Cigrang, L. L. E.; Green, J. A.; Gómez, S.; Cerezo, J.; Improta, R.; Prampolini, G.; Santoro, F.; Worth, G. A. Non-adiabatic direct quantum dynamics using force fields: Toward solvation. *J. Chem. Phys.* **2024**, *160*, 174120.

   (DOI: 10.1063/5.0204911)

   Contributions: Investigation, Methodology, Software, Writing – original draft, Writing – review & editing (equal).

   References: chapter 4

2. Cigrang, L. L. E.; Worth, G. A. Modeling Photodissociation: Quantum Dynamics Simulations of Methanol. *J. Phys. Chem. A* **2024**, *128*, 7546–7557.
   (DOI: 10.1021/acs.jpca.4c03612)

   Contributions: Conceptualisation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing (main).

   References: chapter 3

3. Cigrang, L. L. E.; Curchod, B. F. E.; Ingle, R. A.; Kelly, A.; Mannouch, J. R.; Accomasso, D.; Alijah, A.; Barbatti, M.; Chebbi, W.; Došlíc, N.; Eklund, E. C.; Fernandez-Alberti, S.; Freibert, A.; González, L.; Granucci, G.; Hernández, F. J.; Hernández-Rodríguez, J.; Jain, A.; Janoš, J.; Kassal, I.; Kirrander, A.; Lan, Z.; Larsson, H. R.; Lauvergnat, D.; Dé, B. L.; Lee, Y.; Maitra, N. T.; Min, S. K.; Peláez, D.; Picconi, D.; Raucci, U.; Qiu, Z.; Robertson, P.; Gil, E. S.; Sapunar, M.; Schürger, P.; Sinnott, P.; Tretiak, S.; Tikku, A.; Vindel-Zandbergen, P.; Worth, G. A.; Agostini, F.; Gómez, S.; Ibele, L. M.; Prlj, A. Roadmap for Molecular Benchmarks in Nonadiabatic Dynamics. *J. Phys. Chem.* **2025**, Under review (manuscript available at `https://arxiv.org/abs/2502.14569`)

   Contributions: Writing – section II. A. (main), Conceptualisation, Writing – review & editing (equal).

   References: chapter 5

4. Cigrang, L. L. E.; Saha, S.; Worth, G. A. Including spin-orbit coupling in direct wavepacket dynamics. *Phys. Chem. Chem. Phys.*, In preparation

   Contributions: Conceptualisation, Investigation, Methodology, Software, Conceptualisation, Writing – original draft, Writing – review & editing (equal).

   References: chapter 5

Acknowledgements

First and foremost, I must direct my immense gratitude towards Graham A. Worth. Graham has been my supervisor for nearly four years, and during that time he very generously shared his vast knowledge, experience, and enthusiasm for a scientific discipline that I have now also grown very fond of. His guidance and his approach to scientific research will undoubtedly have a lasting impact on my own future. Thank you Graham, for everything.

Next, I want to acknowledge my parents, SS & CC, and thank them for their unconditional support in all my endeavours, even when they sometimes struggle to understand the things that keep me busy. To both of them, I dedicate the introduction to this thesis and hope that they may find some answers to their many questions about the nature of my research.

Finally, I am also very grateful for the many people with whom I shared my time in London. This includes my colleagues at UCL, past and present, who initially helped me to get started and with whom I could then share my work and discuss exciting scientific ideas. I also had the privilege to be surrounded by many good friends/colleagues, both abroad at conferences or back home in Belgium, and I thank them for making my life so much more enjoyable.

Looking back, all these people have contributed to me thoroughly enjoying my time as a PhD student, and even though I am excited for what is hopefully the start of a long scientifically-minded career, I will certainly miss these last few wonderful years.

Thank you all.

"That which cannot be done perfectly must be done in a manner as near perfection as can be. If exactness cannot, from the nature of things, be attained, then the nearest practicable approach to exactness ought to be made."

- Daniel Webster

Contents

List of Figures **xi**

List of Tables **xiii**

Glossary **xviii**

1 Introduction **1**

1.1 Aims & Outline 2

1.2 Computational modelling 4

1.2.1 Role of mathematics 6

1.2.2 Molecular modelling: classical vs quantum 9

1.3 Photochemistry 12

1.3.1 Theoretical considerations 14

1.3.2 Photodissociation 17

1.4 Role of experimental measurement 19

2 Theory **22**

2.1 Introduction 23

2.2 Electronic Structure Methods 23

2.2.1 The TISE and other generalities 24

2.2.2 Hartree-Fock 26

2.2.3 Post-HF 27

2.2.4 Note on DFT 35

2.2.5 Note on inter-state coupling 36

2.3 Non-adiabatic Quantum Dynamics 37

2.3.1 Born-Oppenheimer and beyond 38

2.3.2 Conical Intersections 44

2.3.3 Vibronic Coupling Hamiltonian . . . 46
2.3.4 Wavepackets & Initial conditions . . . 48
2.3.5 MCTDH . . . 50
2.3.6 vMCG . . . 53
2.3.7 Direct dynamics . . . 55
2.3.8 Note on TSH . . . 60
2.4 Summary . . . 62

3 Modelling photodissociation 65
3.1 Motivation & Background . . . 66
3.2 Computational detail . . . 71
3.2.1 Potential energy surfaces . . . 71
3.2.2 Nuclear dynamics . . . 73
3.2.3 Analysis of the wavefunction . . . 77
3.3 Results . . . 81
3.3.1 Time-independent $\phi^{el}(\mathbf{r};\mathbf{R})$ analysis . . . 81
3.3.2 Time-dependent $\chi^{nuc}(\mathbf{R},t)$ analysis . . . 89
3.4 Reaction scheme . . . 102
3.5 Conclusions & Outlook . . . 103

4 Force Field Quantum Dynamics 107
4.1 Overview . . . 108
4.2 Modelling the environment . . . 108
4.2.1 Implicit solvation models . . . 110
4.2.2 Explicit solvation models . . . 112
4.2.3 Microsolvation . . . 114
4.2.4 Quantum dynamics in solution . . . 115
4.3 FFQD outline . . . 116
4.3.1 Uracil as a benchmark . . . 118
4.4 Theory & Methodology . . . 120
4.4.1 Building a QMD-FF . . . 120
4.4.2 Hierarchical approximations of (DD)-vMCG . . . 122
4.4.3 Workflow - AQUATICS procedure . . . 128

4.5 Uracil benchmark . . . 131
4.5.1 Computational detail . . . 133
4.5.2 vMCG benchmark . . . 137
4.5.3 Direct dynamics benchmark . . . 143
4.6 Uracil in water . . . 147
4.6.1 Set up . . . 147
4.6.2 Effect of solvent configuration . . . 149
4.6.3 Effect of solvent GWPs . . . 151
4.7 Conclusion & Outlook . . . 153

5 Towards accessible quantum dynamics workflows 156
5.1 Motivation & Context . . . 157
5.2 Proposed approach . . . 159
5.3 DD-vMCG workflow . . . 161
5.3.1 Choice of molecule . . . 162
5.3.2 Initialisation . . . 164
5.3.3 Propagation . . . 166
5.3.4 Analysis . . . 170
5.4 Solutions . . . 174
5.4.1 Pruning a QC database . . . 174
5.4.2 Calculation of product ratios . . . 177
5.4.3 Direct dynamics as an exploratory tool . . . 186
5.5 Case studies . . . 189
5.5.1 Case I: Building up a QC database . . . 189
5.5.2 Case II: Population transfer in thioformaldehyde . . . 190
5.5.3 Case III: Pruning the QC database . . . 193
5.5.4 Case IV: Automated calculation of product ratios . . . 195
5.6 Application . . . 195
5.6.1 Electronic motion: state populations . . . 197
5.6.2 Nuclear motion: clustering & product ratios . . . 198
5.6.3 Flux . . . 202
5.6.4 Molecular complexity . . . 203
5.7 Summary & concluding remarks . . . 204

6 Conclusions **209**

Bibliography **215**

Appendices **252**

Appendix A Modelling photodissociation **253**

A.1 Coordinates & Frequencies 254

A.2 CASPT2 intruder states 255

A.3 Local minimum of $\tilde{F}$ state 255

Appendix B Force Field Quantum Dynamics **256**

B.1 Geometry and Frequencies 257

B.2 QVC_0 Hamiltonian Frequencies and Linear Parameters 259

B.3 Intermolecular QMD-FF parameters 260

Appendix C AQUATICS documentation **261**

Appendix D Towards accessible quantum dynamics workflows **272**

D.1 Thioformaldehyde active space 273

D.2 Probability distributions of QC databases 274

D.3 Thioformaldehyde S_1 excitation 275

D.4 Comparison of dimensionality reduction & clustering techniques 277

D.5 Automatic calculation of branching ratios in methanol 278

D.6 Ground state geometries & frequencies 280

D.6.1 Formaldehyde 280

D.6.2 Thioformaldehyde 280

D.6.3 Formic acid 281

D.6.4 Formamide 282

D.7 Validating assignment of reduced coordinates 283

List of Figures

1.1 Sine function (black, dashed) and its Taylor expansion up to 5^{th} (light green) and 7^{th} order (dark blue), expanded around $x = 0$. 8

1.2 Non-radiative photophysical processes induced by photon absorption of energy $h\nu$, illustrated by a system of harmonic oscillators representing different molecular electronic states: singlet ground state (S_0, black), singlet excited states (S_{1-2}, green), and triplet excited state (T_1, purple). Dashed lines represent vibrational energy levels. Red arrows show different non-radiative processes; IC: internal conversion, VR: vibrational relaxation, ISC: intersystem crossing. 13

1.3 Representation of a photochemical reaction, going from Reactants (R) $\rightarrow$ Products (P) via (a) a single excited state surface and (b) internal conversion between two states, mediated by a CoIn. R^* and P^* are the reactants and products in their respective excited states and the label S_n denotes singlet electronic states, with S_0 being the ground state. 16

1.4 Excitation of a molecule from its ground state (S_0) to an unbound electronic excited state (S_1) by a photon. The molecular excited state is denoted by an asterisk. Blue arrows show the direction of nuclear motion. 18

2.1 The RASSCF method. All configurations are allowed in the RAS2 space, whereas only e electrons from RAS1 can be excited into h holes in RAS3. Inactive and virtual orbitals remain doubly occupied and unoccupied, respectively. Frozen and deleted orbitals are not shown. 35

2.2 The crossing of two PESs in (a) the adiabatic picture, leading to an avoided crossing, and (b) the diabatic picture, showing the smooth crossing of the surfaces. The magnitude of the coupling terms are shown in light blue. 43

2.3 Schematic representation of 2 neighbouring adiabatic states, S_n and S_{n+1}, at (a) a conical intersection, and (b) an intersection seam. The branching plane in (a) is defined by two orthogonal vectors **g** and **h**, and in (b) the coordinate along the seam $R \notin \{\mathbf{g}, \mathbf{h}\}$. 46

2.4 Excitation of an initial wavefunction $|\psi_v\rangle$ in its electronic ground state (S_0) to an electronic excited state (S_1), leading to photodissociation of the wavepacket $|\Phi_s(t)\rangle$. x is a nuclear coordinate (e.g. bond stretching) and x_0 the ground state equilibrium geometry. 49

2.5 Spectrum of accuracy vs cost with electronic structure methods in green, and nuclear dynamics methods in red. 63

3.1 Experimental gas-phase UV absorption spectrum of methanol, reproduced from data in ref. [173]. 69

3.2 Optimised geometry of methanol in C_s symmetry, using CCSD//aug-cc-pVDZ. . . 71

3.3 Molecular orbitals of the RAS2 space. (a - d) are initially occupied whereas (e - h) are unoccupied. 72

3.4 Time evolution of a 1-dimensional GWP (initial position $x = 0$) moving on a flat potential. A CAP is defined at $x = 5$ (dashed line), which gradually absorbs the incoming wavepacket. The norm of the wavefunction over time is plotted on the bottom panel. 76

3.5 Potential energy cuts along the O–H stretching mode and the C–O stretching + C–O–H angle bending mode of Methanol at the (a) EOM-CCSD, (b) CASPT2(12,12), and (c) CASPT2(12,12)* levels of theory (* refers to the active space including extra Rydberg orbitals). 84

3.6 Potential energy cuts along the O–H stretching mode and the C–O stretching + C–O–H angle bending modes of Methanol at the RASSCF(12,2+8+2)[1,1] level of theory. Diabatic states have been identified and highlighted based on their chemical character. 85

3.7 Calculated potential energies of the first 4 adiabatic states (top panels, see Figure 3.6 for full PESs), along with CI coefficients (bottom panels) for the S_1 (right) and S_2 (left) states along the O–H stretching (top figure) and C–O stretching + bending (bottom figure) coordinates. 88

3.8 Diabatic cuts along the normal modes of the 4-dimensional LVC model. Surfaces were fitted to points obtained using RASSCF calculations (shown as grey dots). . . 90
3.9 Calculated absorption spectrum of methanol (green) from vMCG simulations, and an experimentally obtained spectrum[173] (black, dashed). The vMCG spectrum is shifted horizontally by -0.16 eV. 90
3.10 Diabatic cuts of the O–H stretching and C–O stretching + bending modes, constructed from the database points generated during DD-vMCG simulations. All other modes are at $q = 0$. Note: only the lowest 5 electronic states are shown, out of a total of 8 states included in the simulations. 91
3.11 Change over time of the magnitude of the total wavefunction, $|\Psi|^2$ (black line), and of the individual diabatic state densities (coloured lines) upon excitation to the (a) first, (b) second, and (c) third excited state of methanol. 93
3.12 $\tilde{A}$ diabatic state population over time, for different numbers of GWPs, N_{GWP}. Thick black line is the result of the simulation analysed in this chapter. 95
3.13 Expectation values of the centres of the individual GWPs for excitations to 3 different initial states. Line widths are proportional to the GGP and the dashed circles indicate regions of the coordinate space corresponding to different photoproducts. Coordinates are given in units of mass-frequency weighted normal modes. 96
3.14 Integrated flux over total simulation time (representing the probability of dissociation) for three initial excitations. Left panels: O–H dissociation. Right panels: C–O dissociation. 100
3.15 Individual GWP trajectories plotted in the momentum space spanned by the C–O stretching and C–O–H angle bending coordinates. 101
3.16 First generation reaction scheme after methanol photodissociation. Light blue arrows are photodissociation steps, while dark blue arrows indicate radical reactions. . 102

4.1 A general solvation model, containing the solute and the environment as separate "layers" of the model, combining to form the overall solvated system. 110
4.2 The PCM solvation model, showing the molecular cavity surrounded by a continuum potential. 111
4.3 The QM/MM solvation model. The individual solvent molecules are described using MM, whereas the central molecule is a solute modelled using a QM level of theory. 113

4.4 The ONIOM solvation model. The central molecule (sol) is described with a high-level theory, the intermediate layer (env') with a lower level and the largest layer (env) uses the lowest level of theory. The green dots represent an approximate way to model a molecular environment, e.g. point charges. 114
4.5 Equilibrium structure of Uracil, with atomic labels. 118
4.6 Summary of the hierarchy of approximations of the vMCG ansatz, and their EoMs. 125
4.7 (a) potential energy surfaces of the 2 modes included in the model system. (b) time-dependent grid populations (bold line: $t = 0$) of an exact wavepacket propagation, used as a reference for the hybrid calculations (c) and (d) in which one mode is described by 2 fully coupled GWPs, as per the vMCG method, and the other mode uses the saddleMCG approximation. 127
4.8 Flowchart of the AQUATICS procedure. 129
4.9 Multi-layer tree structure for the ML-MCTDH dynamics of Uracil on the vibronic coupling Hamiltonian. The numbers on the last layer correspond to the number of primitive harmonic oscillator DVR functions used. The red box indicates a branch containing the 10 system modes, and mode Q_{26}. 134
4.10 Calculated population transfer in Ura: ML-MCTDH, converged (black) and unconverged (dark green), ML-G-MCTDH, converged (light green) and vMCG (red). The computational effort is compared in the inset. 138
4.11 Calculated population transfer in Ura: ML-MCTDH benchmark (black, dashed), vMCG with 61 GWPs (dark green), vMCG with three sets of GWPs (ten modes and 61 GWPs in each) (light green), vMCG with two sets of GWPs (ten modes with 61 GWPs and 20 modes with 41 GWPs) (red), and vMCG with only the ten most important normal modes included (purple). 139
4.12 Calculated population transfer in Ura: fully variational vMCG benchmark (black, dashed), hybrid vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical ones (dark green), hybrid vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical + saddle point ones (light green), vMCG with all classical trajectories (red), vMCG with all classical trajectories and using the saddle point approximation (purple). 141
4.13 Calculated population transfer in Ura: vMCG benchmark (black, dashed), TSH with 500 trajectories (dark green), and TSH with 1000 trajectories (light green). . . 143

4.14 Calculated population transfer in Ura: vMCG benchmark (black, dashed), DD-vMCG with a single-point database, and no updating (dark green), and DD-vMCG with *on-the-fly* updating of the database (light green). 144
4.15 Potentials of the $n\pi^*$ (solid) and $\pi\pi^*$ (dashed) states of Ura after DD-vMCG simulations, both with and without on-the-fly updating. FC point located at $x = 0$. *Top*: cut showing the MECI (marked by a purple dot). *Bottom*: non-totally symmetric normal modes 4 and 15. Black arrows show the nuclear motion associated with the modes. 145
4.16 Calculated population transfer in Ura: fully variational DD-vMCG benchmark (black, dashed), hybrid DD-vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical ones (dark green), hybrid DD-vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical + saddle point ones (light green), DD-vMCG with all classical trajectories (red), DD-vMCG with all classical trajectories and using the saddle point approximation (purple). . . 146
4.17 Energy difference between $\pi\pi^*$ and $n\pi^*$ states of Ura calculated for 100 snapshots (with $N^{solv} = 20$) taken at 100 fs intervals of a 1 ns MD simulation. Dotted line shows energy difference of the states without any solvent present. 149
4.18 Calculated population transfer of Ura + 20 H_2O: hybrid DD-vMCG with 21 GWPs following variational trajectories and 11 GWPs following classical ones (black, dashed), and solvated calculations starting from different initial configurations sampled at 200 ps intervals (coloured lines). Boltzmann averaged density is shown in orange. . . . 150
4.19 Calculated population transfer of Ura + 4 H_2O: hybrid DD-vMCG with 21 GWPs following variational trajectories and 11 GWPs following classical ones (black, dashed), and solvated calculations with increasing number of solvent GWPs (coloured lines). 152

5.1 Set of carbonyl molecules chosen for DD-vMCG analysis and comparison. 160
5.2 Workflow of a DD-vMCG simulation. 163
5.3 *Left:* Singlet electronic excited state energies and characters of formaldehyde, formic acid, and formamide. *Right:* Singlet and triplet states of formaldehyde and thioformaldehyde. Energies computed with XDW-CASPT2//aug-cc-pVDZ (for thioformaldehyde, CASPT2 was used), implemented in OpenMolcas.[192] "g.s." = ground state, "Ryd." = Rydberg orbital. 166
5.4 Graphical representation of the separation of 3 clusters of similarity in a dataset. . 173

5.5 Normal, Lorenz, and Maxwell probability distributions fitted to (from left to right) energies, gradients and couplings (all in atomic units). All data produced by DD-vMCG calculations of formaldehyde, with $N_{rec} = 4420$. 176

5.6 Pseudo code for automated calculation of product branching ratios. `len(X)` returns the length of `X`, i.e. number of geometries if `X` is a trajectory, and number of clusters if `X` = `C`. 184

5.7 Predicted vs actual adiabatic energies (in eV) of database records not included in the neural net training phase. The mean (μ) and standard deviation (σ) of the error is given in the top left corner. 188

5.8 *Top:* Ground (blue) and second excited (green) state of formic acid along the O–H stretch normal mode (units are mass-frequency weighted) after 4 consecutive DD-vMCG simulations. The number of GWPs, N_{GWP}, total simulation time, t_f, and the resulting number of database records, N_{rec}, are given in each panel. *Bottom:* Decay of the population density of the initially excited state (green in top panel), during the first 50 fs. 189

5.9 SOC strengths between selected Singlet-Triplet state pairs of formaldehyde and thioformaldehyde, at their respective FC geometries (see Figure 5.3 for state characters). 190

5.10 Diabatic state populations CASPT2(12,10) level of theory. Inset showing close-up of populations. 191

5.11 *Left:* Expectation values of C–S stretching mode in thioformaldehyde, centre of wavepacket shown as blue line with its width represented by the shaded area. *Right:* PES along C–S stretch, with inset showing variation of S_1-T_2 energy gap, calculated at CASPT2(12,10) level of theory. 192

5.12 Pruning of the (from top to bottom) energies, gradients, and couplings in a QC database. *Left:* individual database entries (all in atomic units). *Right:* probability distributions of Figure 5.5, and cut-off defined by $n \cdot \mathrm{Var}(x)$. Data shown is from formaldehyde calculations. 194

5.13 Automatic computation of branching ratios, illustrated by trajectories obtained after $\pi\pi^*$ excitation of formic acid. Average final geometry of each cluster is shown, with dotted line indicating the dissociating bond. 195

5.14 Diabatic state populations of $\pi\pi^*$, $n\pi^*$, and ground state (g.s.) for three carbonyl molecules during the first 100 fs of DD-vMCG dynamics. The total density is also plotted (thick black line). 197

5.15 ISOMAP and DBSCAN results for three carbonyl molecules. *Left:* trajectories in reduced coordinates. *Right:* product yields and corresponding average structures. Also given are the EPS parameters used in DBSCAN, and the number of clusters, N_C, and noisy points, N_{noise}, identified. 199
5.16 Integrated flux measured along key normal modes for three carbonyl molecules. See Appendix D, section D.6 for mode descriptions. 203

A.1 The CASPT2 reference weights of each electronic state along the O-H stretching coordinate. 255
A.2 PES of the diabatic $\tilde{F}$ state. The local minimum is indicated, along with the corresponding geometry. 255

D.1 Orbitals included in the (12,10) active space of thioformaldehyde. Red and light blue orbitals are initially doubly occupied, whereas green and dark blue orbitals are unoccupied. 273
D.2 Probability distributions fitted to (from left to right) energies, gradients and couplings (all in atomic units) for (a) Thioformaldehyde, (b) Formic acid, and (c) Formamide. 274
D.3 Clustered trajectories of thioformaldehyde, along with the average structure of each cluster, using both DBSCAN and HDBSCAN. In parentheses are the RMSD values of the HDBSCAN structures, compared to the average DBSCAN geometry. 276
D.4 Comparison of clustering (rows) and dimensionality reduction (columns) algorithms, applied to 37 formaldehyde trajectories. 277
D.5 ISOMAP and DBSCAN results for methanol simulations. *Left:* trajectories in reduced coordinates. *Right:* product yields and corresponding average structures. Also given are the EPS parameters used in DBSCAN, and the number of clusters, N_C, and noisy points, N_{noise}, identified. 279
D.6 Correlation plots for reduced coordinates of formic acid. *Left*: RC1 against the C-OH bond distance. *Right*: RC2 against the expectation value of normal mode Q_2. 283

List of Tables

3.1 Summary of experimentally observed branching ratios of methanol at different excitation energies in the gas-phase, and estimated branching ratios in UV irradiated ice. 68

3.2 Vertical excitation energies (eV) and oscillator strengths (unitless) of the first 7 excited states of methanol, obtained using RASSCF(12,2+8+2)[1,1]//aug-cc-pVDZ. 74

3.3 Vertical excitation energies of the lowest lying electronic states of methanol (in eV), calculated using the EOM-CCSD, CASSCF(12,12), CASPT2(12,12) and RASSCF (12,2+8+2)[1,1] methods, compared to experimental values. The aug-cc-pVDZ basis set was used for all calculations. The shaded column contains the data used in the dynamics simulations. 82

3.4 Electronic character of the first 3 diabatic states, along the O–H and C–O dissociation channels. Dissociative states are underlined. (g.s. = electronic ground state) . 87

3.5 Relative yields of photoproducts after excitation to the three lowest lying electronic states of methanol. Ratios are determined from analysis of quantum trajectories, weighted by their GGPs. 95

4.1 Computational effort of the reported Quantum Dynamics simulations in terms of CPU time (Dell PowerEdge R450 Server with Intel Xeon Silver 4316 2.3GHz 20Core CPUs and 384GB RAM). 137

4.2 Total energy (kJ mol^{-1}) of solvated Ura for different initial solvent configurations, calculated using the QMD-FF of the ground state. Normalised Boltzmann factors are calculated in the last column, with $T = 298$K. 151

5.1 Bertz complexity index (C_T). 160

5.2 Comparison of dimensionality reduction methods. 178

5.3 Comparison of clustering methods. 180

5.4 Overview of tasks contained in the DD-vMCG workflow. 206

A.1 Cartesian coordinates (in Å) for methanol used in this work (optimised using CCSD// aug-cc-pVDZ). 254
A.2 Frequencies (in cm^{-1}) of the 12 normal modes of Methanol (CCSD//aug-cc-pVDZ). 254

B.1 Cartesian coordinates (in Å) for Ura used in this work. 257
B.2 Normal modes[a] of Uracil for the ground electronic state, point group = C_s. A' modes are in-plane whereas the A" modes correspond to out-of-plane motions. Frequencies in cm^{-1} 258
B.3 Frequencies (A′ = in-plane, A″ = out-of-plane) and linear parameters of the 2-state vibronic coupling Hamiltonian of Uracil. n in $\kappa^{(n)}$ and $\lambda^{(n,n)}$ refers to the electronic state where $1 = n\pi^*$ and $2 = \pi\pi^*$. Highlighted in red are the 10 system modes used in the 10-D + 20-D partition of the nuclear basis. All units are eV. 259
B.4 Intermolecular parameters obtained for Ura in different electronic states: σ (Å) and ϵ (kJ/mol) are the Lennard-Jones parameters. q (e^-) are the partial charges. Atom types are shown in Figure 4.5. 260

D.1 Sensitivity analysis with respect to EPS. N_C are the number of clusters generated, and N_{noise} the number of unclustered trajectories. 275
D.2 Cartesian coordinates (in Å) for formaldehyde (optimised using CCSD//aug-cc-pVDZ). 280
D.3 Frequencies (in cm^{-1}) of formaldehyde (CCSD//aug-cc-pVDZ). 280
D.4 Cartesian coordinates (in Å) for thioformaldehyde (optimised using CCSD//aug-cc-pVDZ). 280
D.5 Frequencies (in cm^{-1}) of thioformaldehyde (CCSD//aug-cc-pVDZ). 280
D.6 Cartesian coordinates (in Å) for formic acid (optimised using CCSD//aug-cc-pVDZ). 281
D.7 Frequencies (in cm^{-1}) of formic acid (CCSD//aug-cc-pVDZ). 281
D.8 Cartesian coordinates (in Å) for formamide (optimised using CCSD//aug-cc-pVDZ). 282
D.9 Frequencies (in cm^{-1}) of formamide (CCSD//aug-cc-pVDZ). 282

Glossary

(H)DBSCAN (Hierarchical) Density-Based Spatial Clustering of Applications with Noise. 180

ADT Adiabatic-to-diabatic transformation matrix. 42

AI Artificial Intelligence. 207, 208

AIMD Ab Initio Molecular Dynamics. 37, 38

AIMS Ab-Initio Multiple Spawning (method). 53, 56, 124

BOA Born-Oppenheimer approximation. 38–41, 49, 62, 67

CAP Complex Absorbing Potential. 75, 76, 92, 167, 170, 190, 196, 197

CASPT2 Complete Active Space Perturbation Theory (2nd order). 33, 34, 81, 83, 85, 86, 165, 168, 191

CASSCF Complete Active Space Self-Consistent Field (method). 32–35, 70–72, 81, 85, 165, 211

CC Coupled Cluster (theory). 29, 30

CCS Coupled Coherent States, a semi-classical dynamics method. 53, 124

CCSD Coupled Cluster, Singles and Doubles. 29, 71

CI Configuration interaction. 27–32, 50, 70, 87

CIS Configuration Interaction Singles. 28

CISD Configuration Interaction Singles and Doubles. 28

clMCG classical Multi-Configurational Gaussian method where higher order terms of the EoMs relating to the Gaussian parameters are ignored. 124, 128

CoIn Conical Intersection. 16, 23, 41, 43, 44, 46, 49, 62, 87, 119, 168, 193

CSF Configuration State Function. 32, 33, 87

DD-vMCG Direct-Dynamics Variational Multi-Configurational Gaussian (method). An *on-the-fly* quantum dynamics method based on vMCG. 56–60, 63, 72–75, 77, 91, 94, 103–105, 117, 130, 136, 137, 144–147, 159, 161, 165, 166, 169–171, 177, 186, 189, 191, 195, 204, 205, 207, 210–213

DFT Density Functional Theory. 35, 36, 60, 61, 64

DMAP Diffusion Map. 178, 179

DoF Degree(s) of freedom. 38, 50, 52, 111, 112, 117, 120, 123, 130, 140, 151, 153, 154, 164, 172, 198, 200, 202, 207, 283

DVR Discrete variable representation. 51

EoM Equation of Motion. xv, 37, 38, 54, 55, 59, 60, 116, 117, 122, 123, 125, 126, 128, 135, 136, 153

EOM-CCSD Equation of Motion Coupled-Cluster Singles and Doubles. 30, 81, 83, 85

FC Franck-Condon. 48, 57, 75–77, 81, 83, 86, 92, 96, 130, 134, 143–145, 164, 165, 176, 186, 192, 196

FCI Full Configuration Interaction. 28, 29, 31, 32, 34

FF Force Field. 112, 116, 117, 120, 121, 129, 132, 148, 153–155

FFQD Force Field Quantum Dynamics. 108, 116–118, 120, 122, 128–130, 132, 139, 147–149, 151, 153–155

FMS Full Multiple Spawning, a Gaussian-based dynamics method, related to AIMS. 53, 124

FSSH Fewest Switches Surface Hopping method. 60, 61

GGP Gross Gaussian Population. 80, 95, 99, 172, 183, 185, 195

***G*-MCTDH** Gaussian Multi-Configuration Time Dependant Hartree (method) where some of the SPFs of the MCTDH method are replaced by Gaussian functions. 51, 116, 122, 123, 132

GUI Graphical User Interface. 166, 170

GWP Gaussian wavepacket. 53–55, 57, 58, 74–76, 80, 94, 95, 98, 117, 122–126, 128–130, 133, 135, 136, 141, 148, 151, 167, 169, 195, 196, 275

HF Hartree-Fock. 26–31, 33

ISC Intersystem Crossing. 43, 44, 190–193, 196, 205

ISOMAP Isometric Feature Mapping. 178, 179, 183

LCAO Linear combination of atomic orbitals. 26

LHA Local harmonic approximation. 56, 57, 75, 124, 125, 128, 130, 132, 164, 165, 189, 196

LVC Linear Vibronic Coupling. 47, 48, 52, 56, 73, 89, 124

MCE Multi-Configurational Ehrenfest (method). 53

MCSCF Multiconfiguration Self-Consistent Field (theory). 19, 31–33, 62

MCTDH Multi-Configuration Time Dependant Hartree (method). xvi, 50–52, 55, 56, 60, 63, 78, 80, 105, 133, 138, 139, 169, 186, 187, 189, 211, 212

MD Molecular Dynamics. 10, 11, 116, 117, 129, 147–149, 151

MDS Multi-Dimensional Scaling. 178, 179

MECI Minimum Energy Conical Intersection. 144, 145

ML-*G*-MCTDH Multi-Layer Gaussian Multi-Configuration Time Dependant Hartree (method), multi-layer version of *G*-MCTDH. 52, 133, 138, 139

ML-MCTDH Multi-Layer Multi-Configurational Time Dependent Hartree. 132–135, 138, 141

MM Molecular Mechanics. 112, 117, 120, 128, 130, 132, 148, 153, 155

MO Molecular-Orbital. 26, 31, 32

MP2 Møller-Plesset second order perturbation theory (method). 33, 34

MRCI Multi-Reference Configuration Interaction (theory). 30, 31

NAC Non-Adiabatic Coupling. 36, 41–43, 53, 57, 61, 96, 117, 121, 129, 131, 155

ONIOM Our own N-layered Integrated molecular Orbital and molecular Mechanics (hybrid method). 113

PCA Principle Component Analysis. 178, 179

PCM Polarisable Continuum Model. 110, 111, 115

PES Potential Energy Surface. 14–17, 23, 25, 37, 40, 42, 44, 58–61, 69, 70, 85–87, 89, 96, 99, 103, 104, 116, 117, 120, 121, 125, 130, 144, 164, 166–169, 171, 172, 174, 175, 186, 187, 189, 196, 207, 208, 212

QC Quantum Chemistry. 56, 58–60, 75, 117, 120, 121, 132, 168, 174, 187, 189, 207, 213

QD Quantum Dynamics. 10–12, 17, 19, 33, 36, 37, 44, 48–50, 53, 56, 59, 60, 63, 67, 70, 73, 76–78, 80, 81, 83, 85, 86, 89, 91, 103–105, 108, 116–119, 132, 133, 141, 153, 157–162, 164, 170, 173–175, 177, 178, 182, 187, 191, 197, 203–205, 207, 208, 212, 214

QM Quantum Mechanics. 6–8, 10, 11, 19, 77, 111–114, 117

QM/MM Quantum Mechanics / Molecular Mechanics (hybrid method). vii, 112, 113, 116, 153, 155

QMD-FF Quantum-mechanically derived force field. 120, 121, 129, 131–133, 136, 140, 143, 148, 150, 153–155

QVC Quadratic Vibronic Coupling. 47

RASSCF Restricted Active Space Self-Consistent Field (method). 34, 35, 71–73, 81, 85–87, 89, 90, 104

RIC Redundant Internal Coordinates. 121

RMSD Root-Mean-Square Deviation. 183

saddleMCG classical Multi-Configurational Gaussian method where the saddle-point approximation is used for the evaluation of the electronic Hamiltonian terms. 125, 126, 128

SCF Self-consistent field (method). 26, 32

SD Slater Determinant. 25, 26, 30

SOC Spin-Orbit Coupling. 43, 44, 190–193

SPF Single-particle function(s). xvi, 51–53, 122

TD-DFT Time-Dependent Density Functional Theory. 35, 36, 70, 168

TDSE Time-Dependent Schrödinger Equation. 11, 37–40, 48, 50, 211

TISE Time-Independent Schrödinger Equation. 24, 26, 29, 60

TSH Trajectory Surface Hopping method. 59–61, 116, 142, 170, 182

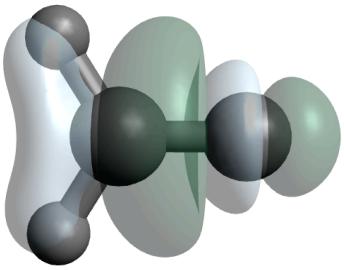
UV Ultra-Violet. vi, xii, 67–70, 102, 105, 119

vMCG Variational Multi-Configurational Gaussian (method). xv, 51, 53, 55, 56, 60, 63, 73, 74, 80, 89, 117, 121–126, 128, 132–142, 144–146, 148, 153, 211, 213

XDW-CASPT2 Extended Dynamically Weighted Complete Active Space Perturbation Theory (2nd order). 165, 196

CHAPTER 1

Introduction



Ceci n'est pas une molécule

1.1 Aims & Outline

This thesis is written after three and a half years of computational research in the field of non-adiabatic quantum dynamics (the exact meaning of these terms is explained in due course) for the study of photophysical and photochemical reactions of small organic molecules. This includes both the use of existing computational approaches, as well as the development of novel techniques to address specific shortcomings of the former. The first aim of this thesis is to provide an accurate account of the motivations, methodologies, and conclusions of the various topics that are discussed. In the big picture, this research demonstrates how and why quantum dynamics simulations can and should be used to study and understand fundamental processes in molecules, such as light-matter interactions.

Specifically, such simulations require the computation of the time evolution of a quantum-mechanical wavefunction, which remains a formidable challenge to this day. It will become clear that methods able to achieve this goal are still actively being developed, and difficult to use. With this in mind, a secondary aim of this thesis is to provide a clear description of the problems involved, highlighting challenging theoretical aspects, discussing suitable examples, and suggesting possible improvements. As such, this work should provide relevant information and guidance to those who wish to study problems resembling the ones discussed here.

Some chapters contain previously published work, but rather than simply repeating what was written in those articles, the opportunity was taken to include more detail and contextual information. Attention is also given to avenues that were not successful, and to how such failures were ultimately resolved.

There are a total of 6 chapters, which are briefly outlined:

1. **Introduction**: This chapter introduces the general ideas underpinning all subsequent chapters. A discussion of computer models and the role of mathematics in the formulation of scientific theories is used to situate and motivate the field of molecular dynamics. An important distinction is made between classical and quantum dynamics, the latter being ideally suited to study photoinduced processes. These are briefly introduced, along with important theoretical considerations. Lastly, the role of experimental results is considered in the context of computational research.

2. **Theory**: The background and theoretical foundations of all relevant computational methods are described. Both electronic structure and nuclear dynamics methods are discussed, with a focus on the latter. The aim here is to define and explain the specific terminology that is used throughout the rest of the thesis.

3. **Photodissociation of Methanol**: A detailed study of the photodissociation reactions of methanol is presented. Understanding these photochemical pathways is atmospherically and astrochemically interesting, but the focus of this chapter lies more on the approach taken to achieve this detailed description, allowing one to model dissociation reactions in general. This work should thus be seen a case-study for the broader application of the methods discussed.

4. **Force Field Quantum Dynamics**: The title of this chapter refers to a novel methodology that addresses the need to describe the effect of environment interactions during some photophysical/chemical phenomena. A procedure that can explicitly include these effects is put forward, along with thorough benchmarking results which demonstrate the validity of the proposed approach. Additionally, the performance of different approximations is evaluated, with the aim of more precisely balancing the trade-off between accuracy and computational cost, and to push the limits of accessible system sizes.

5. **Towards automated quantum dynamics workflows**: Several quantum dynamics simulations of a set of small molecules are used to critically assess the usability, efficiency, and overall performance of a specific method. The emphasis of this analysis lies not in the algorithm itself, but rather the workflow that is required to successfully use it. In that context, each step is evaluated and broken down with the aim of suggesting automated procedures to address current bottlenecks. Some comments are necessarily method-specific, but some considerations also apply more generally.

6. **Conclusions**: The thesis concludes with a brief chapter summarising the contents of the preceding chapters and their main conclusions before giving some general remarks on the lessons learned, insights had, and future directions to be explored.

1.2 Computational modelling

By way of general introduction to this thesis, a broad overview is given of the practice of using models to solve problems. All topics covered in this thesis are examples of how theoretical models can be translated into computational algorithms, and used to gain a deeper and more detailed understanding of particular phenomena. The applications, benefits, challenges and limitations of this approach will be covered and discussed in a general sense, as well as with relation to the specific use-cases considered in this work.

The word "modelling" is defined here as the use of mathematical constructions to describe parts of our world. In more technical terms, the defining characteristic of a model is an *isomorphism* between its mathematical expression and (perceived) reality, defined in whatever way one prefers. This simply means that a mapping must exist between some elements of the model, and the reality it aims to describe. Although very common in this day and age, in applications such as weather predictions or monitoring the stock market, it is surely an astonishing fact that it is possible to apply abstract mathematical concepts and gain such extremely useful knowledge and insight about our complex physical and socioeconomic world. Of course, all models are per definition lacking in terms of what they are able to describe, leading to the saying by statistician George Box that "all models are wrong, some are useful". Thus, an important question needs to be raised; when can a model be considered useful?

One point of view is that if a model is able to predict certain phenomena, without necessarily explaining why they occur, its usefulness is demonstrated. In this case, the underlying theory on which the model is based need not be correct or provide any kind of reason for the predicted outcome. This is an *instrumentalist* approach to scientific theories since the use of theoretical models is seen merely as an instrument to make useful predictions, having no connection to the underlying reality responsible for the observed phenomena.

An opposing view is that a good model should have explanatory power, since the possibility of getting the right answer for the wrong reasons is not acceptable because it doesn't lead to further progress.[1] In the natural sciences specifically, the latter view is often preferred. Modelling is widely used in nearly all areas of physics, chemistry, engineering, and biology in order to try to explain how and/or why certain physical phenomena occur. In order to do this however, it is necessary to have a solid

underlying theory that can be tested and that can provide good explanations. On this point there are some interesting philosophical questions that arise related to the nature of the scientific method itself: namely, the question of how scientific theories are created; pure unbiased observation, or starting from some initial presupposition about the specific subject of study? Naively, one might prefer the former of the two alternatives, and indeed, this view has been very popular in the past after having been initially proposed by Sir Francis Bacon in the 17^{th} century. It suggests that all scientific theories are formulated so as to fit some observation. However, a more modern analysis of the way science progresses has discarded this view, claiming that making observations is only possible with some form of pre-existing theory about the world to make sense of the results. This is the scientific method according to the 20^{th} century philosopher Karl Popper.[2] He rejected the empiricist view that all knowledge is obtained from the senses and emphasised the importance of the interpretation of observations, which necessarily derives from certain beliefs, held by the observer, about reality.[3]

This then is the role of scientific theories; to provide us with a framework to interpret the observations that we make, and derive knowledge from them. In this Popperian view, the idea of *falsification* becomes of central importance . A theory can never be proven right by an observation, but it should be able to be disproved by a contradictory observation to be considered valid. Hence, science can only advance by replacing wrong theories with new ones, until they too are falsified, rejected and replaced. Such a view allows for an infinite growth of human knowledge (constrained only by the laws of physics), through an ever repeating sequence of making mistakes and correcting one's errors.

This philosophical perspective serves to motivate the use of theoretical models in chemistry, and give context to later discussions of specific theories and their relation to experimental observations. Another part of this context is provided in the next section, concerned with the link between scientific theories and mathematics. In the physical sciences, most, if not all, theories are formulated using mathematical foundations since this allows for a robust and unambiguous language in which to define the problem one is trying to solve. Therefore, before introducing the specific class of problems that are the main subject of this thesis, a discussion of the use of mathematics in physical theories/models is given.

1.2.1 Role of mathematics

In 1960, the theoretical physicist Eugene Wigner wrote an article titled "The Unreasonable Effectiveness of Mathematics in the Natural Sciences".[4] In it, he discusses the role of mathematics in the development of our understanding of the physical world, and emphasises that we ought to be grateful for how powerful and successful our theories have been at describing our perceived surroundings. As Galileo put it: the book of nature is written in the language of mathematics. We do not, however, understand *why* our mathematical tools are so well suited to the development of theories concerning our environment. This question is a profound philosophical one, and will not be further explored here. It is simply taken for granted that the use of certain mathematical concepts is appropriate for the description of physical processes.

Specifically, the role of mathematics in the development of the theory of quantum mechanics (QM) has made it possible to not only describe but also understand, to some extent, reality at the scale of atoms and molecules. Quantum theory is at present one of the most fundamental, deep and successful theories ever developed thanks to its unbelievably accurate predictive powers, along with the physical insights it offers about the nature of reality. Although there are, to this day, ongoing debates about which interpretation of the theory most closely matches reality, or provides the best explanations for the phenomena it purports to describe, using the theory in practice always comes down to the same thing; performing mathematical operations (sometimes called the "shut-up-and-calculate interpretation"). As the Oxford University professor David Manolopoulos remarks in his 2$^{\text{nd}}$ year quantum chemistry course: "Quantum mechanics is just matrix algebra". Adopting such a view can be useful because of the often difficult intuitive interpretation of the theory, but if that was all there was to it, a simple computer able to multiply matrices together would be able to figure out everything there was to know. In addition to computation, a certain level of creativity is also needed, to know when and how to apply the ideas of quantum theory. This is the job of research scientists working in related fields.

The relationship between mathematics and QM has been a symbiotic one in the sense that thanks to it, important developments have been made in both subjects. On the one hand, the theory of QM could not have been developed without the use of purely mathematical concepts. One reason for this is that the microscopic world is

very different from the macroscopic one experienced by humans, and hence is much less intuitive. Perhaps the best example to illustrate this point is the fact that in classical mechanics, which can accurately describe the movement of macroscopic objects, one can measure multiple properties of these objects with arbitrary precision, simultaneously. However, when considering the movement of electrons around a nucleus for example, the measurement of one observable – say, the electron's momentum – limits the precision with which another observable can be measured (this is the *Heisenberg Uncertainty Principle*).

Explaining these types of differences between the classical and quantum world called for a radically different theory, i.e. QM. This theory makes use of mathematical ideas which are not intuitively related to the physical world, a fact which makes the role of mathematics even more important. In his paper, Wigner specifically mentions how the counter-intuitive reality implied by the uncertainty principle can be described by simply replacing the variables for position and momentum by matrices of the same. This mathematical trick, proposed by Born, Jordan, and Heisenberg, allowed the theory to make very precise predictions that could not be matched by rival theories, hence its relevance today.

Even when considering one of the most foundational elements of QM one encounters the use of abstract mathematical ideas with no physical analogue. The first postulate of QM requires the state of any system to be fully described by a wavefunction. For a system at position x and time t, this wavefunction is given by a state vector, $|\Psi(x,t)\rangle$, in a Hilbert space (note the use of Dirac notation $|\cdot\rangle$ for state vectors). This is a complex space and all vectors belonging to it are therefore also complex, thus implying a complex-valued wavefunction. Surely the use of complex numbers to describe "real" physical systems is not an intuitive approach, since the natural world does not suggest their existence (hence the term *imaginary* numbers).

Changing perspective, the development of quantum theory has also had major influences on those branches of mathematics that are relevant to it. An example is the use of operator algebra, introduced by von Neumann. Operators, being a central concept in QM, are used to describe measurable quantities by operating on the state vectors in the Hilbert space. As a result of their importance in describing quantum-mechanical observables, the theory of operator algebras is now a well developed mathematical discipline. Another, perhaps less abstract, example is the development of wave mechanics

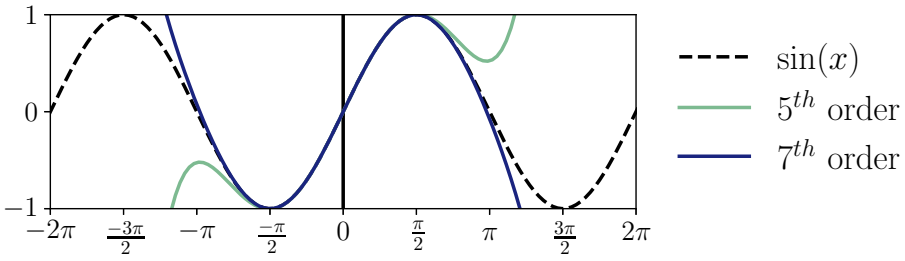


Figure 1.1: Sine function (black, dashed) and its Taylor expansion up to 5^{th} (light green) and 7^{th} order (dark blue), expanded around $x = 0$.

by Erwin Schrödinger, leading to the central equation in QM, able to describe the motion of quantum particles as waves evolving in time. This profound insight sparked an increased interest in the developments of the theory of partial differential equations, a very broad and important field of applied mathematics.[5] The next section introduces how Schrödinger's equation can be used to model the motion of atomic and molecular systems in practice.

Before that, one more important aspect of mathematics is highlighted, and that is its use in the approximation of exact expressions. This is extremely valuable when dealing with "real-life" applications since exact expressions are rarely known. An example that is widely used in theoretical chemistry is the Taylor expansion, the definition of which is given below:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^n(a)}{n!}(x-a) \quad . \tag{1.1}$$

A Taylor expansion expresses a function $f(x)$ as an infinite sum, where each term, n, contains the n^{th} derivative of $f(a)$, given by $f^n(a)$, expanded around a single point a. This way of writing a function allows for an approximation of the exact result by choosing an appropriate point a and truncating the expression up to a desired order of differentiation. Consider for example $f(x) = \sin(x)$ and its Taylor expansion up to 5^{th} and 7^{th} order, given in Figure 1.1. Clearly, the expansions are equivalent to $f(x)$ up to a certain point away from $x = 0$, and the higher the order, the wider the range in which the approximation holds. The use of this method, as well as other ways to approximate exact representations, will be shown to be important in later discussions.

To conclude this section on mathematics, the following assertion is made: knowing a mathematical formula that describes a particular process, does not mean understanding said process. If one adopts the instrumentalist view of science as described

previously, this statement does not matter and one will happily use the formula to make predictions and use it in other ways to serve a particular goal. However, if one is interested in the reality behind the process, simply knowing of a correct way to describe it mathematically is not sufficient. Now there is a desire to look deeper and think about how and why a particular mathematical formulation could explain the process.

Here, the role of computational algorithms comes in. More specifically, the use of simulations allows one to derive knowledge from performing mechanical computations based on certain rules which themselves don't necessarily reveal anything about the system they describe. In a simulation, the final result is not obtained directly, but rather through a process of sequential intermediate steps. These steps are not very telling in isolation, but taken together, they can reveal the underlying details of the transformation from inputs to outputs. This shall indeed be shown in the chapters that follow, where mathematical formalisms are translated into algorithms able to simulate fundamental chemical and physical phenomena.

1.2.2 Molecular modelling: classical vs quantum

Since this thesis is presented in the field of theoretical chemistry, the scope of the discussion of computational models will be narrowed to molecular models in particular. In this context, the role of mathematics has to be re-evaluated. Even though the underlying principles of chemistry are embedded in quantum mechanics, thermodynamics, and statistical mechanics – all well described mathematically – there is a higher level of understanding that emerges from these fundamental concepts. Examples include the idea of chemical bonds, functional groups, aromaticity, resonance etc. Although these lie at the core of chemical language, and allow for an intuitive, pictorial understanding of chemistry, they remain empirical constructs that are not firmly rooted in a mathematical framework. The relationship between these two levels is not clear, and it may seem counter-intuitive to try and extend mathematical paradigms to emergent chemical properties.[6] This thesis will argue however, that many important phenomena can be understood through a bottom-up approach, where the modelling of fundamental physical processes is able to provide insights into complex chemical behaviour.

Before explaining how this is achieved, an important distinction is first made be-

tween empirical and *ab initio* models. The former uses experimental/calculated data as input, the latter derives from first principles (meaning rigorously defined mathematical statements that cannot be deduced from any other axioms). All the work presented in this thesis is the result of *ab initio* modelling, however, chapter 4 introduces an approach that combines features of both. In the following, the difference between the two is explained in terms of molecular dynamics methods.

Both empirical and *ab initio* models are widely used to calculate various properties of molecular systems, whether time dependent or independent. One specific type of time dependent model, classical Molecular Dynamics (MD), uses empirical data in order to simulate the time evolution of potentially very large molecular assemblies according to Newton's equations of motion. Force fields are used to describe the interaction between atoms and molecules, and these are typically parametrised based on experimentally measured or computationally calculated quantities (e.g. bond strengths, expressed as the force constant of a harmonic oscillator between two masses). Since these methods are able to explicitly simulate close to a million atoms, for timescales up to milliseconds, they have proven to offer valuable insights and are becoming ubiquitous in understand many biologically and industrially relevant processes (e.g. protein activity and catalysis).[7,8]

In stark contrast with this class of simulations are molecular Quantum Dynamics (QD) methods. As suggested by the name, these techniques are derived from QM (hence they are *ab initio*) and aim to include the effect of quantum phenomena in the simulation of moving atoms and molecules. Whereas classical MD calculations only explicitly simulate nuclei, and hence only treat nuclear degrees of freedom, QD approaches take into account the role of the electrons as well by explicitly including the electronic structure of the system being studied (calculated using a wide range of electronic structure methods, discussed in the next chapter).

This simple fact means that QD simulations are orders of magnitude more computationally expensive than classical methods and as a result, the system sizes accessible to these methods are orders of magnitude smaller. The obvious upside however is the increased accuracy of QD calculations, which is invaluable when studying phenomena that are inherently quantum mechanical and cannot be approximated classically. One such phenomenon that is quintessential in this thesis is the notion of *non-adiabaticity*. This term is further explained in the next chapter but it essentially refers to the nuclei

entering a state of superposition, which has no classical analogue.

The specific characteristics of the two types of molecular simulations discussed above form a spectrum with MD and QD approaches populating opposite ends; one is limited in its accuracy but handles enormous systems, the other strives for precision but is prohibitively expensive.

Non-adiabatic quantum dynamics

In this work, the dynamical simulations that are performed lie at the QD end of the aforementioned spectrum. Specifically, the methods discussed are all in the family of non-adiabatic QD methods.[†] These methods aim to simulate the time evolution of nuclei as accurately as possible, by considering the coupling between nuclear and electronic motion (and hence allowing for the aforementioned superposition of nuclei). In general, two things are needed to describe the time evolution of a system; the state of the system at an initial time, and a set of rules or parameters that define how the system changes over time. In the previous section, the quantum-mechanical definition of a physical state was already given; a time-dependent wavefunction, $|\Psi(t)\rangle$. The rules governing the time evolution, i.e. the dynamics, of this wavefunction are then given by the time-dependent Schrödinger equation (TDSE):

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \hat{H}\Psi(x,t) \quad . \tag{1.2}$$

A few important things should be noted here. Firstly, the wavefunction is now written as a function, not only of time but also of position, given by x. Next, the Hamiltonian operator, $\hat{H}$, corresponds to both the kinetic and potential energy, making it the total energy operator. Finally, note the presence of the imaginary number i, implying a complex wavefunction.

The TDSE plays a central role in QM in general since it describes the evolution of an arbitrary wavefunction, meaning that solving it would provide one with a complete picture of the motion of a quantum-mechanical object. In fact, only three years after publication of Schrödinger's revolutionary paper in 1926,[9] Paul Dirac summarised the situation as follows: "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to

[†]Whenever "QD" is used, it is implied to mean "non-adiabatic QD", unless otherwise stated.

equations much too complicated to be soluble".[10] Although probably not entirely accurate, he was certainly right about the immense practical difficulties preventing the application of this equation to the vast majority of physical systems. Even though it has the form of a first order differential equation, making it a conceptually simple initial value problem, the definition of the initial wavefunction Ψ at $t = 0$ and at a suitable geometry x_0, is already problematic. The choice of $\Psi(x_0, 0)$ can of course have a large impact on the resulting dynamical behaviour. Another big issue is the fact that as the dimensionality of the system grows (i.e. x becomes multi-dimensional), the effort to solve the equation grows exponentially with respect to the dimensions of x. For molecular systems, the dimensionality grows very quickly as one needs 3 degrees of freedom per atom to fully describe a given molecule in Cartesian coordinates (i.e. $3N$ dimensions, with $N =$ nr. of atoms). Overcoming these challenges in order to use the TDSE to understand a range of fundamental processes is the aim of the field of QD. The Theory chapter of this thesis will discuss, in detail, approaches to solve the TDSE for realistic, multi-dimensional molecular systems.

1.3 Photochemistry

Having introduced the concept of molecular modelling and the aims of the field of QD, some attention will now be given to the choice of application for these methods. In theoretical chemistry one is interested in studying molecules and hence all calculations presented in this work are performed on molecular systems, as opposed to model systems. All chosen molecules are relatively small ($\leq$ 12 atoms) since, as previously mentioned, one needs to consider the trade-off between size and accuracy when using QD methods. However, the size of a molecule is certainly not proportional to how important or interesting it is. As will become clear, the chosen molecules are all of importance in contexts such as atmospheric chemistry or astrochemistry. More specifically, the interest lies in their interaction with light.

Incoming electromagnetic radiation absorbed by a molecule causes energetic changes that affect its rotational, vibrational and electronic character, which in turn can lead to some observable change in the molecule. In particular, if a single electron is excited to a higher lying orbital, creating a molecular excited state, the presence of unpaired electrons can drastically affect a molecule's properties. The way in which a molecule

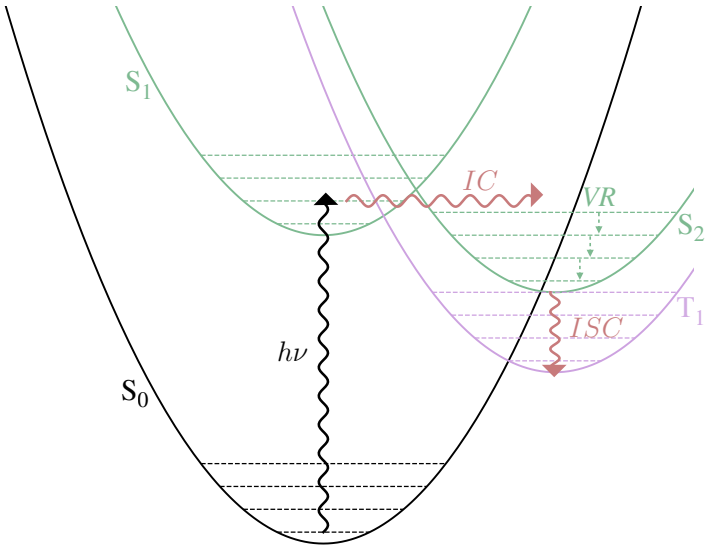


Figure 1.2: Non-radiative photophysical processes induced by photon absorption of energy $h\nu$, illustrated by a system of harmonic oscillators representing different molecular electronic states: singlet ground state (S_0, black), singlet excited states (S_{1-2}, green), and triplet excited state (T_1, purple). Dashed lines represent vibrational energy levels. Red arrows show different non-radiative processes; IC: internal conversion, VR: vibrational relaxation, ISC: intersystem crossing.

is affected by different electronic excitations cannot straightforwardly be generalised because the nature of excited states can vary greatly and can be described using different theoretical frameworks.[11] This fact also makes the calculation of excited states difficult, and hence a number of specialised techniques have been developed to tackle this challenge, discussed in the next chapter.

The possible changes induced by electronic excitation through absorption of a photon are termed photo*physical* processes, and are summarised in a Jablonski diagram (Figure 1.2). Only non-radiative pathways, meaning those that do not produce photons, are shown since those are the ones of interest in this work. There are also two possible radiative pathways a molecule can follow; fluorescence and phosphorescence. The former is the result of a transition between two states with the same spin multiplicity, producing a photon in the process. The latter is the same process but between states of different spin multiplicities. These are not discussed further. The non-radiative analogues of fluorescence and phosphorescence, shown in Figure 1.2, are internal conversion and intersystem crossing, respectively. These phenomena are ubiquitous in the mechanistic explanation of photoinduced processes,[12] and are frequently used in discussions that will follow in later chapters.

It is also possible for a molecule to undergo a chemical reaction/transformation as

a result of the photophysics that takes place after excitation. One then talks about photo*chemistry*. This field of study relates to many modern-day applications like for example technology (solar cells, displays), medicine (therapy, diagnostics) or industry (catalysis).[13] One of the most cited examples of an important photochemical reaction is the isomerisation of retinal, which goes from a *cis* form to an all *trans* configuration upon interaction with visible light, and is responsible for vision in mammalian eyes. Naturally, this relevance in many areas has sparked an increasing interest in the field over recent years. In addition to being of interest for commercial applications, photochemistry is an active academic field as well, because in order to harness the power of photochemical reactions, one must first understand the processes that make them possible in the first place. In this respect, theoretical photochemistry plays a very important role.[14]

1.3.1 Theoretical considerations

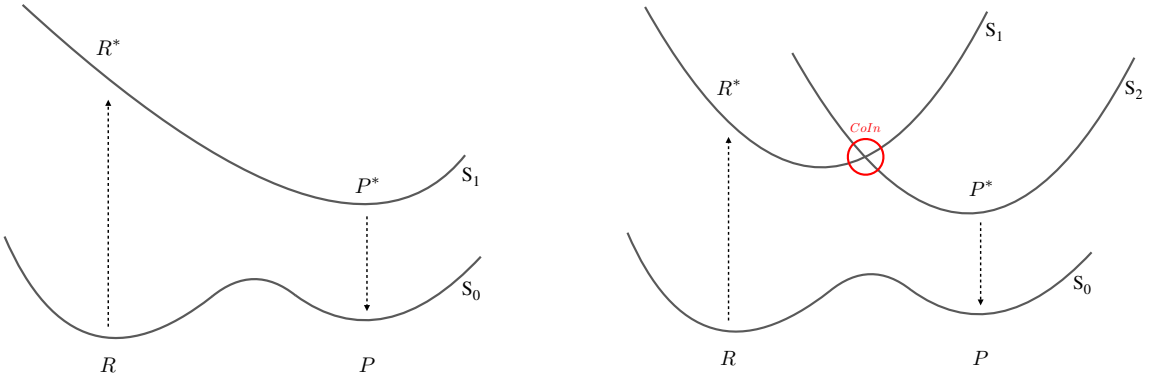
The use of theoretical methods to study photochemical processes can provide mechanistic detail that is often inaccessible to experimental methods. There are, however, several major challenges associated with describing photochemistry computationally. The first such challenge is the fact that many different reaction pathways are potentially accessible when a molecule is excited, all of which need to be described properly. It is therefore necessary to get a sense of the potential energy landscape, which needs to be calculated for different parts of the configuration space of a system, as well as for relevant excited states (which may be a large number). Keeping in mind that most methods in computational chemistry are based on the treatment of the molecular ground state, the scarcity of available theoretical tools also contributes to this challenge.

Next, even if an accurate picture of the energy landscape is obtained, one is faced with the problem of determining what processes are actually possible after photoexcitation, how likely they are, how fast they occur etc. This is where non-adiabatic molecular dynamics simulations come in, but there is also a time-independent way of tackling this problem, which is discussed first. An intuitive picture of the likely dynamics of a system can be obtained from mapping out the potential energy surfaces (PESs) along important coordinates. For example, if a molecule is excited to a certain state, and the minimum energy geometry of that state does not coincide with the

ground state minimum energy structure, one can infer that a configurational change will take place in order to reach the excited state minimum (such a mechanism is responsible for Stokes shifts, for example). Such a qualitative picture is a powerful tool to rationalise photochemical reactions, and by identifying key stationary points along the PESs (minima, saddle points, crossing points between electronic states) and connecting them to each other, one can elucidate mechanisms for complex dynamical processes.[15–17] Applying models such as transition state theory[18,19] or Marcus theory[20] may even allow for quantitative insights into rate constants and excited state lifetimes. Notice the similarity with more traditional descriptions of thermal chemistry, involving reactants, products, and transition states connecting them. However, since one has to characterise the different pathways *a priori*, a potential pitfall is that not all possible photochemical channels are included and hence the complete, multi-dimensional picture may not be revealed.

This problem is especially apparent for photochemical reactions since they involve the complex interplay between multiple electronic excited states. One thing that is always true, by definition, is that a photochemical reaction starts in an excited state but ends with the product(s) in a relaxed ground state. There is therefore necessarily some amount of population transfer between states, leading to de-excitation, either radiatively or not. The question that needs to be answered is thus the following: what drives population transfer between states? At this point a distinction can be made between reactions, based on their particular electronic mechanisms. The first class of reactions are those where a chemical change happens on a single excited state surface. These reactions are called *adiabatic*. Alternatively, population transfer may occur on the reaction path in between the reactants and products, meaning the chemical change happens on multiple electronic states. We then talk about *diabatic* processes.[21] These definitions are slightly outdated (due to the work of Förster in the 1970's) and will be expanded upon more rigorously in the next chapter, but for now, they provide a clear motivation for understanding the exact nature of population transfer processes in photochemical reactions.

More recently, work done in this area over the past few decades has revealed that the vast majority of photoinduced reactions are mediated by regions of the potential energy landscape where two (or more) PESs cross each other. This creates a point of energetic degeneracy between two states, allowing for efficient and radiationless de-



(a) Adiabatic (b) Diabatic

Figure 1.3: Representation of a photochemical reaction, going from Reactants (R) $\rightarrow$ Products (P) via (a) a single excited state surface and (b) internal conversion between two states, mediated by a CoIn. R^* and P^* are the reactants and products in their respective excited states and the label S_n denotes singlet electronic states, with S_0 being the ground state.

excitation. Such crossing regions are termed Conical Intersections (CoIn). Historically thought to be rare occurrences, the all-important role of CoIns in the theoretical description of photochemistry has been unequivocally demonstrated by the early work of Robb, Bernardi and Olivucci.[22,23] How one accurately accounts for these regions of degeneracy, and their role in photochemical dynamics, is dealt with in the Theory chapter. Here, a simple visual representation outlining the difference between diabatic and adiabatic processes, as well as the role of CoIns, is presented in Figure 1.3.

Following the discussion of the time-independent approach to photochemistry, and the description of some of the subtleties that one encounters, it must be concluded that it is usually more appropriate, or at least more complete, to use a time *dependent* approach to model photochemistry generally. The idea is that by propagating an entire molecular wavefunction on a multi-dimensional PES, all accessible pathways can in principle be explored. Provided the electronic structure properties are calculated properly, a quantitative picture of the photophysics/chemistry can be built up. One obvious drawback is the scaling of the numerical effort of such methods, which is at least linear with respect to propagation time. Conveniently though, photophysical/chemical processes can occur on very short timescales and hence the propagation time can be limited.[24] These processes are termed *ultrafast* and occur on the scale of 10s - 1000s of femtoseconds (1 fs = 10^{-15} s).[25,26]

It is exactly this area of photochemistry that is the subject of this work. In a sense,

these processes are perfectly matched with the use of highly accurate QD simulations. On the one hand, the ultrafast nature of many photophysical/chemical processes means that expensive QD methods can be used to simulate the early-time dynamics of a system, and gain a lot of information without having to extrapolate to longer time scales. On the other hand, these complex phenomena require the use of very accurate simulations in order to be understood/modelled properly. The field of QD is thus closely linked with photochemistry and the desire to understand complex, light-induced phenomena. This has been driving the field forward towards ever wider applications, from the study of small astrochemically relevant molecules to large biological chromophores. Seen in this light, the formidable theoretical challenge associated with unravelling precise mechanisms is partly responsible for the growing academic interest in the field, which can only be a good thing.

1.3.2 Photodissociation

One particular photochemical reaction that is central to much of the work discussed later on is photodissociation, i.e. the breaking of one or more chemical bonds as a result of photon absorption. In simple terms, the breaking of a bond can occur when a molecule absorbs a photon of the right energy to excite it to an electronically excited state which has dissociative character (i.e. not bound) with respect to that bond. Pictorially, this process is shown in Figure 1.4, where a single bond between two atoms, or molecular fragments, is broken after an excitation to an unbound electronic state. In the next chapter a more formal description is given of how photoexcitation is modelled in the framework of QD, but here a simple picture suffices to highlight one particularly challenging aspect of simulating photodissociation with computational methods, i.e. the need for a description of the PES far away from the ground state equilibrium structure.

Computing accurate energies of molecules at non-equilibrium geometries is a notoriously hard problem in quantum chemistry. The reason can be understood by looking at a typical non-equilibrium process such as bond dissociation of a diatomic molecule. Taking the simplest example, H_2, the wavefunction of the molecule at equilibrium is solely characterised by the doubly occupied bonding σ orbital, which is delocalised over both atoms. As the internuclear distance increases, the energy gap between the σ and σ^* orbitals decreases until they are eventually degenerate (the dissociation

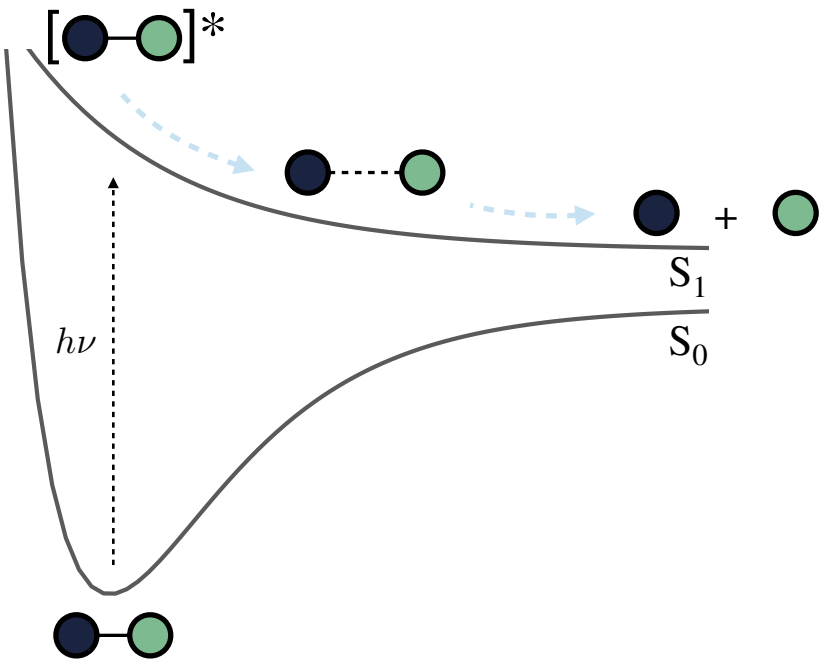


Figure 1.4: Excitation of a molecule from its ground state (S_0) to an unbound electronic excited state (S_1) by a photon. The molecular excited state is denoted by an asterisk. Blue arrows show the direction of nuclear motion.

limit). Consider the wavefunction at this point, it should be equally characterised by the $(\sigma)^2$ and $(\sigma^*)^2$ configurations, since these are now exactly equivalent. This however, is incompatible with the original $(\sigma)^2$ wavefunction characterising the equilibrium ground state geometry. The problem lies in the fact that, historically, most methods to calculate the electronic structure of molecules relied on a single reference wavefunction, which means that when the electronic character deviates significantly from the reference, the method breaks down. As an example, the most prototypical wavefunction-based electronic structure method, Hartree-Fock, fails massively at describing bond dissociations for precisely this reason. The equilibrium bond length is qualitatively well described by a reference wavefunction in which the bonding molecular orbital is doubly occupied, and the anti-bonding orbital is empty. However, if this same reference is used at large bond lengths, the method will predict an ionic state in which both electrons are localised on a single H nucleus, while the other has none. As a consequence, the calculated dissociation energy is much too high.

An obvious solution to this issue is to use reference wavefunctions that are linear combinations of N-electron functions, making them flexible enough to describe both distinct and degenerate electronic configurations.[27] These approaches belong to the

multi-configurational self- consistent field (MCSCF) class of methods and will be used almost exclusively throughout the work presented. The theoretical foundations are discussed in the next chapter.

1.4 Role of experimental measurement

The QD simulations used in this work can in principle provide a full, time-evolving wavefunction. Given this, one has access to a large amount of information about the system, and all that is needed is an appropriate way to extract this information, with the aim of gaining some physical insights. Thus, a transformation is necessary from $\Psi \rightarrow$ *observables*. The postulates of QM tell us how to use the wavefunction to calculate the expectation value of a property, $\mathcal{A}$, by using an appropriate operator, $\hat{A}$, acting on the wavefunction in the following way: $\mathcal{A} = \langle\Psi|\,\hat{A}\,|\Psi\rangle$. Note that if the property in question is a physical one, the operator must be Hermitian to ensure strictly real expectation values. Since the wavefunction is time-dependent, the expectation values can equally be calculated at different times. Properties such as the energy of the system or the position of the atoms over time can be obtained. Slightly more involved techniques also allow for the construction of absorption spectra, which can be compared directly to those measured in a lab. Additionally, one may want to calculate the rates of certain photochemical processes or determine the ratio between different photoproducts. These choices will ultimately be system-dependent.

So, the determination of properties of interest from a calculated wavefunction is relatively straightforward, but how are experiments able to provide the same information if the wavefunction cannot be measured? In the context of photochemistry, it is worth briefly mentioning the achievements of A. Zewail, who pioneered the field of ultrafast laser spectroscopy which allowed for the experimental study of short time-scale, atomistic dynamics of molecules (for which he won the Nobel prize in 1999).[28] He and his colleagues were the first to use femtosecond lasers to study the evolution of chemical reactions, mapping out the nuclear dynamics of, first, the ICN molecule,[29] and later, importantly, the dissociation of NaI.[30] This experiment was the perfect example of how bond breaking can proceed via the interplay between covalent and ionic configurations, through 1 nuclear coordinate and 2 electronic coordinates. The preparation of a system in non-stationary states (i.e. creating wavepackets) was demonstrated, and

its time evolution could be tracked. These results proved, unambiguously, that it is indeed possible to study molecular dynamics in the ultrafast, femtosecond regime at a time where this was thought impossible. Since this seminal work, a wide range of time-resolved spectroscopic techniques have been developed that allow for the study of quantum dynamical phenomena in many different kinds of systems.

The development of experimental techniques went hand in hand with the continuous improvement of computational tools as well, both benefitting from each other. From the theoretical side, once desired properties have been extracted from a calculated wavefunction, a comparison to available experimental data is usually done to validate the accuracy of the chosen theoretical approach. This then, in one word, is the role of experimental measurements in the context of this thesis; *validation* of theoretical results. That word should not be taken too literally though, since it is in principle not possible to derive the absolute validity of a model based on whether or not the predicted results match experimentally determined ones. The reason for this is two-fold. Firstly, there is always the possibility of getting the right answer for the wrong reasons, a point that was already mentioned in the discussion of the instrumentalist approach to scientific theories. Because of this it is important that the methods used to, in this case, characterise the photoinduced dynamics of molecular systems, are able to provide an explanation of the observed phenomena rather than merely predicting the outcomes of experiments. Then, once a mechanistic understanding of a certain process is obtained as the result of performing simulations, it is important to check that this mechanism does indeed lead to accurate predictions of experimental observables. The combination of experimental validation and a good explanation of the observations provides a sound way to test the performance of a given calculation, but is very difficult in practice.

The second reason why validation by experiment should not be seen as the ultimate gold standard is that, rather than getting the right answer for the wrong reason, one may get the wrong answer for the wrong reason. That is to say, the model could be correct, but at the same time in conflict with experimental data. This could simply be because the experimental conditions are not explicitly simulated, leading to a "wrong" answer obtained from a calculation describing the correct physics/chemistry of the system. This brings up the issue of assessing what has actually been measured in an experiment, and whether or not this can be accurately included in a simulation

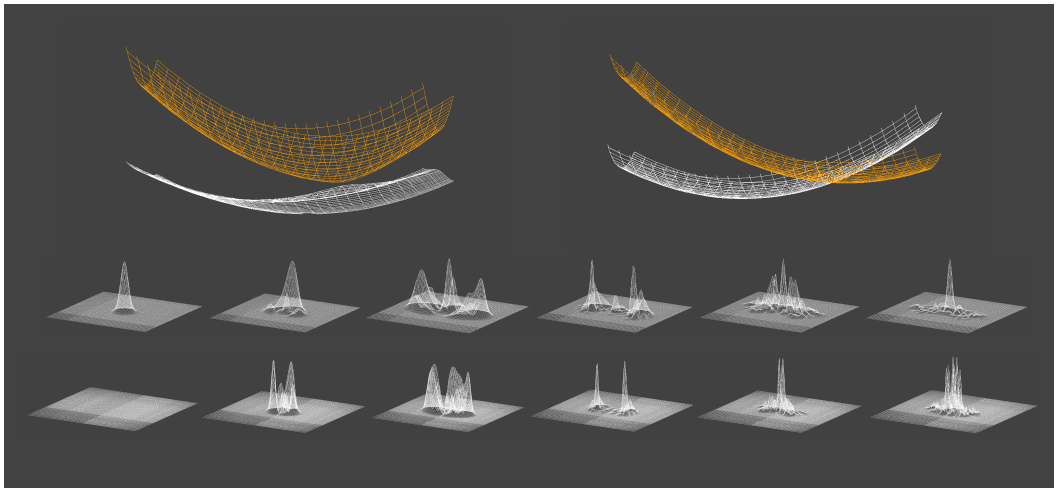
that aims to reproduce that result. One could always make the argument that all systems interact with each other at the quantum level and hence the experimental apparatus is bound to affect the results in a way that cannot be accounted for in simulations, nor in the analysis of the experimental data. Whether or not these effects are significant is another matter. Either way, the more problematic aspect of comparison to experimentally reported data is that the experimental conditions are not always known exactly. For example, in pump-probe experiments it is not straightforward to define $t = 0$, or the exact shape of the pulses used to excite the sample. Furthermore, the effects of thermal motion or zero-point vibrational energy are rarely known or investigated.[31] Lastly, many experimentally determined quantities are the result of some kind of fitting of the recorded data (e.g. lifetimes) which introduce assumptions that, again, are either unknown or hard to account for in computational models.

It is important to keep the limitations of comparison to experiment in mind, when judging the validity of a theoretically obtained result, but of course one can, and should, try to compare data where possible. It would not make much sense to stop comparing results obtained from different sources (whether experimental or theoretical), but the main point of this section is that it should not be seen as the ultimate goal of computational science. The real value of a model is, arguably, its explanatory power and should therefore be judged, not by how well it reproduces or predicts reality, but how well it explains it. In that context, the role of experiments is to help determine which theories lead to *false* conclusions, predictions or explanations. This relates to the ideas of falsification and deductive reasoning proposed by Popper, and others, as the underlying principles of what drives scientific discoveries, and progress in general, forward.[32]



CHAPTER 2

Theory



2.1 Introduction

This chapter describes the theoretical background of all computational methods used in this thesis. No new results are presented here, but rather the aim is to provide a detailed overview of how quantum theory has been used to formulate a variety of different methods that are used to calculate, on the one hand, the electronic structure of a system, and on the other hand, its time evolution. Hence this chapter is split into two main parts, each focusing on a different regime, i.e. the time-independent and the time-dependent domain. Within this structure, apart from just discussing the different possible calculations one can perform, various aspects of theoretical photochemistry/quantum dynamics are highlighted. This includes, for example, the correct description of Conical Intersections (CoIns), and the distinction between diabatic and adiabatic PESs (for an excellent review of these topics which follows roughly the same structure, see ref. [33]). All topics included here will be relevant in one or more of the subsequent chapters, and so this theory chapter essentially serves to unambiguously define all the specific terminology used in order to avoid possible confusion or misunderstanding.

2.2 Electronic Structure Methods

In this first part, related to the electronic structure of molecular systems, the time-independent Schrödinger equation will be discussed. How to solve this equation is the entire premise of quantum chemistry and, since its inception in the late 50's and 60's, many different computational methods have been proposed, developed and applied to do exactly that. A number of these methods are described, starting from the "simplest" one and then moving on to more sophisticated approaches. Each time, the strengths and weaknesses are highlighted. It will become clear that it is certainly not a trivial task to solve the Schrödinger equation generally, hence the wide variety of methods dedicated to this goal. Additionally, one realises that there are many different factors that influence how well a given method is able to describe the electronic structure of a particular molecule. When starting to discuss these factors, one wanders into a minefield of terminologies that are widely used but heavily debated at the same time. In this context, perhaps the most commonly discussed factor is the treatment

of electron correlation, meaning how an electron is affected by all other electrons in the system.

Broadly speaking, there are two distinct types of correlation, static and dynamic. These terms are very widely used but there is still active debate about their specific meaning (in fact, an entire Faraday Discussion conference was recently dedicated to this topic[34–36]). This is because different electronic structure methods are able to capture these contributions to the energy to varying extents. For the purposes of this thesis, it suffices to say that static electron correlation arises as a result of a few close-lying (near-degenerate) electronic configurations with respect to a reference wavefunction, which can lead to states with a fractional excited state occupation number. This is a highly system-specific effect and, if treated poorly, can lead to qualitatively wrong descriptions of the wavefunction, and as a result, its associated properties. Dynamical correlation is defined as involving many different excited configurations, all with small occupations. In this case, the wavefunction will be dominated by a single reference configuration, meaning that the errors introduced by not treating this effect properly will be mostly quantitative.[37] Interestingly, both these forms of correlation are defined based on a comparison with a certain "reference" wavefunction, the choice of which is a second factor that is very relevant in the field of quantum chemistry. Even though many ways exist to construct a suitable reference, most methods use a similar approach, as will become clear.

Both the correlation and the choice of reference wavefunction will be discussed for each of the methods outlined in what follows. This list is by no means exhaustive however, and lacks much detail. The focus of the discussion lies more on the relevant features of each approach in the specific context of this work. For more extensive reviews, see refs. [38–40].

2.2.1 The TISE and other generalities

An appropriate description of the structure of electrons in a static molecule is obtained by solving the electronic time-independent Schrödinger equation (TISE):

$$\hat{H}^{el}\Psi^{el}(\mathbf{r};\mathbf{R}) = E^{el}\Psi^{el}(\mathbf{r};\mathbf{R}) \quad . \tag{2.1}$$

The electronic Hamiltonian $\hat{H}^{el}$ contains only the electronic kinetic energy operator and the potential energy terms related to the Coulombic electron-electron repulsion,

nuclear-nuclear repulsion, and electron-nuclear attraction, hence describing the motion of N electrons in a field of point charges, given by the nuclear positions. The coordinates of all nuclei are denoted by $\mathbf{R}$, whereas the electronic coordinates (both spatial and spin coordinates) are collected in $\mathbf{r}$. The electronic wavefunction itself is then said to be parametrically dependent on $\mathbf{R}$. The electronic energy E^{el} is hence determined for a specific nuclear configuration by finding the eigenvalues of $\hat{H}^{el}$. By solving Equation 2.1 at many different nuclear positions, the potential energy landscape of a molecule can start to be mapped out and indeed, this is how one can obtain the necessary PESs that govern the evolution of a molecule during nuclear dynamics simulations (more on this later).

Before describing specific methods, some general concepts are first explained, starting with the expression of the N-electron wavefunction, $\Psi(\mathbf{r}_1, ..., \mathbf{r}_N)$. A typical approach to break down a high dimensional problem is to use a sum-of-products form. In the case of a N-dimensional wavefunction, one can always expand this as a product of the set of n one-electron basis functions, $\{\chi_i\}$, and their coefficients, C:

$$\Psi(\mathbf{r}_1, ..., \mathbf{r}_N) = \sum_{i,j,...,n} C_{i,j,...,n}\chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_2)\cdots\chi_n(\mathbf{r}_N) \quad . \tag{2.2}$$

Furthermore, the Pauli principle requires an antisymmetric wavefunction,

$$\Psi(\mathbf{r}_1, ..., \mathbf{r}_i, \mathbf{r}_j, ..., \mathbf{r}_N) = -\Psi(\mathbf{r}_1, ..., \mathbf{r}_j, \mathbf{r}_i, ..., \mathbf{r}_N) \quad , \tag{2.3}$$

which allows one to express Ψ in terms of determinants composed of the complete set of N one-electron spin orbitals:

$$\Psi_{SD}(\mathbf{r}_1, ..., \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{r}_1) & \chi_2(\mathbf{r}_1) & \cdots & \chi_N(\mathbf{r}_1) \\ \chi_1(\mathbf{r}_2) & \chi_2(\mathbf{r}_2) & \cdots & \chi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{r}_N) & \chi_2(\mathbf{r}_N) & \cdots & \chi_N(\mathbf{r}_N) \end{vmatrix} \quad . \tag{2.4}$$

This is known as a Slater Determinant (SD). All electrons are included and the motion of one electron depends on all others, making them correlated since swapping any one spin orbital with another yields the correct antisymmetric wavefunction. Clearly, this is a better way of describing the wavefunction than the linear combination of Hartree products in Equation 2.2, but a natural question arises; out of all possible SDs spanning the Fock space†, which one is most suited to describe a given N-electron system? This

†A Fock space is a type of Hilbert space used to represent states of particles.

question lies at the foundation of many different electronic structure methods. It shall become clear that each method takes a slightly different approach at answering it, some using a combination of multiple SDs while others may try to iteratively optimise a single one.

Finally, it ought to be clarified what type of basis functions one might use to construct a molecular spin-orbital. Consider a set of atomic orbitals (AO), φ, which are used in a linear combination to expand the i^{th} spacial molecular orbital (MO), ϕ_i:

$$\phi_i = \sum_{n=1}^{N_{AO}} c_{i,n}\varphi_{i,n} \quad . \tag{2.5}$$

This is known as the linear combination of atomic orbitals approach (LCAO), and is the foundation of molecular orbital theory. Adding the spin contribution to these MOs yields the spin-orbitals introduced earlier.

2.2.2 Hartree-Fock

The discussion of specific methods to solve the TISE must almost necessarily start with one of the simplest ways to represent a many-body ground state wavefunction, i.e. by using a single SD. This can be written in the more compact notation, $|\det ...\rangle$, and with the coordinate variables left out for clarity:

$$|\Psi_{HF}\rangle = |\det \chi_1, \chi_2, ..., \chi_N\rangle \quad . \tag{2.6}$$

This is the Hartree-Fock (HF) expression of a particular molecular state. The energy of the system is simply obtained from the expectation value $\langle\Psi_{HF}|\hat{H}_{el}|\Psi_{HF}\rangle = E_0$. As for the choice of spin orbitals $\{\chi_n\}$, the variational principle dictates that the orbitals leading to the lowest value of E_0 are the best. Hence one can iteratively vary the orbitals until E_0 is minimised, the only criteria being that the orbitals remain orthonormal. This procedure is termed the Self-Consistent Field method (SCF). Because of its simplicity, HF has been instrumental in the development of many electronic structure methods that rely on the same basic idea. However, many of these other approaches aim to correct for the shortcomings of HF, which are briefly outlined.

Although a single SD can often accurately describe the electronic structure of a molecule in its equilibrium ground state, as soon as the geometry is changed sufficiently, or an electron is promoted to a higher lying orbital (i.e. forming an excited state), this

single determinant description may quickly start to fail. Specifically, one of the main drawbacks of HF is that electron-electron interaction is only described by a mean-field approximation, meaning that each electron only experiences the average presence of all other electrons,[41] while inter-electron correlation is not described, except for Fermi correlation. It is therefore not surprising that HF theory may not be suitable in cases where correlation, static or dynamic, becomes important. Such scenarios include the stretching of a chemical bond, where well-separated electronic configurations can become near-degenerate at long bond lengths and hence the static electron correlation becomes relevant. Since a single determinant cannot possibly describe two or more distinct configurations at once (with the possible exception of using complex-valued orbitals[42]), this approach is destined to fail in such cases.[43] One of the main themes of this work is the description of photodissociation reactions, for which the use of HF can safely be ruled out.

2.2.3 Post-HF

The next set of methods that will be discussed can be broadly classified as "post-HF", which refers to the fact that although they can improve massively on the performance of standard HF, the initial starting point remains essentially the same.

Configuration Interaction

To include electron correlation effects in a HF-style calculation, one may use the conceptually simple approach of combining more than one determinant. This idea forms the basis of an important family of methods known as Configuration Interaction (CI). The wavefunction is now expressed as a linear combination of expansion coefficients, c, and expansion functions, $|\Phi\rangle$:

$$|\Psi_{CI}\rangle = \sum_i c_i |\Phi_i\rangle \quad . \tag{2.7}$$

The expansion functions are new determinants, built from excitations out of the HF reference, now $|\phi_0\rangle$, and combined together. A "configuration" then refers to one of the interacting expansion functions, as per the following ansatz[44]

$$|\Psi_{CI}\rangle = c_0 |\phi_0\rangle + \sum_{a,r} c_a^r |\phi_a^r\rangle + \sum_{a>b,r>s} c_{ab}^{rs} |\phi_{ab}^{rs}\rangle + ... \quad . \tag{2.8}$$

Here, $\left|\phi_i^j\right\rangle$ corresponds to a substitution of occupied spin orbitals in the reference (denoted by the set $\{\chi_a, \chi_b, ...\}$) by other, unoccupied, spin orbitals (in $\{\chi_r, \chi_s, ...\}$). This is more elegantly described in the notation of the second quantisation formalism, where annihilation and creation operators (a_i and $a_j^\dagger$, respectively) are introduced to excite an electron out of orbital i, and into orbital j.[45,46] The CI configurations are thus expressed as $|\phi_a^r\rangle = a_a^\dagger a_r |\phi_0\rangle$, $|\phi_{ab}^{rs}\rangle = a_a^\dagger a_b^\dagger a_r a_s |\phi_0\rangle$, and so on for higher order excitations. The number of excitations included in the expansion can be truncated at any point, but if no truncation is made and all excitations up to N-tuply excited determinants are included, the above expression of the wavefunction is exact within the chosen basis set. Calculations which use this expansion are termed Full CI (FCI). They are prohibitively expensive for most applications, making the option of truncation crucial. The most widely used truncations are CI with single excitations (CIS), and CI with single and double excitations (CISD). These have quite different uses however. The only improvement offered by CIS over HF is its description of electronically excited states.[47,48] It does not improve the ground state wavefunction since the HF reference does not interact with singly-excited wavefunctions, according to Brillouin's theorem: $\langle\phi_0| \hat{H} |\phi_a^r\rangle = 0$.[49] CISD on the other hand, does provide an improvement of the description of the ground state by capturing approximately 95% of the (dynamic) correlation energy in small molecules.[50] Generally speaking, the CI approach can be said to improve on HF theory because it allows for both kinds of electron correlation effects to be included, at least to some extent. From Equation 2.8 it is clear that, on the one hand, static correlation is naturally recovered by the inclusion of (near) degenerate configurations in the set of all excitations, $\{\Phi\}$. Dynamical correlation on the other hand, is accounted for by combining the HF reference, $|\phi_0\rangle$, with increasingly higher order excitations (note that these excitations will also increasingly slow down the convergence of the dynamical correlation energy).[51]

Obtaining the electronic energies of a system, in state i, using the CI method is achieved by the following expectation value formula,

$$E_i = \langle\Psi_i| \hat{H}_{el} |\Psi_i\rangle \quad , \tag{2.9}$$

which leads to a standard matrix eigenvalue problem of the form $\mathbf{Hc}_i = E_i\mathbf{c}_i$, with the Hamiltonian and the coefficients now in matrix and vector form, respectively. The eigenvalues of $\mathbf{H}$ correspond to the excited state energies, and the eigenvectors are

the corresponding wavefunctions. Thus, CI methods are easy to use and can provide relatively good accuracy for small molecules but it represents a kind of "brute force" approach to the problem. Because of its problematic scaling with respect to number of orbitals and number of electrons, this approach will quickly become unfeasible.

Coupled Cluster

Moving on to a class of methods that has long been considered the "gold standard" of electronic structure theories, the Coupled Cluster (CC) methods. These approaches are based on a different parametrisation of the wavefunction, which allows for more accurate results to be obtained at a cost similar to CI methods at the same level of truncation of the expansion. However, it is still based on the original HF theory and still uses a single ground state reference, expanded with excited configurations. The expansion of the ground state has the following exponential form:

$$|\Psi_{CC}\rangle = e^{\hat{T}} |\phi_0\rangle \quad , \tag{2.10}$$

where $|\phi_0\rangle$ is still a single Slater determinant, and the excitation operator $\hat{T}$ is expanded in N-electron operators which define the excitations out of the ground state, much like in Equation 2.8:

$$\hat{T} = \hat{T_1} + \hat{T_2} + ... + \hat{T_N} \quad . \tag{2.11}$$

Again, the truncation of the excitation operator to n^{th} order leads to the different levels of CC calculations with the most widely used being CC with single and double excitations (CCSD), where $\hat{T} = \hat{T_1} + \hat{T_2}$. No truncation yields, again, the FCI result. The energy is calculated by inserting the above ansatz in the TISE to obtain:

$$\tilde{H}^{el} |\Psi_0\rangle = E^{el} |\Psi_0\rangle \quad , \tag{2.12}$$

where

$$\tilde{H}^{el} = e^{-\hat{T}} \hat{H}^{el} e^{\hat{T}} \quad . \tag{2.13}$$

More detail is found in the literature[52–54] but one point worth highlighting is that a big advantage of CC methods over CI methods is that they are *size-extensive*, meaning that they scale properly with respect to the number of electrons. In effect, it gives CC a wider applicability since it can calculate more accurate properties, of larger molecules than CI. Another important property of CC that is not found in truncated

CI methods is *size-consistency*, which means that the calculated energy of a system equals the sum of energies of its subsystems (i.e. its constituent atoms, or any number of possible fragments).

Extending the CC formalism to calculate electronic excited state energies is referred to as the Equation of Motion CC method.[54–56] The excited state wavefunction is obtained by applying an appropriate linear operator, $\hat{R}$, to the ground state wavefunction of 2.10 to obtain the i^{th} state

$$|\Psi_i\rangle = \hat{R}(i)\,|\Psi_{CC}\rangle \quad . \tag{2.14}$$

This approach provides a powerful way of obtaining excited state properties in a relatively easy "black-box" way. Although quite expensive as a method (EOM-CCSD for example has a scaling of $\mathcal{O}(N^6)$, whereas HF is $\mathcal{O}(N^4)$, where N is the number of electrons), this should not be an issue for small organic molecules. However, its applicability is limited to systems that are well described by the ground state reference wavefunction. Hence, bond breaking or other potentially open-shell systems may not be accurately calculated.

Possible solutions to this problem have been developed that can include the strong static correlation of such systems, most notably spin-flip methods.[57–60] In essence, these methods describe the (singlet) reference wavefunction as a triplet reference state, acted upon by a spin-flipping excitation operator. This choice of reference provides more flexibility for the method and hence it is able to describe a wider range of systems, specifically those with strong static electron correlation. Another route that can be taken to alleviate the problems associated with the single-reference nature of the method is to use multiple SD as an initial reference, leading to multi-reference CC.[61] The next method that will be discussed is also based on this approach.

MRCI

The idea behind CI, described previously, can be extended to a multi-reference (MR) version, i.e. MRCI,[62] which illustrates the principle behind all multi-reference methods perfectly. This extension was originally developed to deal with systems that were not well described by a single reference wavefunction (i.e. the HF determinant), such as bond dissociations, excited states, and other situations in which orbitals could have an open-shell character. The idea is simple, namely to replace the single reference

function by a set of functions: $|\phi_0\rangle \rightarrow \{\Phi_R\}_{i=1}^{N}$, all acting as a reference. This set of multiple references is then expanded in exactly the same way as single-reference CI (Equation 2.8):

$$|\Psi_{MRCI}\rangle = \sum_{R} c_R |\phi_R\rangle + \sum_{S,r} c_S^r |\phi_S^r\rangle + \sum_{D,r,s} c_D^{rs} |\phi_D^{rs}\rangle + ... \quad . \tag{2.15}$$

The notation here defines $|\phi_R\rangle$ as one reference out of the full set of possible configurations. All single excitations out of all reference functions are given by $|\phi_S^r\rangle$, and similarly, all double excitations by $|\phi_D^{rs}\rangle$. In other words, typical CI excitations are now done for each of the different reference configurations individually. In theory, higher order excitations can be included in the expansion, but for practical reasons related to the cost of the calculations, one never really goes beyond singles and doubles (i.e. MRCISD). Note also that defining the set of possible initial references poses a massive practical challenge for the method.

In the interest of brevity, MRCI approaches will not be further expanded upon and one is referred to the literature.[38,39,51,63] Instead, more time is spent explaining a class of methods that deals with the shortcomings of using single-reference wavefunctions in a slightly different way.

MCSCF

The multi-configurational self-consistent field (MCSCF) family of methods represents a way to go beyond HF, while still essentially using it as a starting point. As the name suggests, MCSCF methods make use of multiple different electronic configurations, generated from an initial reference. In order to avoid facing the prohibitive cost of FCI by considering all possible configurations, a reduction is achieved by dividing the set of molecular orbitals (MO) into subsets with different imposed restrictions on the allowed electronic configurations within them, and the allowed number of excitations out of them. Such a scheme already implies a major practical difficulty in that a choice needs to be made with respect to how the orbitals are divided to form the various subspaces, and which specific excitations are allowed for each subspace. Such a decision is based on a judgement of the importance of a given orbital, and unsurprisingly this decision will vary for different molecules, and is also affected by the particular chemistry that one is aiming to describe. For that reason, no formal rules can be applied to obtain an optimal division of orbitals, only intuition and rules of thumb.

The different types of subspaces are defined as follows: *Frozen* orbitals are those that are always doubly occupied, with no excitations being allowed out of this space. Next, the *inactive* orbitals are doubly occupied in all reference configurations but a number of excitations are allowed when the wavefunction is being generated. Within the *active* orbital space, any number of different configurations is allowed, including all possible configurations (i.e. FCI), or a more restricted set of configurations defined by the user. *Virtual* orbitals are unoccupied in the references but can accept excitations into them, and finally the *deleted* orbitals are always empty and aren't included in the possible excitations. Some of these subspaces are shown in Figure 2.1, along with specific allowed interactions for one particular flavour of MCSCF discussed later on.

Apart from the way the configurations are generated, the main feature of these methods that distinguishes them from every other approach previously discussed is related to what elements are optimised during the SCF procedure. Up until this point, only the expansion coefficients were considered, whereas in MCSCF, both the coefficients and the MOs are optimised simultaneously. Over the years, many efficient algorithms have been developed which can perform these optimisations, both for a single electronic state,[64] and, crucially, also for multiple excited electronic states.[65]

By far the most widely used method in this family is the Complete Active Space method (CASSCF),[66] which employs the following wavefunction (note the similarity with Equation 2.7):

$$|\Psi_{CAS}\rangle = \sum_{i=1}^{N_{CAS}} c_i |\Phi_i\rangle \quad , \tag{2.16}$$

where $|\Phi_i\rangle$ are now linear combinations of Slater determinants, called Configuration State Functions (CSFs)[†], generated from the active orbital subspace containing a fixed number of electrons. All possible configurations of these electrons in the subspace are included in the wavefunction, leading to essentially a FCI expansion within the selected orbitals and number of electrons. Obtaining the energies of the wavefunction is conceptually identical to CI, i.e. solving an eigenvalue problem. However, the fact that both coefficients and orbitals are optimised simultaneously introduces additional complications because of the often strong coupling between the two. Many different algorithms to carry out this optimisation have been proposed,[68] but will not be dis-

[†]Although CSFs and SDs are sometimes used interchangeably, a key difference is that CSFs form a spin-adapted basis, i.e. they are eigenvalues of the square of the spin operator, $\hat{S}^2$ and of its projection along z, $\hat{S}_z$. In contrast, SDs are only eigenfunctions of $\hat{S}_z$.[67]

cussed here. This method has also been formulated for the computation of multiple excited electronic states, leading to the state-averaged CASSCF method. Here, each individual state is given a weight which ensures a balanced description of all states included.[69]

A few remarks are in order. Firstly, the number of CSFs grows factorially with the number of orbitals included in the active space, limiting the ultimate size of calculations that are practically possible. Secondly, the CASSCF method is clearly not "black-box" due to the aforementioned *choice* of which orbitals to include in which subspaces. These decisions can have a large impact on the accuracy and validity of the result, making it a relatively difficult method to use.[70] Further problems arise when different regions of the configurational space need to be described, as is the case in QD simulations. Since the set of optimal orbitals is likely to vary across different geometries, the possibility of orbital rotations from one space to another makes these methods potentially unstable.[71]

Turning now to the treatment of electron correlation, it should be no surprise that static correlation is correctly described (at least within the chosen active space), since a number of (near) degenerate configurations, the CSFs, are used as a reference. Dynamical correlation on the other hand, is hardly accounted for. It has been shown that in MCSCF methods this type of correlation has its origin in double excitations from occupied orbitals in the inactive and active subspaces into the virtual orbitals, and hence it is not included in standard CASSCF calculations.[72]

Various extensions have thus been developed to correct for this missing contribution. The first, and most popular, is the CASSCF method with a 2^{nd} order perturbation theory correction added, the CASPT2 method.[73,74] This approach mirrors the development of the Møller-Plesset second order perturbation theory (MP2) method, which adds a similar correction to the HF wavefunction.[53,75] In typical perturbation theory fashion, an unperturbed (zero$^{\text{th}}$ order) reference state, $\left|\Psi^{(0)}\right\rangle$, is defined and chosen to be the equal to $|\Psi_{CAS}\rangle$, plus a correction expressed as a perturbation:

$$|\Psi_{PT2}\rangle = \left|\Psi^{(0)}\right\rangle + \left|\Psi^{(1)}\right\rangle \quad . \tag{2.17}$$

Specifically, $\left|\Psi^{(1)}\right\rangle$ is constructed from an expansion of the configurations in the space spanned by all single and double excitations out of $|\Psi_{CAS}\rangle$. Similarly, the Hamiltonian

has the form

$$\hat{H}_{PT2} = \hat{H}^{(0)} + \hat{H}^{(1)} \quad , \tag{2.18}$$

where the zero$^{\text{th}}$ order Hamiltonian is a one-electron operator similar to the one used in MP2 (more detail given in ref. [76]). These are then used to give a first and second order correction to the CASSCF energy, respectively:

$$E^{(1)} = \left\langle \Psi^{(0)} \right| \hat{H}^{(1)} \left| \Psi^{(0)} \right\rangle \quad , \tag{2.19}$$

and

$$E^{(2)} = \left\langle \Psi^{(0)} \right| \hat{H}^{(1)} \left| \Psi^{(1)} \right\rangle \quad . \tag{2.20}$$

In practice, the construction of $\left|\Psi^{(1)}\right\rangle$ and $\hat{H}^{(0)}$ are rather involved and cause the CASPT2 method to come with a significantly higher computational cost compared to CASSCF. Further drawbacks of the method include the possible appearance of "intruder states" into the manifold of calculated excited states, and as with all perturbative approaches, its predictions may no longer be valid when the perturbation becomes too large. It is possible, however, to achieve very accurate results of both the electronic ground and excited states, accounting for the missing dynamical correlation, and is therefore widely used in the theoretical prediction of spectroscopic properties of organic molecules, as well as other fields.[77]

The last multi-configurational method that will be discussed explicitly is, again, closely related to the CASSCF approach. Rather than performing a FCI expansion in the active space, some restrictions are introduced and the method is thus referred to as Restricted Active Space SCF (RASSCF).[78,79] These restrictions are imposed on certain orbitals within the active space, giving rise to 3 subspaces. The RAS1 space contains doubly occupied orbitals, out of which only a limited number of excitations is allowed. Next, the RAS2 space is exactly the same as the CAS, where all possible configurations are permitted. Lastly, the RAS3 space consists of unoccupied orbitals which can only accept a limited number of electrons. These subspaces and their interactions are shown in Figure 2.1.

The introduction of these restrictions has of course the effect of speeding up the procedure since a smaller number of total configurations is computed. Additionally, it is also able to recover some degree of dynamical correlation because, in principle, it includes the required double excitations from occupied orbitals into virtual ones.[72]

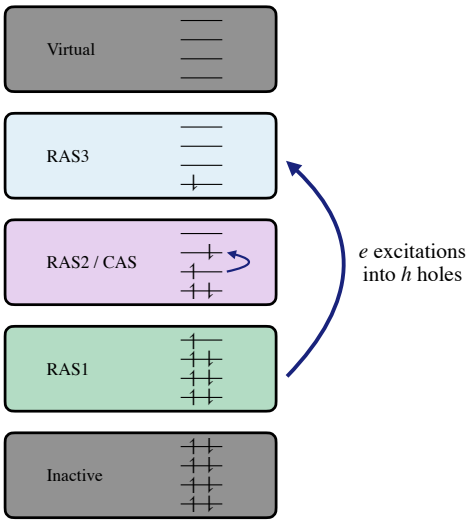


Figure 2.1: The RASSCF method. All configurations are allowed in the RAS2 space, whereas only e electrons from RAS1 can be excited into h holes in RAS3. Inactive and virtual orbitals remain doubly occupied and unoccupied, respectively. Frozen and deleted orbitals are not shown.

Apart from this point and some practical details, the RASSCF and CASSCF methods are equivalent, the former being an extension of the latter. Despite its advantages RASSCF is not as widely used though, presumably because it is even less "black-box" than CASSCF.

2.2.4 Note on DFT

A discussion of electronic structure methods would be incomplete without at least mentioning the most widely used approach to date, Density Functional Theory (DFT),[80,81] and its time-dependent extension TD-DFT[82] for excited states. The main reason for its popularity is arguably its ease of use, and favourable scaling with system size, allowing it to be used for comparatively large molecules (< 1000 atoms). DFT achieves these desirable properties by taking a radically different approach to electronic structure calculation, compared to the methods mentioned so far. Whereas all methods covered up to this point are based on the description of a wavefunction, DFT uses the electron density to obtain the ground state of a molecule, and its properties. This massively simplifies the many-electron problem and leads to major advantages such as the description of both single- and multi-reference systems, and the inclusion of electron correlation effects.[83]

All this does not come for free however, for there is one major drawback to the

method which limits its use for certain applications. DFT derives its validity from assuming an exact exchange functional is available to describe the electron density, which is rarely, if ever, the case. Instead, a plethora of different approximate functionals have been developed over the last 4-5 decades, all of which account for specific effects but none of which can be used universally.[84,85] For this reason, thorough benchmarking against suitable reference calculations (ideally wavefunction-based approaches) is crucial for the application of DFT to accurately study molecular properties. When calculating excited states with TD-DFT, additional problems arise and both the energies and the ordering of the excited states may not be reliable. Additionally, in the commonly used linear-response implementation of TD-DFT, the ground and excited states are treated on a different footing and describing any interactions between the two becomes problematic.[86–88] In summary then, although DFT remains an invaluable tool in computational chemistry, a general rule of thumb would be to avoid it if one can afford a wavefunction-based alternative.

2.2.5 Note on inter-state coupling

Accurately describing the dynamics on multiple electronic states requires, naturally, a notion of the interaction between those states. How to properly include this interaction is described in detail in the next section on QD methods, but it is briefly addressed here in the context of electronic structure calculations. Simply put, the strength of the coupling between two states is given by the so-called non-adiabatic coupling (NAC) vector. The calculation of this property is available for many electronic structure methods, but has always been somewhat obscure relative to properties such as the energy or its gradient for example. Not getting as much attention in the development of new approaches, the theory of the computation of NAC is still evolving.[89]

By mentioning this point, a longstanding disconnect between communities is highlighted. While including NAC is essential in the description of excited state dynamics, researchers focused on the development of electronic structure methods are typically more interested in molecular structure, and hence don't attach nearly as much importance to this quantity. As a result, the difficulty of finding suitable electronic structure methods for use in non-adiabatic dynamics is exacerbated.

2.3 Non-adiabatic Quantum Dynamics

If one is interested in understanding the fundamental mechanisms behind light-matter interactions, the TDSE (see Equation 1.2) provides a complete description of the time evolution of physical systems and can therefore, in principle, be used to study the entirety of photochemistry. The issue is, of course, how to solve it for realistic systems? At first glance, this equation is a simple initial value problem, where, if the wavefunction is known at $t = 0$, one can predict its state at any later time by propagating $\Psi(\mathbf{x}, t)$ according to the Hamiltonian operator, and some equations of motion (EoMs). However, as is often the case when trying to apply theory to a practical problem, this is much easier said than done. Right from the start, one may be discouraged by the fact that an exact solution to the TDSE only formally exists for a 2-particle system, e.g. a hydrogen atom. Fortunately though, the desire to understand the properties of more complex atoms and molecules has driven the development of many different approaches that are able to approximately solve the equation, constituting the field of Quantum Dynamics.

In simple terms, the problem of solving the TDSE is broken down into two parts, electronic and nuclear. The electronic part is needed to provide a description of the energy landscape on which the nuclei can move. This is essentially the difference between quantum and classical dynamics, where in the latter the nuclei are propagated by completely ignoring electrons. Such information is obtained through the aforementioned electronic structure methods, and is then used to parametrise a set of nuclear basis functions which are propagated according to specific EoMs. Once again, there exist a wide range of nuclear QD methods, which also fall into different families according to their formalisms and/or specific traits. A first important distinction is how the PES is represented. In the simplest case, only a single electronic state (usually the ground state) is considered, and the nuclei move classically according to Newton's laws of motion. This is called Ab Initio Molecular Dynamics (AIMD) and relies on the Born-Oppenheimer approximation, hence its equivalent name Born-Oppenheimer Molecular Dynamics. Since this method does not fall within the class of non-adiabatic dynamics, as it only deals with a single state, it is not further discussed here.

As far as methods that do fall in the category of non-adiabatic dynamics, some use adiabatic PESs, while others require diabatic ones (the distinction between the two

representations is clarified later). It is worth mentioning here that one obtains strictly adiabatic data from most electronic structure methods. Next, it is also common to represent the potential on a fixed, discretised grid, on which nuclear basis functions can move. In such an approach, the number of functions used for each degree of freedom (DoF) need to be suitably chosen, since this represents how fine the grid ultimately is and therefore how accurate the results will be. The choice of which nuclear basis functions to use is also an important feature that differentiates methods. A final important distinction relates to the EoMs of the nuclei. Many popular approaches use the central idea behind AIMD and propagate the nuclei classically, according to the forces given by the gradient of the potential energy, but in a manifold of states as opposed to just one. These methods are usually called semi-classical, for obvious reasons. In contrast, fully quantum methods are those where the basis functions are coupled, and follow variationally optimal trajectories.

Several specific non-adiabatic dynamics methods will be discussed in this section, including their classification according to the features that were just described. Before that, it is useful to explain some key theoretical considerations that illustrate the difficulty in accurately describing the coupled motion of electrons and nuclei, and hence justify the ongoing developments in the field.

2.3.1 Born-Oppenheimer and beyond

One particularly challenging aspect of solving the TDSE arises due to the coupling of nuclei and electrons and the need to describe their motion under their mutual attraction. However, a very important insight was arrived at by Born and Oppenheimer in 1927, who realised that since the mass of nucleons is so much greater than that of electrons, it is valid to assume an instantaneous rearrangement of electrons with respect to the slow moving nuclei.[90] This development allows one to imagine the time evolution of a molecule simply as a set of point masses, corresponding to the nuclei, which move on a single potential energy landscape dictated by electrons that follow the nuclei without delay. This is known as the Born-Oppenheimer Approximation (BOA) and has long been the cornerstone of theoretical chemistry. If one is interested in studying thermal reactions, only the ground electronic state needs to be considered and the BOA is perfectly valid. Even when considering dynamics involving excited states, this approximation can be used as long as only a single state, described by

a single electronic configuration, is considered. When, on the other hand, multiple configurations mix together as a function of nuclear position, the BOA can no longer correctly describe this situation. A superposition is created and the nuclei now move in a manifold of coupled electronic states. How to properly describe nuclear dynamics in such a manifold is essentially the subject of this entire thesis.

Before showing what causes the BOA to become invalid, and how to deal with that, it shall be used to derive a useful form of the TDSE. As a first step, the complete *molecular* Hamiltonian, $\hat{H}(\mathbf{x})$, and *molecular* wavefunction, $\Psi(\mathbf{x}, t)$, are split up in their nuclear and electronic terms. The Hamiltonian and the wavefunction can be written as

$$\hat{H}(\mathbf{x}) = \hat{T}^{nuc}(\mathbf{R}) + \hat{H}^{el}(\mathbf{r}; \mathbf{R}) \quad , \tag{2.21}$$

and

$$\Psi(\mathbf{x}, t) = \sum_m \chi_m(\mathbf{R}, t)\phi_m(\mathbf{r}; \mathbf{R}) \quad , \tag{2.22}$$

respectively. The following conventions shall be used: $\mathbf{R}$ and $\mathbf{r}$ are the nuclear and electronic coordinates respectively, and $\mathbf{r}; \mathbf{R}$ indicates a parametric dependence of the electronic wavefunction on the nuclei. Electronic states are denoted by m and n. The nuclear kinetic energy operator, $\hat{T}^{nuc}$, is expressed in terms of the second derivative with respect to $\mathbf{R}$, $\nabla^2_{\mathbf{R}}$, as follows

$$\hat{T}^{nuc}(\mathbf{R}) = -\frac{\hbar^2}{2\mathbf{M}}\nabla^2_{\mathbf{R}} \quad , \tag{2.23}$$

with $\mathbf{M}$ being the effective mass of the nuclei. The electronic Hamiltonian, $\hat{H}^{el}$, is a sum of the electron-electron and the nuclear-nuclear repulsion potentials, the nuclear-electron attraction potential and the electronic kinetic energy operator. In Equation 2.22, also known as the Born-Huang expansion, the set $\{\chi(\mathbf{R}, t)\}$ contains the time-dependent nuclear wavefunctions and $\{\phi(\mathbf{r}; \mathbf{R})\}$ is the orthonormal set of electronic functions, i.e. $\langle\phi_n|\phi_m\rangle = \delta_{nm}$ (Dirac's bra-ket notation, $\langle .|.\rangle$, shall be used throughout to represent an integral over all electronic coordinates). The two expressions in Equations 2.21 and 2.22 can be substituted in the TDSE, and if both sides are multiplied by $\langle\phi_n|$ and the electronic coordinates integrated out, one obtains the nuclear Schrödinger Equation:

$$i\hbar\frac{\partial}{\partial t}|\chi_m(\mathbf{R}, t)\rangle = \sum_m \Big[T_{nm}(\mathbf{R}) + V_{nm}(\mathbf{R})\Big]|\chi_m(\mathbf{R}, t)\rangle \quad , \tag{2.24}$$

where

$$V_{nm}(\mathbf{R}) = \langle \phi_n | \, \hat{H}^{el}(\mathbf{R}) \, | \phi_m \rangle \quad , \tag{2.25}$$

and

$$T_{nm}(\mathbf{R}) = \langle \phi_n | \, \hat{T}^{nuc}(\mathbf{R}) \, | \phi_m \rangle \quad . \tag{2.26}$$

Noticing that at this point no approximations have been made and the expression of Equation 2.24 is still exact, one can invoke the BOA to eliminate the nuclear-electronic coupling and hence simplify the TDSE. Starting with the potential matrix, V_{nm}, the off-diagonal coupling elements can be eliminated by choosing $\{\phi(\mathbf{r};\mathbf{R})\}$ to be the set of eigenfunctions of the electronic Hamiltonian (see Equation 2.1 in the previous section on electronic structure methods). This is referred to as the adiabatic[†] basis set, $\{\phi^{ad}(\mathbf{r};\mathbf{R})\}$, used to compute the adiabatic potential energies of electronic states, $E_m^{el}(\mathbf{R})$. Remembering that the basis is orthonormal, the adiabatic potential energy matrix is defined as

$$V_{nm}^{ad}(\mathbf{R}) = \delta_{nm} E_m^{el}(\mathbf{R}) \quad . \tag{2.27}$$

The adiabatic approximation now allows the kinetic energy term of Equation 2.26 to be similarly simplified to

$$T_{nm}^{ad}(\mathbf{R}) = \delta_{nm} \hat{T}^{nuc}(\mathbf{R}) \quad . \tag{2.28}$$

Finally, the adiabatic Schrödinger Equation, which is sometimes taken to be equal to the BOA, is written as

$$i\hbar \frac{\partial}{\partial t} \left| \chi_m^{ad}(\mathbf{R},t) \right\rangle = \left[T^{nuc}(\mathbf{R}) + E_m^{el}(\mathbf{R}) \right] \left| \chi_m^{ad}(\mathbf{R},t) \right\rangle \quad , \tag{2.29}$$

for every adiabatic electronic state m. One can easily see from this equation how the BOA reduces the coupled motion of electrons and nuclei to nuclear motion on a single PES, defined by E_m^{el} at different geometries $\mathbf{R}$. In fact, without the BOA the very concept of a PES could not exist, and the idea of nuclear geometries representing molecules would not make sense. This further emphasises its importance in theoretical chemistry, not only to make solving an equation easier, but also because it allows for a

[†]The term "adiabatic" derives its meaning in this context from the more traditional use of the word in thermodynamics. Adiabatic processes are idealised scenarios in which no heat or mass is transferred to the environment, based on the assumption that such processes are much slower than the original, fast physical process (e.g. work carried out by the system). In analogy, adiabatic electronic states are also the result of distinguishing between fast (electrons) and slow (nuclei) components of a system.

pictorial and intuitive view of chemical transformations. However, the consequence of using such a framework is that it is not possible to describe nuclear motion on multiple electronic states.

In the context of photochemistry, such a limitation is not acceptable and one needs to go beyond the BOA in order to explicitly include the coupling between nuclei and electrons. This means defining a coupling term, to be added to Equation 2.29, called the non-adiabatic coupling (NAC) operator, $\Lambda_{nm}(\mathbf{R})$:

$$\Lambda_{nm}(\mathbf{R}) = \frac{1}{2M}\left(G_{nm}(\mathbf{R}) + 2\mathbf{F}_{nm}(\mathbf{R})\nabla_{\mathbf{R}}\right) \quad , \tag{2.30}$$

with G_{nm} representing a scalar coupling term,

$$G_{nm}(\mathbf{R}) = \left\langle \phi_n^{ad} \middle| \nabla_{\mathbf{R}}^2 \phi_m^{ad} \right\rangle \quad , \tag{2.31}$$

and $\mathbf{F}_{nm}$ is the derivative coupling matrix, which can be written in a form that follows from the Hellmann-Feynman theorem,

$$\begin{aligned} F_{nm}(\mathbf{R}) &= \left\langle \phi_n^{ad} \middle| \nabla_{\mathbf{R}} \phi_m^{ad} \right\rangle \\ &= \frac{\left\langle \phi_n^{ad} \right| \left(\nabla_{\mathbf{R}} \hat{H}^{el}(\mathbf{R}) \right) \left| \phi_m^{ad} \right\rangle}{E_m^{el}(\mathbf{R}) - E_n^{el}(\mathbf{R})} \quad , \text{ for } m \neq n \quad . \end{aligned} \tag{2.32}$$

Inserting the NAC term in Equation 2.29 yields the full nuclear Schrödinger Equation:

$$i\hbar \frac{\partial}{\partial t} \left| \chi_m^{ad}(\mathbf{R}, t) \right\rangle = \left[T^{nuc}(\mathbf{R}) + E_m^{el}(\mathbf{R}) \right] \left| \chi_m^{ad}(\mathbf{R}, t) \right\rangle \; - \; \sum_n \Lambda_{nm}(\mathbf{R}) \left| \chi_m^{ad}(\mathbf{R}, t) \right\rangle \quad . \tag{2.33}$$

It is easy to imagine that including the coupling term in any kind of computational algorithm that aims to solve this full equation may become expensive, hence why it is useful to ignore it where appropriate. However, the reason that it is usually not possible to ignore the NACs when studying photochemical reactions, is clear from the denominator of Equation 2.32. As the energies of two states m and n approach each other, the magnitude of the derivative coupling increases and cannot be ignored. One can imagine a situation where, if two states become degenerate, a singularity is caused by the denominator approaching zero. This situation, long thought to be a rare occurrence, is now known to be the norm rather than the exception in dynamical molecular systems.[22,23,91–96] Such degeneracies are termed Conical Intersections (CoIns), and are described in more detail in the next section. First however, a new

electronic basis is introduced which avoids the issue of singularities by getting rid of the NAC terms (note that they are not ignored, but simply folded into the new basis). Such a basis can be obtained from the original adiabatic one, although not trivially. First, a subset of states from the infinite manifold needs to be selected, effectively truncating the Hilbert space of electronic states. To this subset of N adiabatic states, a unitary transformation, $\mathbf{S}(\mathbf{R})$, is applied which rotates them into a so-called diabatic representation:

$$\phi_i^d = \sum_{m=1}^{N} \phi_m^{ad} S_{mi}(\mathbf{R}) \quad . \tag{2.34}$$

Similarly, the adiabatic nuclear wavefunctions can also be rotated into the diabatic picture:

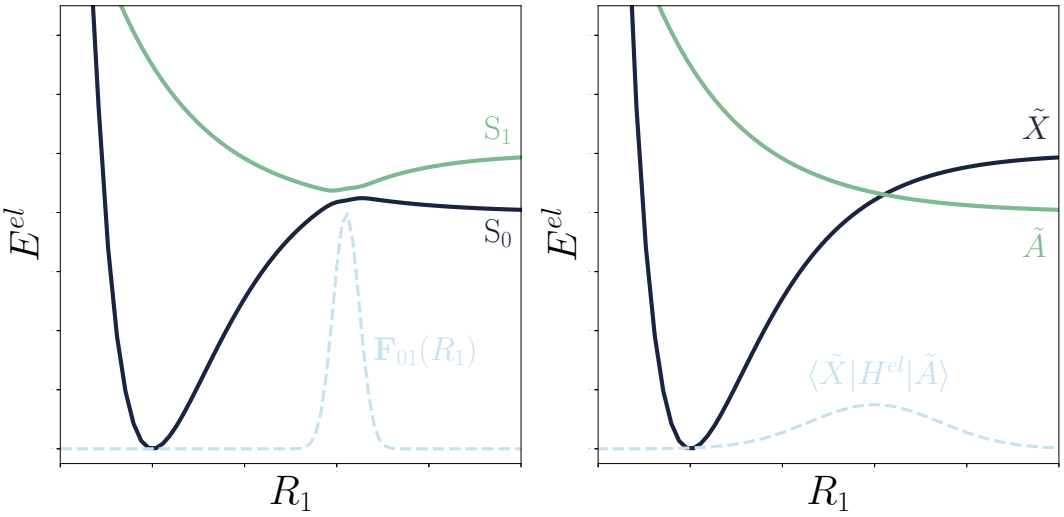
$$\chi_i^d = \sum_{m=1}^{N} S_{im}^{\dagger}(\mathbf{R})\ \chi_m^{ad} \quad . \tag{2.35}$$

The matrix, $\mathbf{S}$, used to transform between the two representations, is suitably called the adiabatic-to-diabatic transformation matrix (ADT). A few comments are in order, the first one being that the diabatic basis functions are not eigenfunctions of the electronic Hamiltonian, meaning that the diabatic Hamiltonian matrix, $H^d(\mathbf{R})$, may have off-diagonal elements, called the potential couplings,

$$H_{mn}^d(\mathbf{R}) = \left\langle \phi_n^d \right| H^{el}(\mathbf{R}) \left| \phi_m^d \right\rangle \quad . \tag{2.36}$$

Next, it is unfortunately impossible to define an exact diabatic basis for molecules containing more than two atoms.[97] This is a result of the so-called extended curl condition, which can only be satisfied for the complete space of electronic excitations, and hence cannot formally be obtained in a truncated subspace. Hence, the NACs will not go exactly to zero and need to be approximated as being completely gone (usually called a *quasi*-diabatic representation). The difference between the diabatic and adiabatic representations is illustrated in Figure 2.2, which shows the crossing of two PESs. In this case, the adiabatic surfaces are not able to become degenerate and hence lead to an avoided crossing, which is a situation that occurs as a result of the so-called *non-crossing rule*.[98] For two adiabatic surfaces to become exactly degenerate, two independent conditions need to be met with respect to the two electronic states, ϕ, at a particular nuclear geometry, R:

$$V_n(R) = V_m(R) \Leftrightarrow \begin{cases} \langle \phi_n | H^{el}(R) | \phi_n \rangle = \langle \phi_m | H^{el}(R) | \phi_m \rangle \\ \langle \phi_n | H^{el}(R) | \phi_m \rangle = 0 \end{cases} \quad . \tag{2.37}$$



(a) Adiabatic (b) Diabatic

Figure 2.2: The crossing of two PESs in (a) the adiabatic picture, leading to an avoided crossing, and (b) the diabatic picture, showing the smooth crossing of the surfaces. The magnitude of the coupling terms are shown in light blue.

An avoided crossing occurs when at least one of these conditions is not met. Diatomics for example, can never simultaneously fulfil both conditions and will therefore never exhibit adiabatic degeneracies. More interestingly however, are cases in which adiabatic states *are* able to become equal in energy, leading to CoIns.

Note on spin-orbit coupling

The inclusion of NAC terms in the Hamiltonian allows different states to couple together and hence interact, but only when the states in question have the same spin multiplicity. However, this selection rule may be relaxed in some cases due to a relativistic effect known as spin-orbit coupling (SOC), first observed in the 1940's by Kasha, and others.[99,100] SOC arises due to the interaction between the magnetic field generated by moving electrons, and their spin. The most important consequence of this is that transitions between electronic states of different spin multiplicities become weakly allowed. This phenomenon is known as intersystem crossing (ISC), and is analogous to internal conversion between states of the same multiplicity, facilitated by NAC.

The possibility of spin-forbidden processes gives rise to a whole class of chemical reactions proceeding via changes in spin-state.[17,101] Many of these reactions involve transition metals since strong SOC is most often associated with heavy elements.[102–104]

Nevertheless, it is also possible for smaller organic molecules to exhibit non-negligible SOC, especially when heavier elements such as sulphur or selenium are present.[105,106] Thus, including the effect of SOC in QD simulations may be important, even when only studying relatively simple organic molecules. Although this effect is not prevalent in the present work, there is one example in chapter 5 where SOC was explicitly included in the simulation of a sulphur-containing molecule, and ISC is observed. Therefore, a brief comment is made about how to account for the possibility of spin-forbidden processes in non-adiabatic QD simulations.

The SOC terms need to be of the form $\langle\phi_m|\,\hat{H}^{SOC}\,|\phi_n\rangle$, and can be readily calculated by many electronic structure methods. However, a subtlety arises related to how these terms are included in the nuclear dynamics. The first option is to add the effect of SOC as a correction to the propagation, essentially introducing off-diagonal elements in the system Hamiltonian, which ensures that the electronic states remain spin-pure eigenstates. The dynamics are then said to proceed in a spin-diabatic basis. Alternatively, a spin-adiabatic basis can be used where the SOC terms are folded into a diagonalised Hamiltonian, leading to spin-impure electronic states. More thorough descriptions of these different bases are given elsewhere, but are not relevant here.[107–109] In this work, when SOC effects are included, it will always be in the spin-diabatic representation, which has the advantage of being chemically more intuitive.

2.3.2 Conical Intersections

CoIns were long thought to be exotic features, but are now known to play an important role in photochemical dynamics by providing efficient pathways to radiationless decay in the form of internal conversion. The last few decades of quantum chemical computations have solidified their place as ubiquitous features of excited state PESs of molecular systems. A comprehensive theoretical description of CoIns can be found in the early work of Köppel, Domcke and Cederbaum,[110] which will be briefly summarised here. For clarity, a system with only 2 states is considered, but in principle, any number of states could have points of degeneracy.

The diabatic potential matrix of a 2-state ($|1\rangle$ and $|2\rangle$) system is written as follows

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \quad , \tag{2.38}$$

which can be diagonalised to obtain the adiabatic potential energies, V_1 and V_2. Ac-

cording to the non-crossing rule, states $|1\rangle$ and $|2\rangle$ can only become degenerate provided that $W_{11} - W_{22} = 0$ and $W_{12} = W_{21} = 0$. Now, to investigate what happens at such a point of degeneracy, one can first take the diabatic and adiabatic representations to be equal at this point, in which case the matrix elements of $\mathbf{W}$ are simply

$$W_{ij} = \langle \phi_i | \, H^{el} \, | \phi_j \rangle \quad . \tag{2.39}$$

Next, $\mathbf{W}$ can be expressed as a Taylor expansion (see Equation 1.1 in the Introduction) around a point R_0, where the degeneracy occurs:

$$\mathbf{W} = \mathbf{W}^{(0)}(R_0) + \mathbf{W}^{(1)}(\vec{R} - R_0) + \mathbf{W}^{(2)}(\vec{R} - R_0) + ... \quad , \tag{2.40}$$

where $\vec{R}$ represents a displacement vector relative to R_0. The zero$^{\text{th}}$ order term is now a diagonal potential matrix, with $\mathbf{W}^{(0)}(R_0) = \mathbf{V}(R_0)$ and $V_1(R_0) = V_2(R_0)$. The first order expansion term contains on-diagonal elements representative of the gradient of the potential, labelled κ, and off-diagonal elements corresponding to the derivative coupling between the states, labelled λ:

$$\kappa_i(R) = \langle \phi_i | \, \frac{\partial H^{el}}{\partial R} \, | \phi_i \rangle \quad , \tag{2.41}$$

and

$$\lambda_{ij}(R) = \langle \phi_i | \, \frac{\partial H^{el}}{\partial R} \, | \phi_j \rangle \quad . \tag{2.42}$$

Substituting these expressions in the first order expansion of $\mathbf{W}$, one finally obtains:

$$\mathbf{W} = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} + \begin{pmatrix} \kappa_1 \cdot \vec{R} & \lambda_{12} \cdot \vec{R} \\ \lambda_{21} \cdot \vec{R} & \kappa_2 \cdot \vec{R} \end{pmatrix} \quad . \tag{2.43}$$

From this equation, two vectors are defined; the gradient difference vector, $\mathbf{g} = (\kappa_2 - \kappa_1)$, and the derivative coupling vector, $\mathbf{h} = \lambda_{12}$. These are used to reformulate the non-crossing rule, which leads to some physical insights:

$$V_1(\vec{R}) = V_2(\vec{R}) \Leftrightarrow \begin{cases} \mathbf{g} \cdot \vec{R} = 0 \\ \mathbf{h} \cdot \vec{R} = 0 \end{cases} \quad . \tag{2.44}$$

In geometric terms, this condition for degeneracy tells us that the two states will remain degenerate so long as $\vec{R}$ is orthogonal to $\mathbf{g}$ and $\mathbf{h}$. If however, $\vec{R}$ falls within the so-called branching plane $\{\mathbf{g}, \mathbf{h}\}$, the degeneracy will be lifted.

Schematically, the description of a 2-state intersection given above can be represented by a double-cone, where the lower and upper part represent the first and

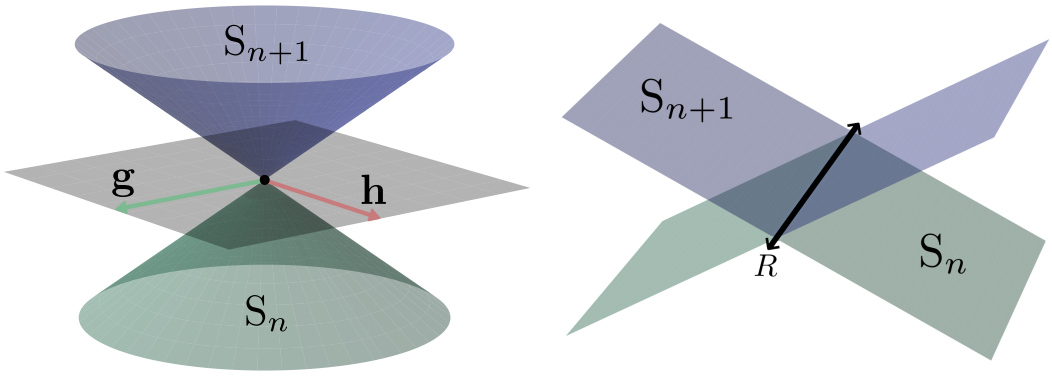


Figure 2.3: Schematic representation of 2 neighbouring adiabatic states, S_n and S_{n+1}, at (a) a conical intersection, and (b) an intersection seam. The branching plane in (a) is defined by two orthogonal vectors $\mathbf{g}$ and $\mathbf{h}$, and in (b) the coordinate along the seam $R \notin \{\mathbf{g}, \mathbf{h}\}$.

second adiabatic states, respectively. From this diagram, given in Figure 2.3a, the term "Conical Intersection" seems to make much sense. However, it is important to remember that this singular crossing point only occurs in the branching plane. In the full-dimensional reality, two crossing states will form an intersection seam in all orthogonal directions to the branching plane (Figure 2.3b).

2.3.3 Vibronic Coupling Hamiltonian

The ideas presented above help to understand how CoIns arise, but they can also be applied in a practical sense to derive one of the most widely used types of model Hamiltonians in molecular photophysics and/or photochemistry: the vibronic coupling Hamiltonian. In effect, vibronic coupling models corresponds to Taylor expansions of Equation 2.43, but applied to the full molecular Hamiltonian matrix, $\mathbf{H}^{mol}$:

$$\mathbf{H}^{mol}(\mathbf{Q}) = \mathbf{H}^{(0)}(\mathbf{Q}) + \mathbf{W}^{(0)}(\mathbf{Q}) + \mathbf{W}^{(1)}(\mathbf{Q}) + ... \quad , \tag{2.45}$$

where the coordinates $\mathbf{Q} = \{q_1, q_2, ..., q_{3N-6}\}$ are the $3N-6$ dimensionless normal modes ($3N-5$ for linear molecules), taken as such to simplify the expression of the kinetic energy operator. As before, it is assumed that at the point around which the expansion is done, the diabatic and adiabatic basis is the same, which allows the zero$^{\text{th}}$ order Hamiltonian to be expressed using the following separation of kinetic and ground state potential terms:

$$\mathbf{H}^{(0)}(\mathbf{Q}) = \mathbf{T}_N(\mathbf{Q}) + \mathbf{V}_0(\mathbf{Q}) \quad , \tag{2.46}$$

which are both expressed as a simple harmonic oscillator, parametrised by the harmonic vibrational frequency of the α^{th} normal mode. Equation 2.46 is thus re-written as

$$\mathbf{H}^{(0)}(\mathbf{Q}) = -\frac{1}{2}\sum_{\alpha}\omega_{\alpha}\frac{\partial^2}{\partial q_{\alpha}^2} + \frac{1}{2}\sum_{\alpha}\omega_{\alpha}q_{\alpha}^2 \quad . \tag{2.47}$$

The zero$^{\text{th}}$ order $\mathbf{W}$ term, which is still the diagonal potential energy matrix, defines the excited state surfaces by vertically displacing the ground state potential, V_0, by the excitation energy of the different states. To horizontally displace the harmonic oscillators, the linear (on-diagonal) first order terms are used, whereas the non-adiabatic coupling between states is given by the off-diagonal terms in $\mathbf{W}^{(1)}$. Both these first order terms are the same ones introduced in Equation 2.43, which take the following form in normal mode coordinates, for two electronic states n an m:

$$W_{nn}^{(1)}(\mathbf{Q}) = \sum_{\alpha}\kappa_{\alpha,n}q_{\alpha} \quad , \tag{2.48}$$

$$W_{nm}^{(1)}(\mathbf{Q}) = \sum_{\alpha}\lambda_{\alpha,nm}q_{\alpha} \ , \ n \neq m \quad . \tag{2.49}$$

If only first order terms are included in the Taylor expansion of Equation 2.45, the resulting operator is the familiar Linear Vibronic Coupling (LVC) Hamiltonian. Truncating up to second order gives rise to a Quadratic Vibronic Coupling (QVC) model, and so on.

There are several benefits of using this type of LVC model to parametrise a Hamiltonian. One is the simple form of the kinetic energy operator when normal mode coordinates are used. Another relates to molecular symmetry, allowing the number of terms in Equations 2.48 and 2.49 to be drastically reduced in some cases. Based on group theoretic arguments, it is possible to determine *a priori* which modes will give rise to non-zero values of the Hamiltonian parameters. An integral is only non-zero if the integrand function is totally symmetric, meaning the following rules must apply for the linear parameters to be non-zero:

$$\kappa_{\alpha,n} \neq 0 \Leftrightarrow \Gamma_{\alpha} \supset \Gamma_A \quad , \tag{2.50}$$

and

$$\lambda_{\alpha,nm} \neq 0 \Leftrightarrow \Gamma_n \otimes \Gamma_{\alpha} \otimes \Gamma_m \supset \Gamma_A \quad . \tag{2.51}$$

Here, A is the totally symmetric irreducible representation of the point group of the molecule. One can imagine that in highly symmetric molecules the number of vanishing

linear terms is significant and will greatly reduce the computational effort of evaluating the full Hamiltonian matrix (e.g. a LVC model of benzene gives rise to 796 linear terms, only 93 of which are non-zero thanks to its D_{6h} symmetry[111]). This fact makes the LVC model ideally suited for use in grid-based QD calculations.

2.3.4 Wavepackets & Initial conditions

A final general theoretical consideration is given, before moving on to describe specific methods for solving the TDSE. Since the work in this thesis is concerned with excited state processes, it is appropriate to first formalise the notion of photoexcitation. The Franck-Condon principle is taken as a starting point, stating that no major nuclear rearrangement takes place when an electron is excited.[112] This is known as a vertical excitation. The point at which the excitation takes place is usually taken to be the ground state equilibrium geometry, and is often referred to as the Franck-Condon (FC) point, or region. These two terms will hereafter be used interchangeably.

Putting this in quantum-mechanical terms, a molecule in its electronic ground state equilibrium is represented by a stationary wavefunction of a given vibrational state, ν, $|\psi_\nu(x_0)\rangle$. At time $t = 0$, a photon is absorbed by the system and the molecule is vertically excited into some higher lying electronic state s, according to the FC principle. On this state and at the FC geometry, a non-stationary (i.e. time-dependent) wavepacket, $|\Phi_s(x_0, t)\rangle$, is constructed as a superposition of stationary states, given by the vibrational eigenfunctions of the ground state wavefunction with energy $= E_\nu$, and a time evolution factor:

$$|\Phi_s(x_0, t)\rangle = \sum_\nu c_\nu \, |\psi_\nu(x_0)\rangle \; \mathrm{e}^{-iE_\nu t/\hbar} \quad . \tag{2.52}$$

Note that all possible $|\psi_\nu\rangle$ are eigenfunctions of the molecular Hamiltonian, and since linearity applies, the wavepacket, $|\Phi_s(x_0, t)\rangle$, is necessarily an eigenfunction as well.[113] The initial excitation, which defines the starting point for QD simulations, is given by the condition

$$|\Phi_s(x_0, 0)\rangle = \hat{\mu}_{0s} \, |\psi_\nu(x_0)\rangle \quad , \tag{2.53}$$

where $\hat{\mu}_{0s}$ is the transition dipole operator between the ground and excited state (for an ideal vertical excitation, $\hat{\mu}_{0s} = 1$). This process of photoexcitation followed by the evolution of the wavepacket is illustrated in Figure 2.4.

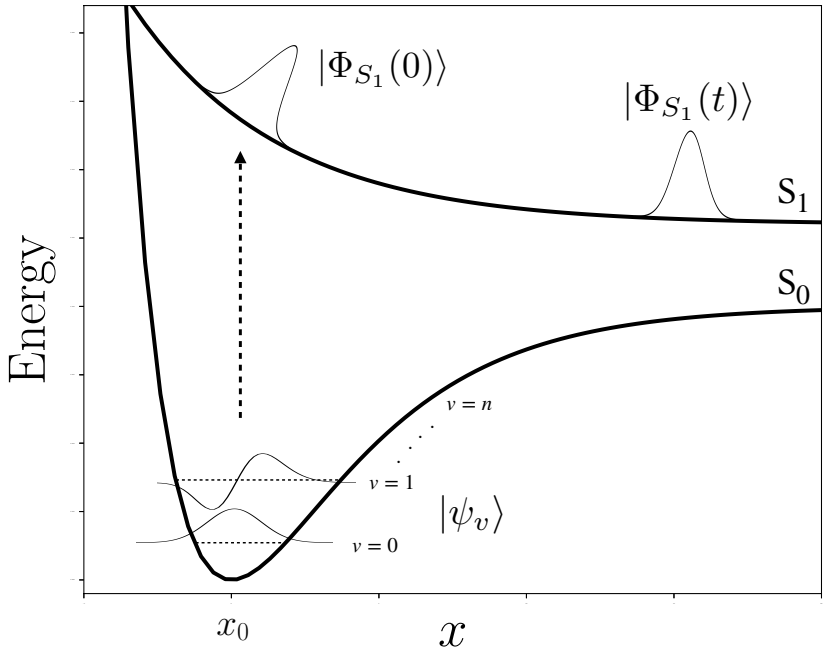


Figure 2.4: Excitation of an initial wavefunction $|\psi_v\rangle$ in its electronic ground state (S_0) to an electronic excited state (S_1), leading to photodissociation of the wavepacket $|\Phi_s(t)\rangle$. x is a nuclear coordinate (e.g. bond stretching) and x_0 the ground state equilibrium geometry.

In practice, these ideas can be applied in a number of different ways, often making approximations in terms of which initial wavefunction is excited, or how to construct the wavepacket on the final state. It is important to consider these initial conditions since they can have drastic effects on the outcome of QD calculations.[114,115] Especially in the context of simulating realistic experiments, preparing the correct initial state of a system is crucial. In that case, the condition of Equation 2.53 is no longer exactly valid since it assumes an infinitely short excitation pulse, which cannot realistically be produced.[115] Explicitly considering the excitation pulse in QD simulations is not part of the work presented in this thesis, but it is an active area of research and different suggestions have been made on how to properly account for different pulse shapes.[116–120]

Summarising this introductory backgound section, some fundamental ideas are presented, from the BOA and its breakdown, to the theoretical foundations of CoIns, the formulation of simple vibronic coupling Hamiltonians, and the creation of wavepackets. These are all crucial historical developments that allowed QD calculations to be pushed to their limit and applied to study real molecules in their full dimensionality. More comprehensive accounts of these concepts, and all the details that were omitted

here, can be found in refs. [110,113,115,121–123]. Having established this background, some methods are now discussed which allow one to actually solve the TDSE and perform the QD calculations which are foundational to the research presented in subsequent chapters.

2.3.5 MCTDH

At present, the Multi-Configurational Time-Dependent Hartree (MCTDH) method has been around for over 3 decades, with the original work having been published in 1990 by Meyer, Manthe and Cederbaum.[124] Since then, this approach has been widely used to accurately model the time evolution of realistic molecular systems of ever increasing size. The theoretical basis and applications of the method has been extensively reviewed in articles[125–127] and textbooks,[128–130] to which one is referred for all details not provided in what follows. The focus here will be on the reason for the success of MCTDH over its predecessors, and the current limitations of the method. As is often the case, MCTDH was an improvement on existing techniques because it allowed for the reduction in computational cost while still remaining numerically exact, and hence larger systems could be studied with high accuracy.

To understand the nature of this improvement, the method on which MCTDH is based must first be understood. It is commonly referred to as the "standard method" for solving the TDSE, and is equivalent to a Full CI electronic structure calculation in the sense that it includes all possible configurations within a given basis. The full wavefunction in the standard method is expanded as a direct product of time-independent nuclear basis functions, $K(\mathbf{x})$:

$$\Psi(\mathbf{x},t) = \sum_{j_1 \cdots j_f} C_{j_1 \cdots j_f}(t) \prod_{\alpha=1}^{f} K_{j_\alpha}^{(\alpha)}(x_\alpha) \quad , \tag{2.54}$$

where j_α is the set of basis functions used for the α^{th} DoF and f is the total number of DoFs. A summation over all electronic states is implied here, as well as throughout the rest of this section. The problem with such an expression is that to propagate this wavefunction, the number of coupled equations that needs to be solved is exponentially dependent on the number of DoFs, as well as the basis size, which needs to be relatively large (at least 10 basis functions per DoF, i.e. 10^f in total). In MCTDH, a very similar ansatz is used, which still scales exponentially, but the size of the basis is contracted

and hence the method can be pushed to higher dimensional systems. The contraction scheme takes the following form,

$$\Psi(\mathbf{Q},t) = \sum_{j_1 \ldots j_p} A_{j_1 \ldots j_p}(t) \prod_{\alpha=1}^{p} \varphi_{j_\alpha}^{(\alpha)}(Q_\alpha, t) \quad , \tag{2.55}$$

where the f sets of time-independent functions are replaced by p sets of orthonormal, time-dependent basis functions $\{\varphi^{(i)}\}$, called Single Particle Functions (SPF). The Hartree product of these SPFs is what is meant by a "configuration" (i.e. the "C" in MCTDH). Additionally, the nuclear coordinates are also replaced by p sets of composite coordinates, $\{\mathbf{Q}\}$, where $Q_\alpha = \{q_{\alpha_1}, q_{\alpha_2}, ..., q_{\alpha_d}\}$ and $\{q\}$ is often chosen to be the set of normal mode coordinates. The SPFs themselves are then expanded in a time-independent basis of so-called primitive basis functions, $\chi(Q_\alpha)$,

$$\varphi(Q_\alpha, t) = \sum_{k_1 \ldots k_d} a_{k_1 \ldots k_d}^{(\alpha)}(t) \prod_{v=1}^{d} \chi_{k_v}^{(v)}(q_v) \quad . \tag{2.56}$$

These functions are in fact the same basis functions as those used in the standard method (Equation 2.54), thus making the connection between both methods clear; MCTDH is simply a contraction scheme for the nuclear wavefunction. The primitive basis is used to form the product grid, which is populated by the SPFs. This leads to an effective dimensionality reduction because there can be fewer SPFs than the size of the grid.

The primitive basis is generally expressed in terms of a Discrete Variable Representation (DVR), both in MCTDH and in the standard wavepacket propagation method.[131,132] Many different DVRs have been developed, usually tailored to specific coordinates. The harmonic oscillator DVR for example, is used for vibrational motion. Another type of primitive basis function that will be specifically mentioned here because of its importance in the context of this thesis is the use of Gaussian functions. Replacing some of the SPFs in Equation 2.55 by Gaussian functions leads to the *G*-MCTDH method, developed by Burghardt and co-workers to model system-bath interactions.[133] The vMCG method described in the next section also uses such a Gaussian basis and hence a straightforward comparison between the two is possible, and will indeed be shown in chapter 4.

The MCTDH equations of motion are outside the scope of this discussion, but it is noted that they are variational and therefore conserve the total norm and energy, and

lead to favourable convergence properties.[126] Furthermore, the level of convergence in MCTDH calculations can be properly determined by checking, for example, the wavefunction density at the edges of the grid, which should be negligible. Another important metric for convergence is the so-called natural weight of the underlying basis functions, indicating the relative level of importance for each individual function. For converged calculations, the weight of the least important function should be small. Finally, the most rigorous check for convergence is comparing two wavefunctions from calculations using a different number of SPFs (but same primitive basis), and verifying that the overlap is close to unity.

The power of MCTDH was first demonstrated when a simulation of pyrazine was performed using all 24 normal modes, and an accurate absorption spectrum was produced.[134] Although impressive at the time, it is not possible to go to much larger systems than that, even with present day computers. However, the development of a Multi-Layer (ML) version of MCTDH has provided ways to study systems of thousands of DoFs. It works in much the same way as standard MCTDH, but uses layers of time-dependent, multi-dimensional SPFs which are expanded each time in a direct product, until the final layer constitutes the time-independent grid, equivalent to Equation 2.56.[135,136] Note here that one can again choose the primitive grid to be populated with Gaussian functions, yielding the ML-*G*-MCTDH method.[137,138]

Finally, the drawbacks of the MCTDH family of methods need to be discussed. Simply put, the definition of the Hamiltonian poses practical challenges. The first hurdle is that the kinetic energy operator is coordinate dependent and therefore potentially system-dependent. For this reason, a common choice is to use mass-frequency weighted normal modes because this massively simplifies the expression (see the first term of Equation 2.47). Next, the potential on which the wavefunction is propagated needs to be defined globally across the grid, which grows exponentially with each added dimension. Additionally, for MCTDH to work efficiently, the analytical expression of the potential, V, needs to be in a specific separable form (i.e. sum-of-products):

$$V^{separable}(\mathbf{Q}) = \sum_{j_1 \ldots j_p} c_{j_1 \ldots j_p} \prod_{\alpha=1}^{p} v_{j_\alpha}(Q_\alpha) \quad . \tag{2.57}$$

Conveniently, a Taylor expression has this form, making the LVC Hamiltonian described previously a suitable choice for representing the potential. Nevertheless, the

need for such an analytical description hinders the study of larger, more complex, realistic molecular systems because it would require the evaluation of energies, gradients and NACs at many points to cover the configurational space adequately, followed by a procedure to fit multi-dimensional functions to those points. This very quickly becomes impractical, if not impossible. To overcome this challenge, there are two possible avenues one may take; make the calculation of the required properties more efficient and improve fitting procedures, or design a QD method that does not rely on the pre-computation of the potential. In the following section, an example of the latter approach is discussed.

2.3.6 vMCG

The method discussed here is referred to as variational Multi-Configurational Gaussian (vMCG). Its name derives from the fact that Gaussian basis functions are used, rather than SPFs, to express the wavefunction. This change offers a way to remove a constraint of grid-based approaches, namely that the potential has to be known in advance. Avoiding this requirement is a major practical advantage, and has motivated the development of several different methods that exploit the useful properties of Gaussian functions in order to efficiently propagate a wavefunction. These functions are used to construct Gaussian Wavepackets (GWPs) which can be propagated in different ways depending on the specific method. Examples of such methods include Full and Ab Initio Multiple Spawning (FMS[139,140] and AIMS,[141] respectively), Coupled Coherent States (CCS)[142] and Multi-Configurational Ehrenfest (MCE),[143,144] just to name a few. They all have in common that the wavepackets follow classical trajectories. In the vMCG method on the other hand, the functions follow variationally optimal trajectories which, in principle, makes convergence faster (meaning a smaller number of basis functions is necessary) and increases accuracy.

The wavefunction in the vMCG method is expanded in the usual way, as a product of coefficients and basis functions. The basis consists of Gaussian wavepackets, which are time-dependent, as are the expansion coefficients, A. This expansion leads to a linear combination of multi-dimensional GWPs, G_i, given by the following ansatz:

$$\Psi(\mathbf{x}, t) = \sum_{i=1}^{N} A_i(t) G_i(\mathbf{x}, t) \quad . \tag{2.58}$$

Here, N is the total number of GWPs and a summation over all electronic states is still implied. The wavepackets themselves are separable into a product of one-dimensional functions,

$$G_i(\mathbf{x},t) = \prod_\alpha g_i(x_\alpha,t) \quad , \tag{2.59}$$

that each take the form of a general Gaussian function:

$$g_i(x_\alpha,t) = \exp\left(-\zeta_{i\alpha}x_\alpha^2 + \xi_{i\alpha}(t)x_\alpha + \eta_i(t)\right) \quad , \tag{2.60}$$

where ζ, ξ and η are the width, the linear and the scalar parameters, respectively, and α runs over the DoFs. The use of such functions in nuclear dynamics calculations is originally derived from the pioneering work of Heller in the late 1970's.[145,146] The vMCG formalism employs the Frozen Gaussian approximation, keeping the widths fixed and hence the time-dependence is included in the evolution of the linear and scalar parameters only.

The wavepackets in vMCG propagations obey a set of coupled EoMs, derived from the Dirac-Frenkel variational principle,

$$\langle\delta\Psi|H - i\hbar\frac{\partial}{\partial t}|\Psi\rangle = 0 \quad . \tag{2.61}$$

This ensures that the evolution of the GWPs is variationally optimal. Applying the principle to the wavefunction is done by varying either the coefficients, collected in a matrix $\mathbf{A}$, or the Gaussian parameters collected in $\mathbf{\Lambda}$ (where $\mathbf{\Lambda}_i = \{\boldsymbol{\zeta}_i, \boldsymbol{\xi}_i, \eta_i\}$), leading to two coupled equations of motion, written here in vector notation:

$$i\dot{\mathbf{A}} = \mathbf{S}^{-1}(\mathbf{H} - i\boldsymbol{\tau})\mathbf{A} \tag{2.62a}$$

$$i\dot{\mathbf{\Lambda}} = \mathbf{C}^{-1}\mathbf{Y} \quad . \tag{2.62b}$$

with the overlap matrix $S_{ij} = \langle G_i|G_j\rangle$, the Hamiltonian matrix $H_{ij} = \langle G_i|\,\hat{H}\,|G_j\rangle$ and the (time-)differential overlap matrix $\tau_{ij} = \left\langle G_i\middle|\frac{\partial G_j}{\partial\lambda_j}\dot{\lambda}_j\right\rangle$, where λ_j is a single parameter of the j^{th} Gaussian. Note that the overlap matrix needs to be computed because the Gaussian basis is not orthonormal, which is a drawback of all Gaussian-based methods. The compact form of Equation 2.62b, when expanded, becomes much more complicated but will not be discussed in great length here. The original literature provides the necessary detail.[133,147–149] However, one point will be expanded upon since it will become important in chapter 4. Both $\mathbf{C}$ and $\mathbf{Y}$ contain the overlap matrix in

their expansion, which depends on the Gaussian parameters. This means that these overlaps can represent the Gaussian *moments* up to different orders. When using frozen width functions, as is the case here, the overlap matrix can only represent moments up to first order. In a simplified form, those moments are thus

$$M_{ij}^{(n)} = \langle G_i | \, x^n \, | G_j \rangle \quad , \quad \text{for } n = 0, 1 \quad . \tag{2.63}$$

This representation is important because part of the Hamiltonian can be expressed as a combination of these moments and a vector of linear Gaussian coefficients $\mathbf{X}$. This vector does not contain terms which mix different sets of GWPs and hence is said to collect all classical terms, which do not couple the basis functions together. Using this partition of terms, $\mathbf{Y}$ can now be re-written as

$$\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{Y}_R \quad , \tag{2.64}$$

where $\mathbf{Y}_R$ contains the residual part of the Hamiltonian described by the higher order moments (i.e. $n \geq 2$). This expression, inserted in Equation 2.62b, gives the following variation of the EoM for the Gaussian parameters,

$$i\dot{\mathbf{\Lambda}} = \mathbf{X} + \mathbf{C}^{-1}\mathbf{Y}_R \quad . \tag{2.65}$$

The above scheme is referred to as the "CX" formalism, and provides advantages such as improving the numerical stability of the propagation.[147] One other major advantage relates to the separation of classical and quantum terms, contained separately in $\mathbf{X}$ and $\mathbf{C}^{-1}\mathbf{Y}_R$ respectively. To clarify this point, consider the EoM of the position, q, and momentum, p, of the basis functions, written according to the "CX" formalism:

$$\dot{\mathbf{q}} = \frac{\mathbf{p}}{M} + \frac{1}{2\zeta}\mathrm{Im}\left[\mathbf{C}^{-1}\mathbf{Y}_R\right] \tag{2.66a}$$

$$\dot{\mathbf{p}} = -\frac{\partial V(\mathbf{x})}{\partial \mathbf{x}} + \mathrm{Re}\left[\mathbf{C}^{-1}\mathbf{Y}_R\right] \quad . \tag{2.66b}$$

Notice that the first term in each expression is the corresponding classical EoM.[150] This result is further discussed in chapter 4, where its use as an approximation in vMCG dynamics is evaluated.

2.3.7 Direct dynamics

The vMCG method as described above is closely related to MCTDH, and thus suffers from many of the same drawbacks. For example, a primitive grid is still required

which implies a pre-computed potential. However, the method is now perfectly suited to replace its fixed Gaussian basis set by a flexible, dynamic Gaussian basis. In that case, the potential can be obtained during the nuclear propagation itself, and only in regions visited by the moving wavepackets.[151] This method is known as the direct dynamics version of vMCG, i.e. DD-vMCG, first developed in 2004 by Worth *et al.*.[148] The use of Gaussian functions in direct, or *on-the-fly*, QD methods has long been exploited in semi-classical methods such as AIMS, and it is known that reaching a converged result using such approaches often requires too large a number of functions to be feasible. To make things worse, even defining suitable convergence criteria within this framework is not trivial. In principle then, a fully coupled, variational, Gaussian-based direct dynamics method such as DD-vMCG should require fewer basis functions to converge, since the trajectories followed by the individual wavepackets are optimal. In reality however, vMCG doesn't have well defined convergence criteria either, but it does demonstrably converge to MCTDH,[152] making it possible to assess convergence by comparison.[153]

The representation of the potential in DD-vMCG is now briefly discussed. Naturally, the analytical potentials provided by, for example, a LVC model are not applicable in direct dynamics approaches. Rather, the potential will be locally defined, and subsequently updated, by QC calculations (this is the meaning of the term *on-the-fly*, or *direct* dynamics). Hence, the form of the potential should be compatible with the information obtained from QC. Conveniently, a suitable expression can once again take the form of a second-order Taylor expansion around the centre (q_0) of the i^{th} basis function, G_i, at time t:

$$V^{(i)}(\mathbf{q}) = \mathbf{V}(q_0) + \sum_{\alpha} \frac{\partial \mathbf{V}(q_0)}{\partial q_\alpha} \cdot (q_\alpha - q_0) + \frac{1}{2} \sum_{\alpha\beta} (q_\beta - q_0) \cdot \frac{\partial^2 \mathbf{V}(q_0)}{\partial q_\alpha q_\beta} \cdot (q_\alpha - q_0) \quad , \quad (2.67)$$

with α and β representing different normal modes. This widely used representation is known as the Local Harmonic Approximation (LHA). Its elements are readily obtained from standard QC calculations in the form of the adiabatic potential energies $\mathbf{V}$, the nuclear gradients $\mathbf{V}'$, and the force constants $\mathbf{V}''$. As suggested by its name, the potential given by the LHA is locally approximated to be harmonic. If and when this becomes a bad approximation of the true potential, a new electronic structure calculation is carried out to update and correct the approximation. The criteria which define whether or not the LHA holds are discussed below. Before that, a comment is

given on how the LHA is used to define the initial conditions of a simulation.

To generate a wavepacket on an excited state, one starts by calculating the ground state frequencies, which are then used to construct the ground state potential using the LHA. Assuming the molecule is entirely in its lowest vibrational state (a reasonable assumption given that simulations are typically performed at 0K) and at the FC point, the eigenfunction of this state is projected, vertically, onto the specified excited state. This excited wavepacket is represented by a number of GWPs. To ensure an exact representation of the wavepacket, the basis functions are distributed in momentum space, but their positions are all centered at the FC point. Initially, one GWP carries 100% of the population, which can be transferred to other wavepackets upon starting the propagation.

Since its original implementation in the QUANTICS package, the DD-vMCG method has been further developed to include many improvements. Notably, the use of a diabatisation scheme, called *propagation diabatisation*,[154] has improved the usability of the approach and made it possible to simulate relatively complex systems for which many excited states need to be computed. Remembering that the potential energies given by electronic structure calculations are the adiabatic ones, such schemes are a necessity. The basic principle is to propagate, along with the wavepackets, the adiabatic-diabatic transformation matrix, $\mathbf{S}$, using the following relationship,[155]

$$\nabla \mathbf{S} = -\mathbf{F}\mathbf{S} \quad , \tag{2.68}$$

where $\mathbf{F}$ is the derivative coupling matrix from equation 2.32. This expression allows one to determine $\mathbf{S}$ at any coordinate $\mathbf{x} + \Delta\mathbf{x}$, from a previously obtained $\mathbf{S}(\mathbf{x})$. A question naturally arises concerning the definition of $\mathbf{S}$ at $t = 0$. This is referred to as setting the *gauge* for the propagation, i.e. defining a point at which the NACs vanish. The gauge in DD-vMCG always takes the diabatic and adiabatic representations to be equivalent at the FC point. Although this choice is somewhat arbitrary, it is assumed to be as good as any other possible gauge since defining exact diabatic states is never possible in more than one dimension, so an approximate diabatisation (i.e. *quasi*-diabatisation) is performed either way.[122,156] The use of diabatisation-propagation has been shown to work relatively well in previous studies on the photochemistry of small organic molecules,[147,157–159] and hence its validity will be assumed (a critical analysis of the diabats obtained in this way is given in chapter 3, as an example of how to

evaluate the quality of the diabatisation).

Moving on to another major improvement of DD-vMCG; the introduction of the QC *database*. With this feature, the output of every *on-the-fly* calculation performed during the propagation is stored as a separate record in a SQL database. Each record can contain the following elements:

- Geometry (Cartesian and Normal modes)
- Potential energies (diabatic and adiabatic)
- Nuclear gradients (diabatic and adiabatic)
- Hessian matrix (diabatic and adiabatic)
- Non-adiabatic coupling vectors
- Spin-orbit coupling vectors.
- Adiabatic-diabatic transformation matrix
- Molecular Orbital coefficients

As different wavepackets explore different parts of the configuration space during a DD-vMCG simulation, the database is consulted at every time-step to check that the potential at a given point is still valid. This evaluation is made based on a distance criterion between the centre coordinate of a given GWP and the closest point that was previously calculated and stored in the database. If this distance exceeds a given threshold value, a new QC calculation is performed at that geometry and the database is updated accordingly. In this way, the database gets progressively more populated during a propagation, as the potential energy landscape is explored and computed.

Since the information in the database represents a collection of single points, there is a question of how to represent a continuous surface based on those points. Such a representation is necessary to allow for wavepacket propagation. One approach is fitting functional forms to the points, but this may require much effort and would go against the principle of direct dynamics. Instead, a Shepard interpolation is performed between each point, achieving a global PES given by a weighted sum of Taylor expansions around the calculated points.[160] As such, the potential at any point, $V(\mathbf{x})$, is

given as

$$V(\mathbf{x}) = \sum_{i=1}^{N_{Db}} w_i(\mathbf{x}_i) T_i(\mathbf{x}_i) \quad , \tag{2.69}$$

where N_{Db} is the total number of records (calculated points) in a given database. The Taylor expansion around each point, $T(\mathbf{x})$, is typically done to second order, and the weights, $w(\mathbf{x})$, are calculated according to a function based on the distance between $\mathbf{x}$ and $\mathbf{x}_i$. This scheme gives the DD-vMCG method its flexibility and ultimately reduces the total number of explicitly calculated points required for accurate propagations, hence massively reducing the overall computational cost. One of the main drawbacks of such an interpolation however, is that it becomes impossible to demonstrate convergence in the same way as can be done using grid-based methods, i.e. based on the populations across the entire regular grid. To see why, imagine the interpolated surface as a PES evaluated across the same regular grid in configuration space. There will be regions that have not been explored by any wavepackets during the propagation, and hence no QC calculations will have been performed there. The description of the surface in those regions is given by an interpolation of distant points, which will be very poor. This is not expected to influence the quality of the nuclear dynamics themselves, but it does render the use of traditional grid-based convergence criteria rather unreliable.

On a different note, some advantages of the combination of the database and the interpolation between points are discussed. Firstly, a new electronic structure calculation does not necessarily need to be performed at every time-step, but only when a wavepacket encounters a region where the potential is poorly described. Compared to other methods, this means that fewer calculations are performed overall. Next, a very useful feature of having a database of points is that one can use exactly the same data to run different QD calculations. For example, one might want to compare DD-vMCG simulations with different numbers of basis functions, or using different EoMs. Going a step further, one could use an entirely different *on-the-fly* method (e.g. TSH) to propagate the nuclear basis functions on exactly the same underlying potential given by a database. This is very valuable in the context of developing common benchmarks for the entire non-adiabatic dynamics community (see chapter 5 for a discussion of the development of benchmarks). One more advantage worth mentioning is that, in principle, all the points in the database could be used in a fitting procedure to obtain

analytical expressions for the PESs. These could then be used in more traditional, grid-based dynamics methods to, once again, allow proper comparison between QD methods (also briefly discussed in chapter 5). It is worth pointing out that despite its many advantages, the use of a QC database has not been adopted by other *on-the-fly* non-adiabatic QD methods, where instead the data parametrising the potential is thrown away after the simulation has finished.

Lastly, an advantage of using variationally coupled basis functions is highlighted. Generally speaking, the trajectories will optimally explore the relevant configuration space and thus a smaller basis set is typically required, as compared to semi-classical Gaussian-based methods for example. This then means that fewer QC calculations will have to be performed, which could lead to a major reduction in computational cost, or allow for the use of more expensive electronic structure methods. However, it should be said that despite this point, DD-vMCG is never expected to be competitive to other methods in terms of cost. The idea is that it makes up for this weakness with its superior accuracy.

2.3.8 Note on TSH

The Trajectory Surface Hopping (TSH) method is for the non-adiabatic dynamics community what DFT is for electronic structure theory. This analogy refers not the theoretical foundations of the methods, but to the fact that TSH, like DFT, is the most widely used method in its field thanks to its relative ease of use and its ability to tackle large systems. Hence, a brief description ought to be given which outlines the main features and shortcomings. Note also that in chapter 4, some results obtained with TSH dynamics are presented and compared directly to MCTDH and vMCG.

In a seminal paper by Tully from 1990 (incidentally, the same year as the first MCTDH paper), he developed a stochastic algorithm that could be used to simulate the complex excited state processes induced by photoexcitation of molecules.[143] This approach, named Fewest Switches TSH (FSSH), was an extension of previous work on trajectory based dynamics.[161] The central idea was the introduction of a "hopping" probability from one electronic state to another during nuclear dynamics simulations, in order to approximate non-adiabatic phenomena. The method relies on solving the electronic TISE to obtain the PESs, but uses classical EoMs to propagate the position of an atom with mass M_A, as individual trajectories according to Newton's second law

($F = ma$):

$$M_A \frac{\partial^2 \mathbf{R}_A}{\partial t^2} = -\nabla_A E_m^{el}(\mathbf{R}_A, t) \quad , \tag{2.70}$$

where $\mathbf{R}_A$ still represents a nuclear coordinate and the term on the right-hand side is the negative gradient – the force – of the PES of electronic state m. This approximation is exact in the classical limit of $\hbar \to 0$. A quantum wavepacket is approximated by N_{traj} trajectories, each with different initial conditions and completely independent of one another, moving on one adiabatic surface at a time. At each time-step, a probability is calculated (based on NACs and electronic coefficients) to allow the possibility of a given trajectory hopping to a different adiabatic state. In this respect, many different approaches have been developed which define different hopping conditions, but historically the most popular method remains FSSH.

Since its inception, many ongoing developments have been made to address the limitations of classically approximating nuclear motion. Among the most notable limitations is proper sampling of the initial conditions. This is usually done by Wigner sampling of the ground state,[162,163] but is known to be inadequate for some systems.[164] Next, the classical nature of the motion along PESs means nuclear tunnelling effects cannot be described.[165] Another well-known topic in the surface hopping community is the decoherence problem, arising from the bifurcation of the nuclear wavepacket.[166] These problems can all be individually corrected for, but no universal surface hopping algorithm exists that avoids all of them.[167] Again, this is reminiscent of the problem of choosing the right functional in DFT.

A further drawback which poses a more philosophical problem is the fact that the TSH formalism has not been rigorously derived from the Schrödinger equation. One can therefore make an argument about the *ad hoc* nature of the scheme, making it potentially unreliable by being more prone to obtaining the right answer for the wrong reasons.[168] In conclusion then, this method for non-adiabatic dynamics is very powerful and can be applied to a wide range of systems, but it is important to properly benchmark one's results or at least critically evaluate the possible inaccuracies introduced by the independent classical trajectory approximation. Finally, making once again a link with DFT, if one can afford the use of more expensive but accurate methods, it may be more appropriate to avoid TSH.

2.4 Summary

This chapter served to introduce, outline and explain the theoretical background behind both electronic structure methods and approaches to compute the dynamics of nuclei, coupled with their electrons. Although an emphasis is placed on the latter, both topics need to be covered before presenting any results in the later chapters. Even in the field of dynamics, it is important to appreciate the wide range of electronic structure methods available to computational chemists, and the reasons behind their various advantages and drawbacks. After all, one cannot describe how molecules move without knowing their properties in different parts of the configuration space. The theoretical considerations of photochemical reactions given in the first chapter also highlights this fact, and it should now be clear that some methods are well suited to this challenge, and some are not.

Notice also that only wavefunction based approaches have been discussed in any detail. There are other formalisms to approximate the electronic structure of molecules, but the ones presented here are the most applicable in the present case. Of particular importance in this thesis are the multi-configurational methods, specifically MCSCF. Their advantages have been outlined and it is clear that they are invaluable for the description of dissociating molecules (their use for that purpose will be presented in chapters 3 and 5). It should however be noted that the accuracy and flexibility offered by such methods comes at a high computational and practical cost, the latter often being the more problematic of the two.

Having established the tools for time-independent calculations, the history and early development of the field of non-adiabatic molecular dynamics was introduced. This discussion covers the theoretical foundation of much of this thesis, and will be referred back to in later chapters. In a regime where one can no longer rely on the BOA, it is shown how the nuclear Schrödinger equation is transformed to include non-adiabatic coupling terms which describe the inter-state interactions in a manifold of coupled electronic states. The ubiquitous role of CoIns is introduced and the difficulty in describing them theoretically is explained. The vibronic coupling Hamiltonian is shown to be a simple but effective way of describing the complex interplay between nuclear and electronic motion, allowing efficient grid-based simulations of molecules. The theoretical treatment of photo-excitations is also briefly outlined.

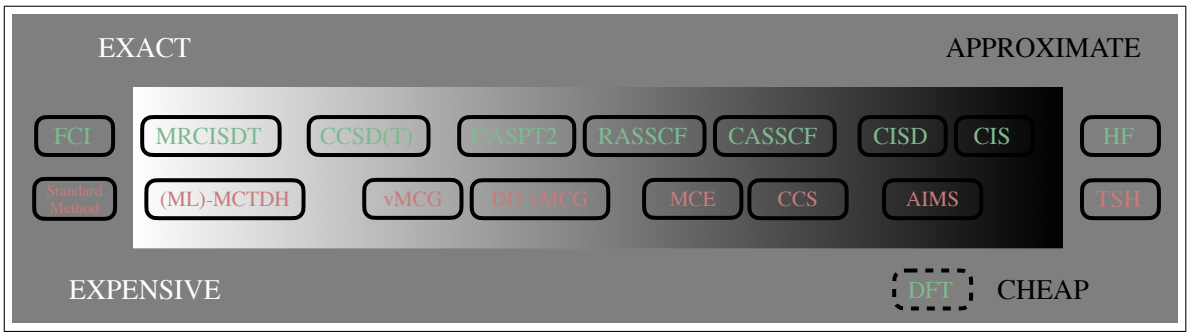


Figure 2.5: Spectrum of accuracy vs cost with electronic structure methods in green, and nuclear dynamics methods in red.

Following the account of these fundamental concepts in the field, an overview is given of some practical QD methods used in this work. As with electronic structure theories, one is faced with many different algorithms, which can be based on completely different formalisms. Again, the focus lies on those methods that are particularly relevant in the context of the following chapters, as opposed to giving an accurate account of all flavours of QD. A natural starting point is the well-known MCTDH method, which sparked a lot of the early interest in the field. Additionally, the two most relevant methods, vMCG and DD-vMCG, derive directly from the MCTDH ansatz. Although the theoretical foundations of these methods is clear, they are much younger than MCTDH meaning that there are still many open questions related to their use in molecular photochemistry. Because they have been used in relatively few studies to date, much remains to be discovered in terms of their properties such as convergence, numerical stability, accuracy, and cost. With respect to this final point, it should be clear that, on the spectrum of cost vs accuracy, these methods lie nearly all the way at the accurate but costly side. They are thus applied to small molecules in chapters 3 and 5, but following recent developments described in chapter 4, much larger systems are tackled as well.

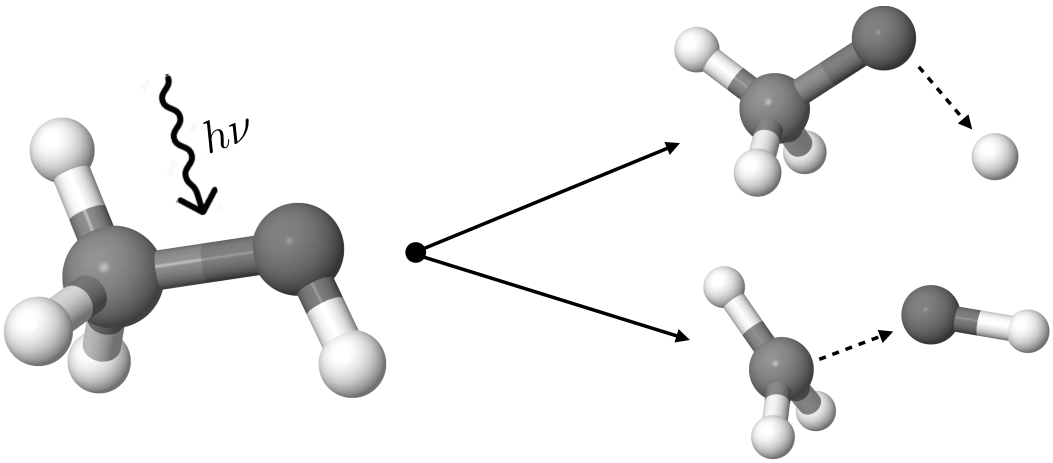
Taken together, the two parts of this chapter provide an overview of the tools required to accurately describe photophysical and photochemical phenomena. Although both the electronic structure section and the nuclear dynamics one start with well established theoretical foundations, it is important to appreciate that the current state of the art is still very much evolving. A good way to think about the nature of the current method developments happening in both fields is illustrated in Figure 2.5, where the electronic structure and nuclear dynamics methods are placed on a sim-

plified spectrum of accuracy vs cost (note that DFT is an outlier since it is cheap, but its accuracy depends strongly on the chosen functional). In this way, it becomes clear that new developments consist of exploring the enormous grey area that exists between (near) exact methods, and more approximate ones. As such, it becomes very important to keep up to date on new developments as they may allow the treatment of systems that were previously inaccessible.



CHAPTER 3

Modelling photodissociation



3.1 Motivation & Background

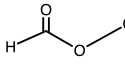
This chapter presents the results of a comprehensive quantum dynamics study on the photodissociation of methanol (CH_3OH). The primary aim of this chapter is to show how one can study photodissociation reactions using theoretical methods. Since methanol will be used as an example, a secondary aim is of course also to unravel the detailed mechanisms behind the dissociation reactions of this molecule. The motivation for the first aim is clear, in the sense that a proper description of dissociation processes (photoinduced or not) is desirable for understanding much of chemistry, which is concerned with reactions that, more often than not, break and/or form bonds. One could argue that the problem of calculating bond dissociation energies for the ground state of a molecule is essentially solved, at least for relatively small species. Several wavefunction-based methods exist that can deal with the multi-configurational nature of a stretched bond, even though these methods typically still require careful thought and consideration, and should not be taken for granted. In contrast, it would not be fair to say that the calculation of electronically excited states across a range of configurations, including and especially ones with highly stretched bonds, is a solved problem which has become trivial. In fact, quite the contrary. Furthermore, even if an accurate set of excited state potentials is obtained, the question still remains: what are the consequences for a system when it finds itself in a particular excited state?

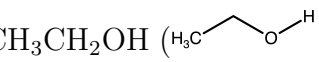
By definition, an excited system is out-of-equilibrium and will therefore try to lose excess energy by relaxing to a lower energy state. Methods exist to calculate the minimum energy path between the initially excited molecule and known ground state products. This constitutes essentially a steepest descent calculation and allows one to characterise photochemical reaction mechanisms in an analogous way to standard "thermal" chemistry (i.e. Reactant → Transition State → Product). Two important limitations of this method are apparent. Firstly, at least some of the so-called "stationary points" of the reaction have to be known in advance (from previous investigations), and secondly, the minimum energy path assumes the molecule has no kinetic energy. In more realistic situations, large deviations from the minimum energy path are not only possible, but expected.[169]

Resolving both these drawbacks and providing a detailed, mechanistic answer to the previously posed question requires a dynamical treatment of the nuclear and electronic

motion. Since excited states are involved, such a simulation must also take into account the potential coupling between nuclear and electronic motion, i.e. it cannot rely on the BOA. The non-adiabatic QD methods discussed in chapter 2 are capable of doing exactly that, and hence this study provides an example of how to use them (specifically, the (DD)-vMCG method). How to set up the system initially and how to ultimately extract useful data is explained and shown by presenting simulations of photo-excited methanol.

Why methanol? Organic chemists will recognise this small molecule as a commonly used solvent in the laboratory, as well as in industry, but probably don't spend much time thinking about its properties beyond polarity or freezing/boiling point. In recent years, methanol is also being talked about more and more in the context of alternative fuels for the global shipping industry for example, due to its desirable combustion properties as well as the possibility to source it from sustainable sources (so-called "green methanol"). However interesting or promising this may be, it was not the motivation for this study. Computational chemists might look at methanol as "the first alcohol", i.e. the smallest carbon-based molecule containing the OH group. In that context, studying its behaviour may provide insights into the properties of all alcohols by extrapolating the conclusions to larger, more complex molecules that don't lend themselves so easily to computational investigations. While it is true that methanol's small size played a role in the decision to use it for this study, allowing computationally expensive methods to be used, it was ultimately an astronomical observation that motivated the choice.

In a review paper on astrochemistry published by Öberg in 2016, she highlights the fact that chemical complexity can build up in the interstellar medium through photochemical reactions occurring on small ice particles.[170] These icy grains are composed mainly of water but also contain a layer rich in CO, CO_2 and, perhaps surprisingly, CH_3OH. Located in outer space and therefore not shielded by an atmosphere, these grains are constantly bombarded by radiation, including, significantly, UV light in the range of 100-200 nm. When these photons are absorbed by small organic molecules like methanol, reactive radicals are formed and a number of chain reactions can follow ultimately resulting in the formation of larger molecules such as $HCOOCH_3$ (),

CH_3OCH_3 () or CH_3CH_2OH ().[171] Such reactions have been investigated experimentally by recreating ices in the appropriate conditions (10-20K and under high vacuum) and irradiating them with UV, while monitoring the photoproducts being formed. Obviously it is very difficult to exactly match interstellar conditions in the lab but nevertheless, these studies are able to provide some insights into the reaction kinetics on the surface of these ices. For example, the branching ratios of the original species that made up the ice can be estimated from the abundances of the final photoproducts, .

Arriving now at the main point of this discussion, it has been estimated that in methanol-rich ices the photolysis of the methyl hydrogen (producing H + CH_2OH) was about 5 times more likely than the production of methoxy radicals via H + CH_3O. This observation is very surprising when one compares this with the quantum yields observed for the gas-phase photodissociation of methanol (results summarised in Table 3.1). The reason for this matrix-dependent difference in the yields of the

Table 3.1: Summary of experimentally observed branching ratios of methanol at different excitation energies in the gas-phase, and estimated branching ratios in UV irradiated ice.

λ (nm)	**State**	**CH_3O + H**	**CH_2OH + H**	**CH_3 + OH**	**CH_2O + 2H**
193[172]	S_1	0.86	< 0.01	0.14	
157[173]	S_n[a]	0.57	0.26		0.17
122[174]	S_n	0.18	0.11		0.71
Methanol in interstellar ice (approximated)					
115-170[171]	S_n	0.17	0.83	< 0.1	

[a] "S_n" refers to any electronic state higher than S_1.

various dissociation products is not understood and is therefore worthy of study since the final products are determined by the initial fragmentation. In a big-picture sense, understanding the chemistry on these interstellar ice grains may help answer some of the many questions surrounding the origin of complex chemical structures, such as those responsible for life on earth. To explain the differences between the gas- and condensed-phase dissociation reactions of methanol, one must model both and compare the respective mechanisms of fragmentation. The work presented here is thus the first step towards this goal, i.e. modelling the gas-phase photodissociation of methanol.

In addition to being astrochemically relevant, methanol also plays a role in the

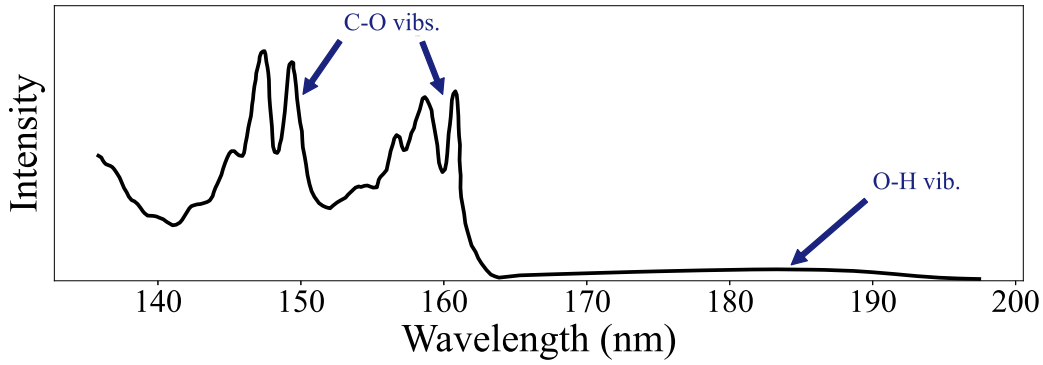


Figure 3.1: Experimental gas-phase UV absorption spectrum of methanol, reproduced from data in ref. [173].

chemistry of our atmosphere, providing another reason for its potential research interest. Unsurprisingly then, this is not the first time that the photochemistry of methanol is being investigated, whether experimentally or computationally. On the experimental side, a range of spectroscopic techniques such as photofragment translational spectroscopy and time resolved photoelectron spectroscopy have been used to study the dissociation characteristics at different UV wavelengths.[173–183] These experiments served to fairly unambiguously characterise the absorption spectrum, reproduced in Figure 3.1. It shows one fairly weak, broad and continuous feature centered around 184 nm which corresponds to a transition to a 3s Rydberg state resulting in the fast dissociation of the hydroxyl hydrogen and giving rise to this typical signature of rapid "ballistic" dissociation. In the region between 140-165 nm, a finer vibronic progression is observed which is associated with transitions to 3p Rydberg orbitals and corresponds mainly to vibrational states of the C–O bond.[179] Furthermore, possible topographies of the excited state PESs have been inferred from these experiments, which conclude that unbound surfaces must exist along both the O–H and the C–O bond stretching coordinates, corresponding to possible dissociation channels.[174] No such channels have been identified for the C–H bond breaking. In fact, the C–H bond has only been reported to dissociate on long timescales (picoseconds), either after the cation is created, or through the $O(^1D)$ + CH_4 reaction.[176] Hence, these reactions are outside the scope of the present investigation since only the ultrafast direct dissociation dynamics on unbound states will be of interest here.

Complementary to these experimental studies, various computational investigations have also been performed.[179,180,184–188] As mentioned, the small size of methanol

allows for the use of fairly high level electronic structure methods. As early as 1984, Buenker *et.al.*[188] performed multi-reference double-excitation CI calculations to study the photodissociation of methanol at 185 nm. The potential energy surfaces clearly show the that the first excited state is repulsive along the O–H bond coordinate (i.e. the state is unbound). The surfaces also confirm that this same state is repulsive along the C–O coordinate, although a small energetic barrier is present which could explain why the O–H bond is preferentially broken, compared to the C–O bond. The second excited state is shown to be bound along both of these coordinates. These results were later confirmed by TD-DFT and CASSCF calculations.[180,184–186]

This characterisation of different excited state surfaces has provided a qualitative explanation for the observed branching ratios in the gas phase. Still lacking however, are time-dependent calculations that allow for a more quantitative and mechanistic picture of the photophysical and photochemical processes occurring upon excitation. To this end, the only examples of dynamical studies are a trajectory surface hopping simulation which seems to be severely undersampled, with only 35 trajectories yielding dissociation products,[185] and a recent study employing quasi-classical trajectory simulations to investigate the dynamics on the ground state,[187] based on previously published ground state PESs.[189,190] These simulations employed over 45,000 trajectories but the work does not overlap with the present research question related to the excited state dynamics.

In conclusion, the excited state dynamics of methanol are of interest in the context of photodissociation reactions in astrochemical and atmospherical environments. Although the relevant dissociation channels have been determined in previous studies, no quantum dynamical description exists of the photophysical and photochemical processes occurring in the manifold of low-lying electronic states following excitation in the UV range. This chapter aims to fill this gap by providing high dimensional non-adiabatic QD simulations, including the 8 lowest excited states of methanol after excitation to its first, second and third electronic states. The quantum yields of the various dissociation channels are calculated and rationalised based on the calculated surfaces and the state population dynamics. A rather lengthy and detailed description of the work is provided here, but a more concise account of the procedure, results and conclusions can be found in a paper titled *"Modelling Photodissociation: Quantum Dynamics Simulations of Methanol"*, published in J. Phys. Chem. A, in 2024.[191]

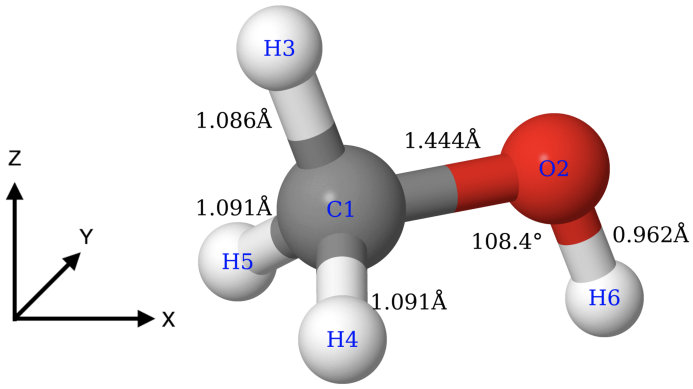


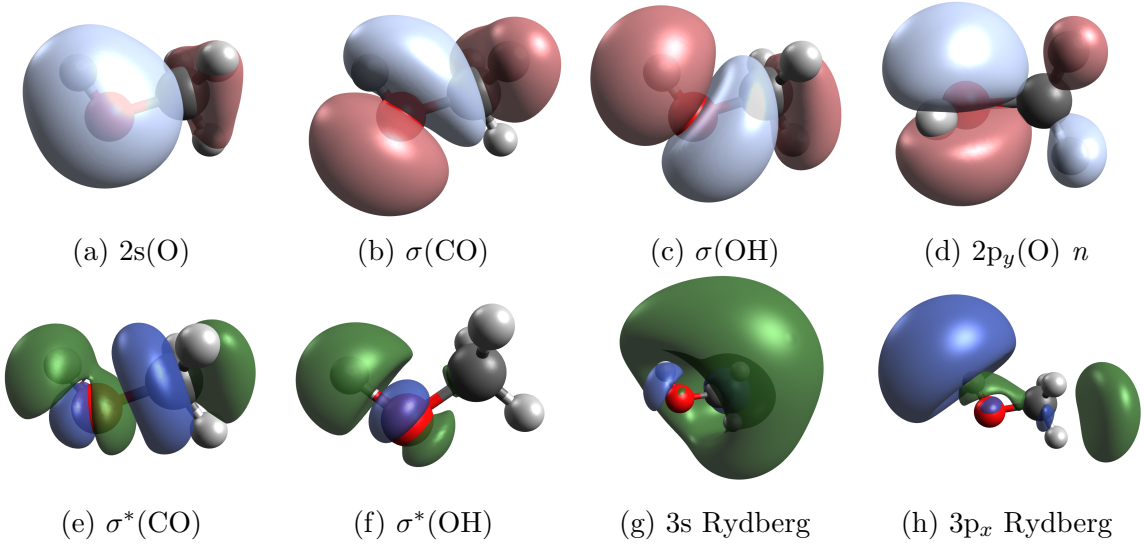
Figure 3.2: Optimised geometry of methanol in C_s symmetry, using CCSD//aug-cc-pVDZ.

3.2 Computational detail

The reliable CCSD method was used to optimise the geometry of the ground state, shown in Figure 3.2. The corresponding Cartesian coordinates can be found in Appendix A (Table A.1), along with the calculated frequencies and descriptions of the $3N-6$ normal modes (Table A.2).

3.2.1 Potential energy surfaces

Having characterised the molecule at its equilibrium ground state geometry, the search for an adequate description of the configurational space of the first few low-lying electronic states was started. Faced with a large number of possible electronic structure methods, one should first think about what needs to be calculated and hence rationalise which methods might be able to achieve these calculations. In the present case, bond-dissociations need to be described which already implies the need for an electronic wavefunction which includes some multi-configurational character. Methanol, only consisting of 6 atoms, lends itself to the use of more expensive methods which can achieve high accuracy. Thus, active space based methods were the obvious choice, although thorough benchmarking against other methods was performed (see section 3.3.1). The ultimate "winner" of the comparative study was the RASSCF method, which is related to the more widely used CASSCF method (see chapter 2, section 2.2.3 for the theoretical background). Looking at the orbitals of the RAS2 space (Figure 3.3), i.e. the subspace in which all possible electronic configurations are computed, one clearly sees how the chosen orbitals fit with the research question. Specifically, to



(a) 2s(O) (b) σ(CO) (c) σ(OH) (d) 2p$_y$(O) n

(e) σ^*(CO) (f) σ^*(OH) (g) 3s Rydberg (h) 3p$_x$ Rydberg

Figure 3.3: Molecular orbitals of the RAS2 space. (a - d) are initially occupied whereas (e - h) are unoccupied.

describe the dissociation of the O–H and the C–O bonds, the respective bonding and anti-bonding σ orbitals need to be included. Additionally, the traditional "frontier orbitals" are included, i.e. the lone pair on the oxygen, which is the HOMO, and two Rydberg states, the LUMO and LUMO+1. All together, the 8 orbitals in the RAS2 space were complemented by 4 additional orbitals (not shown), 2 occupied and 2 unoccupied ones, forming the RAS1 and RAS3 space, respectively. A total of 12 electrons were included in the active spaces to arrive at the following definition of the calculation performed: RASSCF(12,2+8+2)[1,1] (the notation RASSCF($n, m^I + m^{II} + m^{III}$) $[h, e]$ refers to n active electrons in m^I RAS1 orbitals with a maximum of h holes, m^{II} RAS2 orbitals and m^{III} RAS3 orbitals with a maximum of e electrons). As will be shown, these calculations are able to describe the potential energy landscape of the first few electronic states of methanol within the relevant configurational space.

A caveat that should be mentioned here is that, although the potential energies were calculated at RASSCF level (as implemented in OpenMolcas[192]), this method is not presently able to provide nuclear gradients or non-adiabatic coupling vectors. These properties are required during the nuclear dynamics and were thus calculated at the CASSCF level. The active space was defined by the 8 orbitals and 8 electrons of the RAS2 subspace and are therefore expected to be consistent with the RASSCF results. In any case, the implementation of the DD-vMCG method in Quantics (see chapter 2, section 2.3.6) ensures that as long as accurate energies are obtained from the electronic

structure calculations, the potentials on which the wavepackets will be propagated will be reasonably smooth. The potentials are ultimately obtained from an interpolation between the calculated points, and the couplings merely serve to represent the surfaces in the diabatic picture. The gradients are only used to correctly interpolate between the points. All in all, even though some inaccuracies are inevitably introduced by not being able to calculate all required properties at the exact same level of theory, the fact that the surfaces are almost entirely based on the RASSCF calculations means that the dynamics are not expected to suffer much under this hybrid approach. However, performing a proper evaluation of its performance would be valuable since it may open up doors to performing highly accurate potential energy calculations using one method, complemented by gradient and coupling calculations using a cheaper method. Lacking an obvious procedure to carry out this analysis, it remains an avenue to be explored in the future and is not further discussed.

3.2.2 Nuclear dynamics

The theory and potential applications of QD calculations have been discussed in chapter 2. Here a concrete example will be given of the practical aspects of setting up such calculations. Specifically, the use of vMCG, and its *on-the-fly* version DD-vMCG, will be presented. Despite the conceptual simplicity of these methods it is probably not surprising that there are many practical subtleties that limit their general usability. After all, theory and practice are the same in theory, but in practice they are not. This topic is further explored in chapter 5, while this section simply lays out the different steps that were taken in this particular investigation of methanol.

Starting with the vMCG method, the most important step is the formulation of the Hamiltonian. Mass-frequency scaled normal modes are used throughout, allowing the kinetic energy operator to be expressed in a straightforward manner:

$$\hat{T}^N(\mathbf{Q}) = -\frac{1}{2}\sum_{\alpha}\frac{\partial^2}{\partial q_\alpha^2} \quad . \tag{3.1}$$

The potential energy terms on the other hand, require more work. A LVC model, described in chapter 2, section 2.3.3, is used in this case. It was constructed using the VCHAM program, implemented in QUANTICS, which has been used in the past to fit vibronic coupling models for a variety of different molecules.[111,193–196] The data used for the fitting was obtained using the previously mentioned RASSCF calculations,

performed at regular intervals of 0.5 mass-frequency weighted normal mode units along the positive and negative directions of 4 key modes. Initially, a first order fitting was performed to obtain the linear parameters corresponding to the gradients and couplings (κ, and λ, respectively). Next, the diabatic potentials were fitted using higher order analytical functions including Morse curves and exponentials, as well as regular harmonic oscillators. A straightforward conjugate gradient method is used to iteratively optimise the variables defining each of these expressions.

The 4-mode model was used for 100 fs vMCG simulations with 128 GWPs. All Gaussian widths were set to $\frac{1}{\sqrt{2}}$, which corresponds to the exact ground state vibrational wavefunction within the harmonic approximation, when mass-frequency scaled normal modes are used. Based on the energies and oscillator strengths in Table 3.2, it was decided to initiate simulations on the first 3 excited states, and use the combined results to obtain an absorption spectrum (the next section outlines how this is achieved).

Table 3.2: Vertical excitation energies (eV) and oscillator strengths (unitless) of the first 7 excited states of methanol, obtained using RASSCF(12,2+8+2)[1,1]//aug-cc-pVDZ.

Label	**Energy**	**Osc. strength**
S_1	6.33	0.012
S_2	7.51	0.028
S_3	8.44	0.010
S_4	9.31	0.058
S_5	11.00	0.019
S_6	11.54	0.072
S_7	11.62	0.100

Moving on to the preparation of the DD-vMCG calculations, which do not require any prior fitting of surfaces and use the same kinetic energy operator. Despite this simplification, one must nevertheless be careful when setting up the system. Firstly, even though no primitive grid is used, there are still defined limits as to how far away from their initial position the wavepackets can travel. In a dissociating molecule such as methanol, it was necessary to extend this range along the coordinates corresponding to the dissociative motion. Additionally, the CH_3 rotational mode (Q_1) was frozen during the simulations. This is often done when using normal mode coordinates since the description of the curved path of the rotating atoms during torsional motion is not properly converted to cartesian coordinates because of the use of linear displacement

vectors.[197] The GWPs of this simulation are thus 11-dimensional and in total, 33 wavepackets were propagated, again with a width of $\frac{1}{\sqrt{2}}$.

The initial conditions are defined such that all wavepackets are centered at $q = 0$ for all modes (i.e. the FC point), and are vertically projected onto the desired electronic state. The vertical excitation of the molecular wavepacket is exactly represented by distributing the GWPs in momentum-space. Only one of the wavepackets is initially populated but as the simulation progresses, population density is quickly transferred to the others. It is also noted that the zero-point energy of the vibrational ground state is included in the excitation, giving the wavepackets some internal energy. The surfaces at this point are given by a LHA around the FC point, and are then updated once the Gaussians start moving. In practice, one tends to run multiple sequential *on-the-fly* simulations to gradually populate the database containing the calculated QC information. Once adequate surfaces are obtained, a final run is performed on the database without adding any new points. For the simulations presented here, a final database of about 1,500 points was needed for the surfaces to be properly described. It should be emphasised that thanks to the Shepard interpolation between all points, analytical expressions of the potential energy landscape are obtained which span much more space than only along the pure normal modes of the system.

In principle, the procedure described above is all that is necessary to initialise a DD-vMCG calculation. In practice, some system-dependent considerations may need to be made. When studying dissociation reactions for example, one will want to avoid the description of the molecule at very highly stretched geometries. The reason for this is two-fold. Firstly, electronic structure calculations are likely to break down if two molecular fragments are too far apart. The other reason is that as a fragment accelerates away from the parent species (via an unbound surface), the integrator will reduce the time-step of the simulation, causing the total simulation time to increase massively.[157,159] To avoid either of these scenarios, the dissociating wavepackets can be uncoupled from each other at a certain threshold value, causing them to behave classically. These classical GWPs then simply continue their motion without new QC points being generated, until their population drops to zero. This is achieved by defining a Complex Absorbing Potential (CAP) along dissociative coordinates, at carefully chosen positions, that acts as a dividing surface between a fragment being bound or being fully dissociated. CAPs are imaginary potentials, W, of the following

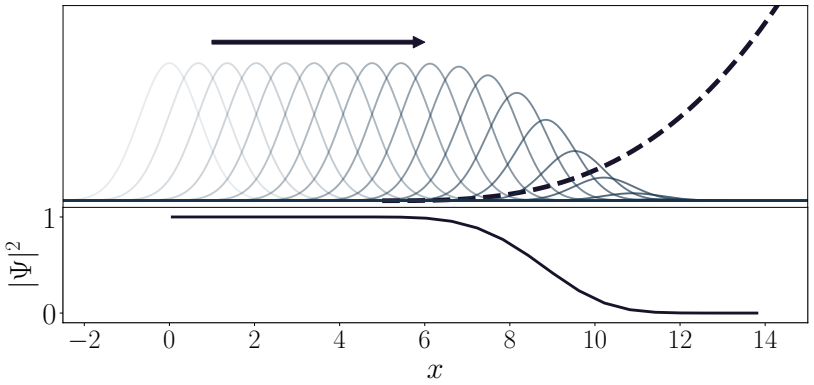


Figure 3.4: Time evolution of a 1-dimensional GWP (initial position $x = 0$) moving on a flat potential. A CAP is defined at $x = 5$ (dashed line), which gradually absorbs the incoming wavepacket. The norm of the wavefunction over time is plotted on the bottom panel.

form

$$iW = \eta\Theta(k(x - x_0))^n \quad , \tag{3.2}$$

with an order, n, a strength, η, and Θ is the Heaviside step function. The k parameter defines whether the CAP is positioned in the negative or positive direction with respect to its location, defined by x_0. Such potentials are also commonly used in grid-based QD methods, but in that case they are used to avoid unphysical reflections of wavepackets from the end-points of the grid.

To illustrate the effect of a CAP on an incoming wavepacket, a simple model system is set up; a 1-dimensional GWP moving on a flat potential, with some initial momentum, encounters a CAP (Figure 3.4). The wavepacket is gradually absorbed by the imaginary potential and consequently, the total norm of the wavefunction, $|\Psi|^2$, is ultimately reduced to zero. This same procedure was applied to the dissociative modes of methanol. All CAPs were parametrised in the same way, with $n = 3$, $\eta = 0.01$, and placed along the following modes (x_0 given in brackets, both in normal mode coordinates and in Å for bond lengths, and degrees for angles): O–H stretch (20 = 2.9 Å), C–O stretch (13 = 2.7 Å), and C–O–H angle bend (15 = 151°). In terms of the positions of the CAPs, one should make sure to place them far enough from the FC point to avoid interference with any wavepackets that are oscillating in a bound potential (i.e. vibrational motion). Note that it was necessary to put a CAP on the C–O–H angle since the dissociation of OH goes along with an opening of this angle. As with the model system, wavepackets that reach these CAPs will cause the norm of the total wavefunctions to drop. The simulations were run until $|\Psi|^2 \approx 0$, meaning that

the molecule has fully dissociated, or until a simulation time of 100 fs was reached.

In the end, three identical DD-vMCG calculations were prepared, each initiated from a different electronic state (S_1, S_2 or S_3). By convention, the adiabatic states will be labelled with the symbols S_0, S_1, S_2 ... whereas the diabatic states are denoted $\tilde{X}$, $\tilde{A}$, $\tilde{B}$ ... These are chosen to be equivalent at the FC point and can therefore be used interchangeably at that point. Note that in reality, excitations of molecules typically result in the partial population of multiple close-lying states, which is why it is always informative to simulate the dynamics from multiple initial states. In this way, it is also possible to extract wavelength-dependent properties (e.g. photolysis rates[198]) from a system, although this was not explicitly done here.

3.2.3 Analysis of the wavefunction

The discussion is now turned towards the problem of extracting insightful information from a wavefunction once it is obtained from successfully performing a QD simulation. This is not exactly trivial, and more attention will be given to this point in chapter 5. For now, the elementary theoretical basis for some commonly used analysis methods will be outlined, and subsequently used to extract useful data from the methanol calculations.

Given the principles of QM, it is not difficult to relate a given wavefunction to a measurable property. It boils down to a transformation from $\Psi \rightarrow$ *observables*, using suitable operators to compute expectation values. A quantum-mechanical observable need not always relate to a physical property however, and in fact, much mechanistic information can also be obtained by calculating expectation values of a range of operators that do not correspond to measurable observables. Therefore, the distinction will not be given much thought here.

Photoabsorption spectra

One of the most general and widely used analysis methods for photochemical simulations is the computation of an absorption spectrum, and comparing this with an experimentally obtained one. An obvious reason for this is the availability of spectra for a wide range of molecular species, collected over the last century or so. Incidentally, it is very straightforward to calculate such a spectrum from a wavefunction propagation. Although not the the only possibility, taking the Fourier transform of

the autocorrelation function is by far the most suitable method in the present context. The autocorrelation function at a given time, $C(t)$, is defined as

$$C(t) = \langle \Psi(0) | \Psi(t) \rangle \quad , \tag{3.3}$$

from which the spectrum, $\sigma(\epsilon)$, can be obtained in the desired energy range (by varying the value of ϵ):[113]

$$\sigma(\epsilon) \propto E \int_{-\infty}^{+\infty} \mathrm{d}t \; C(t) \; \mathrm{e}^{i(E_0+\epsilon)t} \quad . \tag{3.4}$$

The ground-state energy, E_0, is required since that is where the vertical photoexcitation takes place. Notice also that the integral is bound by infinity, whereas the simulation time is necessarily finite and strictly positive. This first point presents a problem that will introduce errors in the spectrum, except in cases where $C(t) \to 0$. Such situations occur during, for example, photodissociation dynamics. In more common, bound-state systems, the issue can be alleviated by introducing a damping function, which forces $C(t)$ to smoothly reach zero after a given time. Finally, the problem of integrating over negative time is solved by setting $C(-t) = C^*(t)$ (where * denotes the complex conjugate).

State populations

Calculating the change in density/population of individual electronic states can often reveal important mechanistic insights in photochemical processes. In theory, to obtain the population of a state m, P_m, from the total wavefunction, a projection operator, $\hat{P}_m = |\phi_m\rangle \langle \phi_m|$, can be defined and its expectation value calculated as

$$P_m = \langle \Psi | \, \hat{P}_m \, | \Psi \rangle \quad . \tag{3.5}$$

Here, the type of electronic state is of no importance. However, in methods such as MCTDH, (DD)-vMCG, and most other wavefunction-based QD methods, the propagation is carried out in the diabatic picture. The wavefunction can therefore be made to have state-specific components corresponding to individual diabatic states, which can be extracted. Thus, to calculate the diabatic state populations, it suffices to compute the norm of the component pertaining to a specific state (indicated by $^{(m)}$):

$$P_m^{diab} = |\Psi^{(m)}|^2 \quad . \tag{3.6}$$

Extracting adiabatic populations from methods that use a diabatic representation is also possible, but relies on approximate and non-trivial procedures that will not be discussed here.[153,199]

Finally, it is noted that since diabatic states are not eigenstates of the electronic Hamiltonian, their populations are not considered to be physical observables. Nevertheless, they will be shown to provide valuable insights.

Flux

Flux refers generally to the measure of a given quantity changing over time, per unit area. In this context, it is taken to be a measure of the amount of population per electronic state going into a given channel. Specifically, the channels of interest in this chapter are dissociation pathways. The flux going into such a pathway provides insights into the relative importance of different channels, and information on what electronic states are ultimately responsible for each of them. It is measured by first defining a dividing surface between bound and dissociative fragments and evaluating the population density crossing the surface over time. Thus, if the surface is placed along a dissociative coordinate, at a bond length long enough to assume full dissociation of the two fragments (i.e. where asymptotic motion is assumed), the density going into a specific photoproduct channel is obtained. The time-dependent flux $f(t)$ through these dividing surfaces is calculated for each individual state m as follows

$$f_m(t) = \frac{\partial}{\partial t} \langle \Psi(t) | \hat{P}_m \Theta(x(t) - x_d) \hat{P}_m | \Psi(t) \rangle \quad , \tag{3.7}$$

using the same projection operator, $\hat{P}_m$, as in Equation 3.5. Θ is the Heaviside step function and x_d is the position of the dividing surface. The probability, p_m, of dissociation via a given state is then given by the integrated flux over the total simulation time ($t = 0 \rightarrow t_f$):

$$p_m(t_f) = \int_0^{t_f} \mathrm{d}t \; f_m(t) \quad . \tag{3.8}$$

This expression is equivalent to the diabatic state population of m.[126]

Wavepacket trajectories & GGP

It is often desirable to analyse the nuclear motion of a molecule following a dynamics calculation. One of the most intuitive ways to interpret nuclear motion is to generate

trajectories in configuration space. In this context, Gaussian wavepacket methods such as vMCG have certain advantages over other wavefunction-based methods like MCTDH, in which the wavepackets are delocalised and cannot straightforwardly be interpreted in a classical way. Gaussian functions on the other hand, are localised around their centre coordinate and hence one can compute the expectation value of the position of that centre coordinate and obtain a "quantum trajectory" for individual GWPs. Importantly, these should be contrasted against the trajectories obtained from semi-classical QD methods, in which an ensemble of independent trajectories approximates a moving wavepacket. A statistical distribution is then used to obtain weights for the individual trajectories.

The corresponding weights for the quantum trajectories in the vMCG method cannot be obtained in the same way and thus a different, suitable metric needs to be defined that properly describes the contribution of individual wavepackets to the global wavefunction. One possibility would be to use the square norm of the coefficients, $|A_i(t)|^2$, to give the weight of the i^{th} GWP, were it not for the fact that the nuclear basis set is not orthogonal and thus causes problems due to non-zero overlaps between Gaussians. Instead, an approximate metric referred to as the Gross Gaussian Population (GGP) is used to provide the weights, following previous work.[200] Its calculation is derived from a normalisation requirement of the vMCG formalism, using the overlap matrix, $\mathbf{S}$, to account for the non-orthogonality of the basis functions:

$$\langle\Psi|\Psi\rangle = \sum_{ij} A_i^*(t)\ S_{ij}(t)\ A_j(t) = 1 \quad . \tag{3.9}$$

Dividing the overall density between the individual GWPs, one obtains

$$\mathrm{GGP}_i(t) = \mathrm{Re}\sum_j A_j^*(t)\ S_{ji}(t)\ A_i(t) \quad , \tag{3.10}$$

where only the real-valued part is taken since the imaginary components sum to zero. Note that a summation over all electronic states is implied, although state-specific weights can be obtained as well. In general, the expectation value for any quantity a can now be approximated as a weighted sum over the GGPs:

$$\langle a(t)\rangle = \sum_i \mathrm{GGP}_i(t)\ a_i(t) \quad . \tag{3.11}$$

In short, this metric allows for the proper quantitative determination of various properties obtained from a Gaussian wavepacket calculation.

3.3 Results

3.3.1 Time-independent $\phi^{el}(\mathbf{r};\mathbf{R})$ analysis

Firstly, the results of electronic structure calculations are discussed. The reason for performing these calculations prior to running any dynamics simulations is two-fold. On the one hand, it serves as an important validation of the chosen electronic structure method (RASSCF) which is necessary especially when using active space based methods, due to their non-"black box" nature. Secondly, careful analysis of potential energy cuts along key normal modes offers an intuitive, visual picture of important dissociative states that will likely play a role during the dynamics. Identifying the chemical character of those states further solidifies one's understanding of the mechanisms responsible for bond breaking. Such intuitions are invaluable when eventually analysing the outcome of QD simulations.

Electronic structure benchmark

Summarised in Table 3.3 are the predicted energies of several low-lying electronic states of methanol according to 5 different electronic structure calculations, all performed at the FC point. Experimental values are also given. The nature of the transition of each state is provided and makes it clear that not all methods predict exactly the same states. Specifically, the CASSCF, CASPT2 and RASSCF results seem to be missing 2 states when compared with the EOM-CCSD and the experimental results, corresponding to excitations to 3p Rydberg orbitals. Those states are not calculated simply because the active space does not include those particular 3p orbitals. If one does include them, the CASPT2* results are obtained and the two states are correctly captured (see 6^{th} column in the table). Naively, one may conclude from this observation that the best active space is the one which includes the extra two 3p orbitals since it allows for the reproduction of experimental values, as well as the highly accurate EOM-CCSD results. This reasoning however, ignores the nature of the problem at hand. The aim is to describe bond dissociations, which implies the need for a description of the molecule at geometries far away from the FC point. It follows that the process of bond breaking can, at least within a Molecular Orbital theory framework, only be described correctly if both the bonding and anti-bonding σ orbitals are included in the calculation. Hence, it is not appropriate to make judgements about the suitability of

Table 3.3: Vertical excitation energies of the lowest lying electronic states of methanol (in eV), calculated using the EOM-CCSD, CASSCF(12,12), CASPT2(12,12) and RASSCF (12,2+8+2)[1,1] methods, compared to experimental values. The aug-cc-pVDZ basis set was used for all calculations. The shaded column contains the data used in the dynamics simulations.

Label [a]	**Excitation**	**EOM – CCSD**	**CASSCF**	**CASPT2**	**CASPT2*** [b]	**RASSCF**	**Exp.** [179]
S_1	$2p_y$(O) n $\rightarrow$ 3s Ryd.	6.7014	6.3223	6.7981	6.5217	6.3343	6.7596
S_2	$2p_y$(O) n $\rightarrow$ $3p_x$ Ryd.	7.8981	7.4904	7.9547	7.6860	7.5056	7.7268
·	$2p_y$(O) n $\rightarrow$ $3p_y$ Ryd.	8.4861			8.2491		8.3133
·	$2p_y$(O) n $\rightarrow$ $3p_z$ Ryd.	8.5791			8.2637		8.3133
S_3	σ(OH) $\rightarrow$ 3s Ryd.	8.7349	8.4232	8.8367	8.6212	8.4374	/ [c]
S_4	σ(OH) $\rightarrow$ $3p_x$ Ryd.	9.7099	9.2984	9.7562	9.5900	9.3143	/
S_5	$2p_z$(C) $\rightarrow$ 3s Ryd.		10.9836	11.3136		11.0046	

[a] "S_n" refers to the adiabatic singlet states at the FC point of the RASSCF results. [b] Active space is changed to include all 3p Rydberg orbitals, instead of the σ^* orbitals. [c] Not explicitly seen experimentally due to low oscillator strength.

an active space solely based on calculations performed at the FC geometry, but rather one should perform energy scans along some bond stretching coordinate to adequately determine the validity of a given active space across a range of out-of-equilibrium geometries.

Such scans are presented in Figure 3.5, where the EOM-CCSD, CASPT2 and CASPT2* results are plotted along the two key coordinates of this study; the O–H and C–O bond stretching modes. Note that whereas the O–H coordinate corresponds to a pure normal mode, the C–O coordinate is defined by a linear combination of both the C–O bond stretching and the C–O–H angle bending normal modes, in order to reveal the dissociative channels which are a result of both bond elongation and angle opening. The two CASPT2 results (Figures 3.5b and 3.5c) clearly exemplify how instabilities in the active space lead to discontinuities in the calculated surfaces. These calculations also suffered from large contributions of intruder states (i.e. singularities caused by a small or negative energy denominator in the PT2 formalism) which is a known problem of perturbative approaches, especially when Rydberg orbitals are involved and a diffuse basis set is used.[201] One indicator of intruder states is the so-called "reference weight" (a measure of how much a single configuration contributes to a state), which is plotted in Appendix A, Figure A.1 for all states along the O–H stretching coordinate. These instabilities would not be too problematic if one plans to use grid-based QD methods, where analytical functions are fitted based on the quantum chemistry data. One could fit smooth functions by simply ignoring the discontinuous points. However, in a direct dynamics approach such failures of the electronic structure will be detrimental to the outcome of the dynamics and hence an alternative method needs to be found that produces smooth surfaces. The EOM-CCSD cuts meet this requirement and are thus suitable for use in an *on-the-fly* approach, however the quality of the surfaces should be questioned. Although in general qualitative agreement with the multi-configurational results, even basic visual inspection reveals that the energies of the dissociative channels predicted by EOM-CCSD do not match the CASPT2 calculations. They are systematically higher, and increasingly so as the bonds get longer.

In such a situation, one should naturally have more confidence in multi-configurational approaches, as opposed to methods which are designed to be used at or near the equilibrium geometry. Furthermore, the overestimation of excitation energies is a known

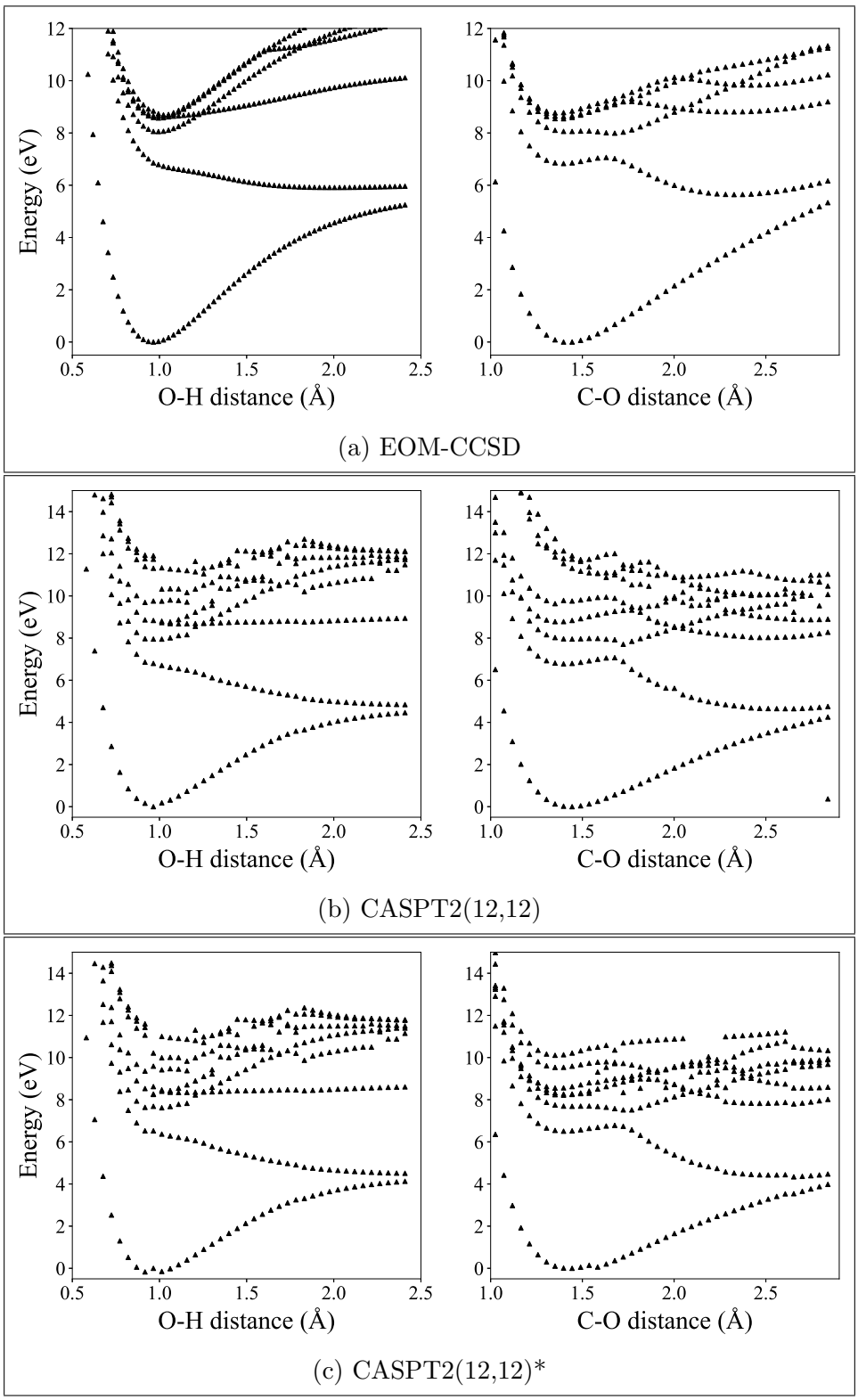


Figure 3.5: Potential energy cuts along the O–H stretching mode and the C–O stretching + C–O–H angle bending mode of Methanol at the (a) EOM-CCSD, (b) CASPT2(12,12), and (c) CASPT2(12,12)* levels of theory (* refers to the active space including extra Rydberg orbitals).

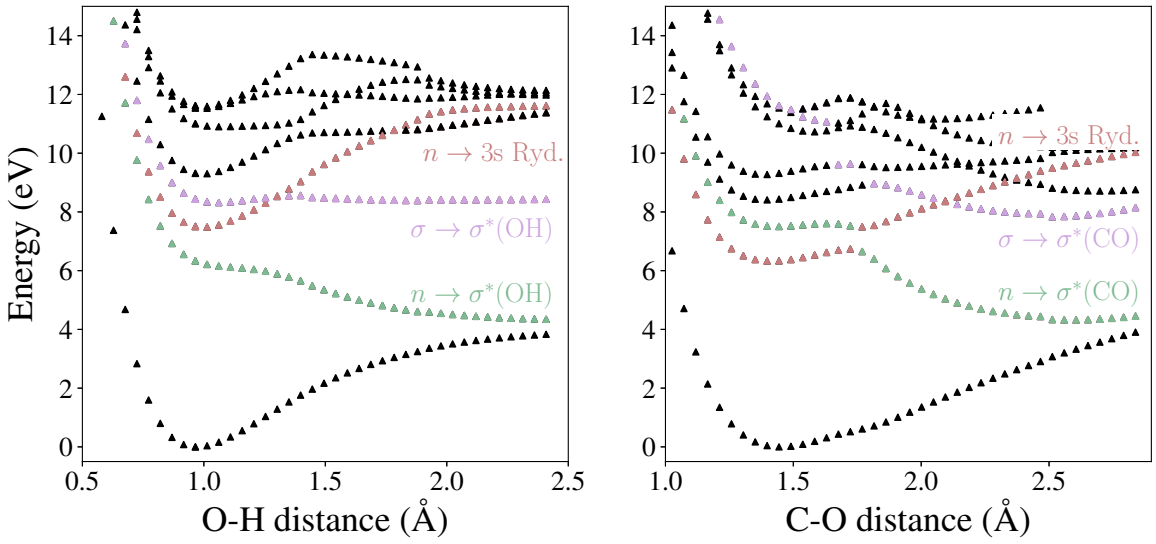


Figure 3.6: Potential energy cuts along the O–H stretching mode and the C–O stretching + C–O–H angle bending modes of Methanol at the RASSCF(12,2+8+2)[1,1] level of theory. Diabatic states have been identified and highlighted based on their chemical character.

problem with EOM-CCSD, and can usually be resolved by including triple excitations into the wavefunction, i.e. EOM-CCSD(T).[202,203] Such calculations would be prohibitively expensive for application in the present case, and the EOM-CCSD method was therefore not used for any QD simulations. Finally, a set of smooth points along the main dissociation coordinates was obtained by employing the RASSCF method (Figure 3.6), which proved to be more stable than the equivalent CASPT2 calculations, and yielded similar results at a computational cost of at least one order of magnitude less. They are, additionally, in good agreement with previously obtained PESs found in the literature.[178–180,184,188] It is also pointed out that even though regular CASSCF calculations would not necessarily be too expensive for use in *on-the-fly* dynamics, it is suspected that not including any dynamical correlation in the calculated energies would, in this particular case, not be valid. Compare for example the CASSCF energies with CASPT2 in Table 3.3, and notice the relatively large difference between the two, hinting at the importance of dynamical correlation. Although RASSCF does not explicitly correct the energy to include this effect, it has been argued that some amount of correlation is recovered through the interaction of the orbital subspaces, yielding more accurate results overall.[72]

For completeness, other alternative methods that could potentially yield appropriate surfaces are given. Going beyond standard CASPT2, it is possible to use variants

of the "multi-state" formalism such as Rotated Multi-State (RMS-CASPT2),[204] Extended Multi-State (XMS-CASPT2)[205] or Extended Dynamically Weighted (XDW-CASPT2),[206] which can correct for some of the shortcomings of the PT2 approach. Some preliminary calculations were performed using these alternatives but they were ultimately abandoned because of the increased computational cost and the fact that some of the instabilities in the standard CASPT2 calculations persisted. Moving away from active space based methods, there exist other approaches that should be able to handle the multi-reference nature of the problem at hand. These include the Density Matrix Renormalisation Group (DMRG)[207] or Quantum Monte Carlo (QMC) type methods.[208,209] They were not considered since derivatives and couplings may not be readily available, as well as the fact that their computational cost may limit their use in direct dynamics simulations. Additionally, there is no interface to any of these two methods in Quantics, at least for now.

The above discussion clearly highlights the subtleties and difficulties of using non-"black-box" electronic structure methods, while also explaining why more standard methods may not be suitable for a given problem, such as (photo)dissociation. Multi-configurational approaches, like the family of active space-based methods, are powerful but require careful consideration of the problem one is looking at. Proper benchmarking should be carried out before any simulations are attempted, especially when using an *on-the-fly* approach, to avoid unwanted changes in the active space leading to discontinuous PESs. Furthermore, it is not usually enough to merely consider the performance of a given method at the FC point as the conclusions drawn may not hold in the extended configurational space. All this to say that one of the biggest bottlenecks in the field of QD has been and remains to this day, the electronic structure calculations. Finding an appropriate method from the large amount of possible choices can be very time-consuming, but is necessary since it can massively affect the quality of the final dynamics calculations.

State characters

Having determined the suitability of the potential energy cuts produced by RASSCF calculations, further analysis of the surfaces is carried out. More specifically, the focus lies on the nature of the different dissociative channels. Figure 3.6 clearly shows a number of unbound states, both along the O–H and the C–O stretching coordinates.

Determining the chemical character of these states, i.e. what electronic transition they correspond to, is done here based on the leading CI coefficient of the electronic configurations (i.e. CSFs) contributing to a particular state in the RASSCF calculations. This measure of state character is deemed valid since it has been established that the coefficients in active space-based methods can also be used as a measure of the multi-reference nature of excited states (because the CI calculation is complete within the RAS2 active subspace).[210,211] In effect, this characterisation allows one to represent the energy-ordered adiabatic states in a diabatic picture. Lacking of course a proper mathematical formalism, the concept of "diabatisation by inspection" is introduced, whereby the diabatic states are constructed by inspecting the leading CI coefficients of the adiabatic states. Attempts at formalising such a diabatisation procedure have been previously suggested, first by Ruedenberg and Achity,[212,213] and later extended by Nakamura and Truhlar,[214–216] but its application has not been demonstrated on a manifold of more than 2 electronic states.

In the present case, the aim was not to obtain actual diabatic wavefunctions, but rather to simply identify the diabatic states along the two key coordinates of methanol and compare this analysis to the PES generated during the direct dynamics simulations (see later). The results of the CI analysis are plotted in Figure 3.7, where the top panels show again the calculated RASSCF surfaces, and the adiabatic state being tracked is highlighted. The bottom panels show the changes in CI coefficients for different transitions. This data nicely shows how an avoided crossing or a CoIn, identified by a sudden change in the gradient of the potential of a particular state, coincides with a different electronic transition starting to dominate the character of that state. This analysis was used to identify the three lowest lying diabatic states, which have been labelled in Figure 3.6 with their corresponding state character, and are summarised in Table 3.4. The results show two distinct dissociation channels which both correspond

Table 3.4: Electronic character of the first 3 diabatic states, along the O–H and C–O dissociation channels. Dissociative states are underlined. (g.s. = electronic ground state)

	$\tilde{\boldsymbol{X}}$	$\tilde{\boldsymbol{A}}$	$\tilde{\boldsymbol{B}}$	$\tilde{\boldsymbol{C}}$
O - H	g.s.	$\underline{n \rightarrow \sigma^*(\mathrm{OH})}$	$n \rightarrow$ 3s Ryd.	$\underline{\sigma \rightarrow \sigma^*(\mathrm{OH})}$
C - O	g.s.	$n \rightarrow$ 3s Ryd.	$\underline{n \rightarrow \sigma^*(\mathrm{CO})}$	$\underline{\sigma \rightarrow \sigma^*(\mathrm{CO})}$

to an excitation of the σ^* orbital of either the O–H or the C–O bond. This is entirely consistent with a chemist's intuition of a bond-breaking process, in which an initially

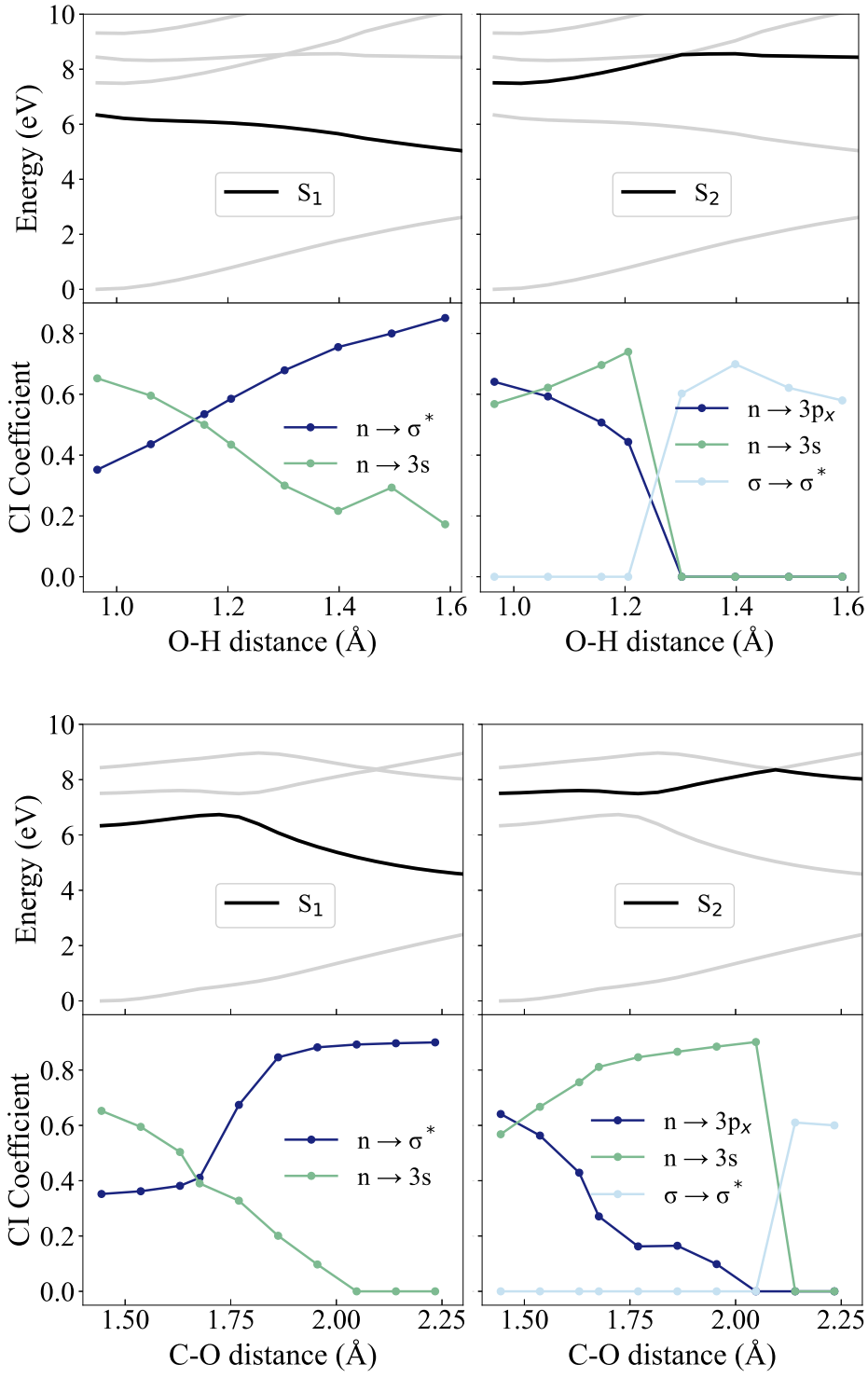


Figure 3.7: Calculated potential energies of the first 4 adiabatic states (top panels, see Figure 3.6 for full PESs), along with CI coefficients (bottom panels) for the S_1 (right) and S_2 (left) states along the O–H stretching (top figure) and C–O stretching + bending (bottom figure) coordinates.

high lying anti-bonding orbital comes down in energy as the bond length increases. As such, this investigation of relevant cuts along the excited state PESs of methanol already provides a qualitative picture of the mechanisms of dissociation of the molecule, which will be useful when analysing the outcome of the dynamics simulations in the next section.

3.3.2 Time-dependent $\chi^{nuc}(\mathbf{R}, t)$ analysis

All results of the QD simulations are described here. Each section presents one particular way in which a time-evolving wavefunction, made up of coupled wavepackets, can be analysed. By calculating different observables, one obtains different insights into the dynamics of the system and ultimately a complete picture of the photodissociation can be built up.

Absorption spectrum

The absorption spectrum of methanol is calculated from a vMCG calculation based on a LVC model. It is compared to an experimentally obtained one, but as already alluded to in the introductory chapter, such comparisons are not straightforward and shouldn't be seen as the ultimate validation. For example, many experiments are performed at room temperature in the condensed phase, whereas calculations tend to be done at 0K in the gas phase (technically, in a vacuum). Techniques do exist to account for such disparities, and methodologies have been developed that allow for accurate comparison of experimental results with computational data.[217] For the purposes of the current investigation however, the computation of a spectrum serves merely as a first check that the dynamics are capturing the important excited state processes. Such a test is especially useful when one is planning to use *on-the-fly* approaches, where it is initially not known what part of the configurational space will be explored by the nuclear basis functions.

The first 5 diabatic states of the 4-mode LVC model obtained from fitting RASSCF calculations are plotted in Figure 3.8 (see Table A.2 in Appendix A for a description of the modes). The quality of the fits is far from perfect but the idea behind performing a vMCG calculation was simply to approximate the absorption spectrum in order to get a sense of the initial dynamics of the system and make sure that all key features are well described by the electronic structure calculations, i.e. the dynamics along the

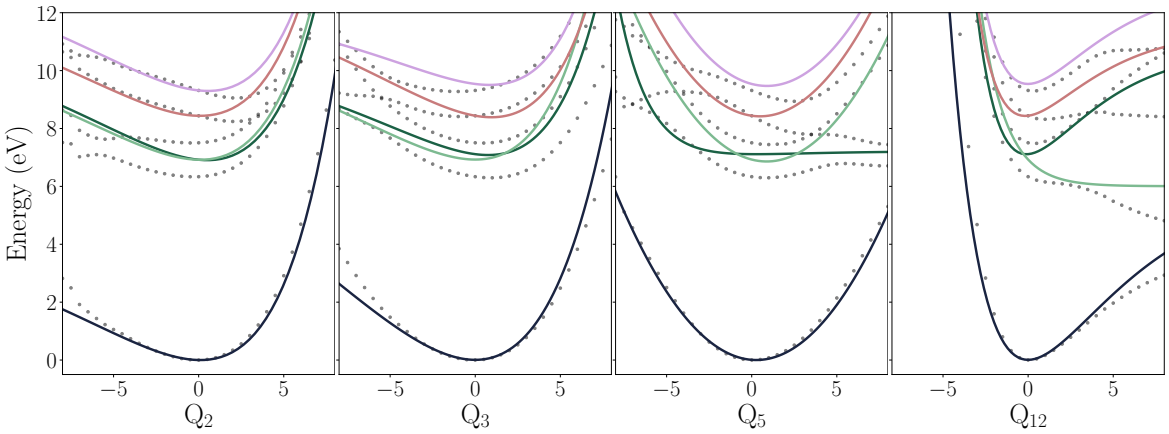


Figure 3.8: Diabatic cuts along the normal modes of the 4-dimensional LVC model. Surfaces were fitted to points obtained using RASSCF calculations (shown as grey dots).

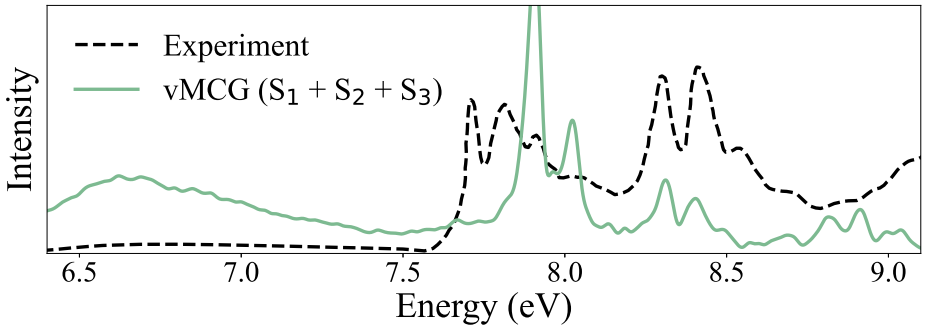


Figure 3.9: Calculated absorption spectrum of methanol (green) from vMCG simulations, and an experimentally obtained spectrum[173] (black, dashed). The vMCG spectrum is shifted horizontally by -0.16 eV.

O–H and the C–O bonds. The direct dynamics calculations then serve to provide the actual details of the observed phenomena.

A spectrum was generated following the method described in section 3.2.3 for 3 calculations, each starting on a different initial state. Then, the spectra were summed together and the resulting combined spectrum is compared to the experimental one in Figure 3.9. Despite the fact that the fitted potentials do not accurately represent the calculated RASSCF surfaces, the spectrum obtained from this simplified model captures the two main features seen in the experimental spectrum (albeit shifted by -0.16 eV, to account for the error in the calculated energies). Firstly, the broad low-intensity feature centered around 6.8 eV is reproduced thanks to the inclusion of an excited state dissociation channel of the hydroxyl hydrogen (see Q_{12} in Figure 3.8). With respect to the higher energy vibronic features corresponding to C–O vibrational

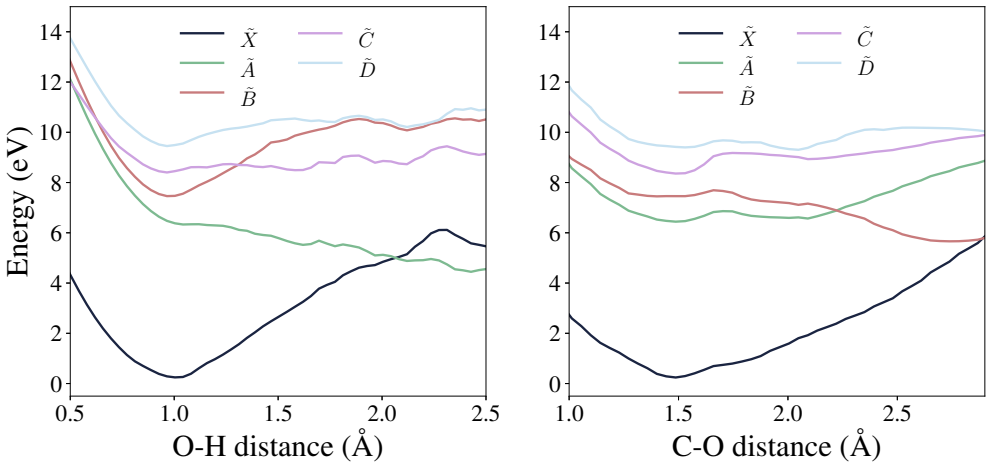


Figure 3.10: Diabatic cuts of the O–H stretching and C–O stretching + bending modes, constructed from the database points generated during DD-vMCG simulations. All other modes are at $q = 0$. Note: only the lowest 5 electronic states are shown, out of a total of 8 states included in the simulations.

motion, the calculation only recreates some of the peaks. This would indicate that either the 4 modes included in the model are not sufficient to describe the motion of the molecule, or that the relevant surfaces (i.e. along Q_2, Q_3, and Q_5) are not adequately fitted. Both of these alternatives are certainly possible, but in any case, the calculated spectrum does show that the electronic structure used to obtain the surfaces is at least a reasonable choice. Knowing this, direct dynamics simulations were performed using this same method and then analysed in much greater detail in the following sections.

Potential energy surfaces

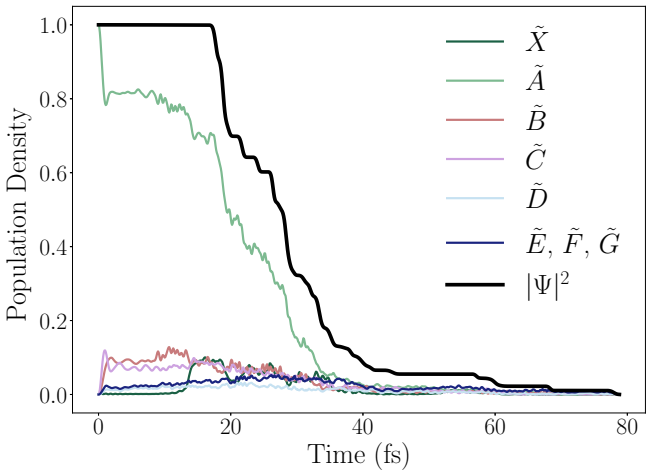
To check whether the calculated points in the database of the DD-vMCG calculations accurately describe the dissociation channels, 1-dimensional cuts along the O–H stretching and C–O stretching + bending coordinates are inspected (Figure 3.10). Comparing these surfaces with the ones of Figure 3.6, a good match is observed. Parts of the potential appear slightly discontinuous or uneven but this is mostly an artefact of the interpolation between points that may not lie exactly along the 1-dimensional coordinates that are being plotted. The two most important things to note are that the different dissociation channels are well represented (i.e. for O–H: $\tilde{A}$ and $\tilde{C}$, for C–O: $\tilde{B}$ and $\tilde{C}$), and that the diabatisation is done correctly (this can be judged based on the previous "diabatisation by inspection"). As a general rule, these criteria should be satisfied before initiating the final QD calculations.

Diabatic state populations

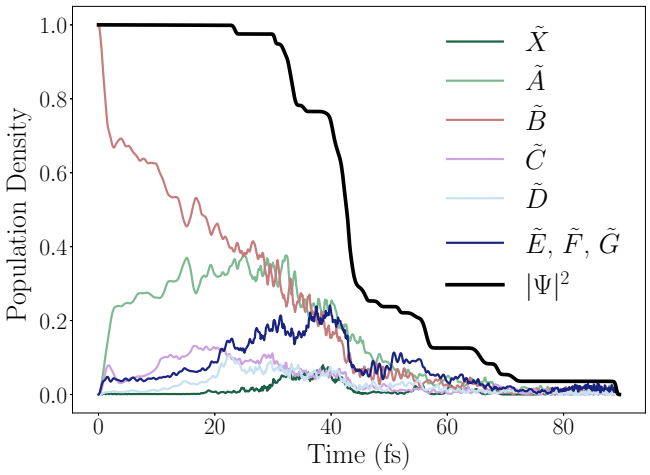
The use of diabatic state populations for analysis of quantum dynamics simulations is very common but is sometimes criticised for not representing an experimental observable. However, it is still believed to be a very useful tool to unravel the exact mechanism of excited state processes, and is often used to complement experimental studies. As discussed in the first chapter, reproducing experimental data should not always be the goal and furthermore, the information yielded by the individual (and total) state populations potentially has a lot of explanatory power and hence it is deemed appropriate for discussion in this context.

The plots in Figure 3.11 show the time-dependent population of the first 5 states of methanol (the 3 highest lying states are summed together in a combined dataset, for clarity), as well as the change in total magnitude of the molecular wavefunction, Ψ, calculated as the square of its norm, $|\Psi|^2$. One of the most obvious features that all three plots have in common is the rapid decrease of the total wavefunction density after a few tens of fs of the simulation. In a "closed system", the norm should of course be conserved but, as discussed, the CAPs will absorb any incoming wavepacket density that goes past the point at which it is defined. Since the CAPs are placed far enough away from the FC point, such that any wavepacket that reaches them can be said to have fully dissociated, one can infer from the black line in Figure 3.11 how rapid the overall dissociation process is and after how much time it starts. As the excitation energy increases, the wavepackets take longer to initially reach the CAPs which makes sense given that $\tilde{A}$ is the steepest dissociative state. If this state is not directly excited, it will take some time before any density is transferred to it via internal conversion processes, hence the delay in the sudden drop in total density. The step-like features in the decay are a sign that multiple pathways to one or more dissociative surfaces exist and are being reached at different times.

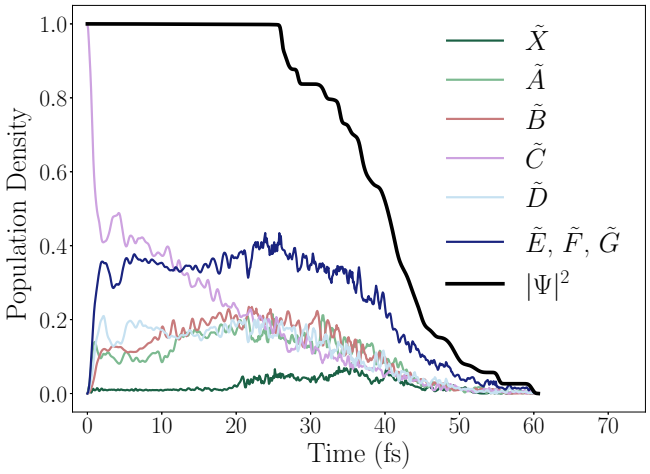
A more detailed analysis is obtained by looking at the coloured lines, representing the change in population of each individual state. Again, all three plots have one feature in common, namely the immediate transfer of population between the state being excited and some combination of other states. At first, this may look like unphysical behaviour since the molecule is not likely to move on a sub-femtosecond timescale, but this is resolved by realising that these initial population dynamics are



(a) $\tilde{A} \leftarrow \tilde{X}$



(b) $\tilde{B} \leftarrow \tilde{X}$



(c) $\tilde{C} \leftarrow \tilde{X}$

Figure 3.11: Change over time of the magnitude of the total wavefunction, $|\Psi|^2$ (black line), and of the individual diabatic state densities (coloured lines) upon excitation to the (a) first, (b) second, and (c) third excited state of methanol.

not the result of any kind of nuclear motion. Rather, it is a sign of the appropriate electronic eigenstate being formed according to the diabatic Hamiltonian, which is correctly described as a mixing of multiple states through the couplings between them. Upon excitation of $\tilde{A}$, not much internal conversion is seen after the initial mixing of the states and the decay of $\tilde{A}$ is essentially equal to the decay of the total density. This clearly indicates a rapid, ballistic dissociation which would give rise to a broad, low intensity feature in the absorption spectrum, and hence hydroxyl hydrogen loss can be fairly confidently said to be the main contributor to this decay, based on the previous experimental evidence. Excitations to the two higher lying states present a more complicated picture. There is a lot more internal conversion happening, to many different states. The overall dissociation is still very rapid once it starts however, indicating the presence of other, higher lying dissociative channels which are now being accessed (these channels are discussed later). In terms of relaxation to the ground state, it does not seem to be a major reaction channel in any of the three simulations, and ground-state reactions are therefore not further investigated.

Finally, the diabatic state population can also be used as a proxy for convergence.[218] However, no fully converged calculation is currently available for methanol, meaning no benchmark result exists. In such cases, it is common practice to converge a property of interest by performing a number of simulations with increasing basis set size.[152] In this case, the change in population of the initially excited S_1 state is compared across calculations using different numbers of GWPs (Figure 3.12). On the one hand, a clear similarity between the curves indicates a fast, ballistic dissociation in all cases, as expected. On the other hand, there are two important differences. Firstly, the smoothness of the curves increases as the size of the basis is increased. This is a clear sign of an evolution towards convergence. Secondly, the curves clearly start diverging after about 20 fs. Somewhat surprisingly, the deviations are not in the same direction making it difficult to determine how well the $N_{GWP} = 33$ simulation, used in this work, performs. For instance, the result obtained with 21 GWPs reproduces the $N_{GWP} = 67$ result better than any of the other simulations. Overall, one has to conclude that the simulations discussed and analysed in this chapter are relatively far from convergence. This is not unusual for DD-vMCG calculations but it does mean that calculated properties may not be quantitatively reliable. That being said, some properties are more sensitive to convergence than others, and appreciating this

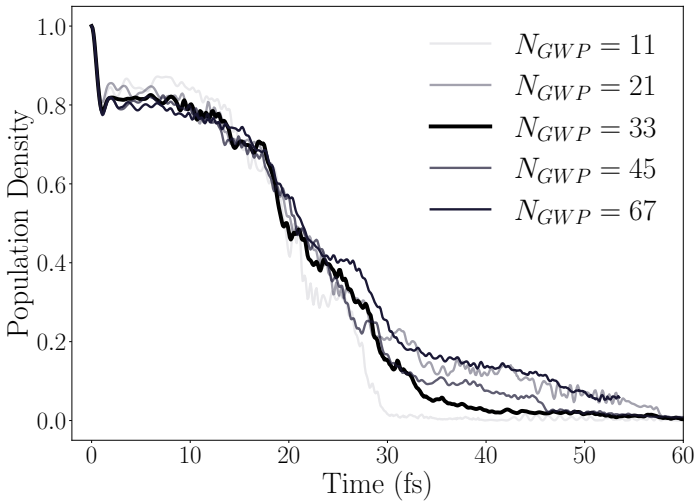


Figure 3.12: $\tilde{A}$ diabatic state population over time, for different numbers of GWPs, N_{GWP}. Thick black line is the result of the simulation analysed in this chapter.

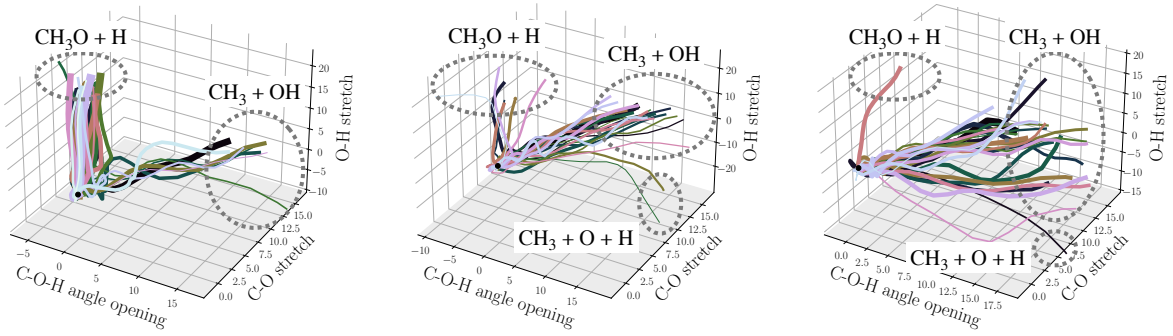
can nevertheless allow one to draw sensible conclusions even if fully converged results aren't available.

Wavepacket trajectories & product ratios

The next set of data that will be analysed is related to the movement of the individual wavepackets themselves. Using the GGPs to obtain suitable weights for the trajectories followed by individual GWPs, a quantitative determination of the relative product yield can be made for every channel explored during the propagation. The quantum trajectory for each GWP is plotted in Figure 3.13, where the positions of the centres of each wavepacket are given in the space spanned by the O–H stretching coordinate, and both the C–O bond stretching and C–O–H angle opening coordinates. This coordinate space allows for a clear and intuitive view of the separation of wavepackets into different product channels. A summary of the branching ratios of the three simulations is presented in Table 3.5. For excitation to the first excited state, the dominant

Table 3.5: Relative yields of photoproducts after excitation to the three lowest lying electronic states of methanol. Ratios are determined from analysis of quantum trajectories, weighted by their GGPs.

Excitation	**CH_3O + H**	**CH_3 + OH**	**CH_3 + O + H**
$\tilde{A} \leftarrow \tilde{X}$	0.82	0.18	0.00
$\tilde{B} \leftarrow \tilde{X}$	0.26	0.70	0.04
$\tilde{C} \leftarrow \tilde{X}$	0.05	0.93	0.02



(a) $\tilde{A} \leftarrow \tilde{X}$ (b) $\tilde{B} \leftarrow \tilde{X}$ (c) $\tilde{C} \leftarrow \tilde{X}$

Figure 3.13: Expectation values of the centres of the individual GWPs for excitations to 3 different initial states. Line widths are proportional to the GGP and the dashed circles indicate regions of the coordinate space corresponding to different photoproducts. Coordinates are given in units of mass-frequency weighted normal modes.

process is loss of hydroxyl hydrogen (as seen experimentally when exciting methanol at 193 nm,[172] see Table 3.1). Looking back at the PES cuts of the two dissociative coordinates (Figures 3.6 and 3.10), this observation is easily rationalised by comparing the topographies of the lowest lying dissociative channels along each of the two bonds. Whereas the diabatic $\tilde{A}$ state along the O–H coordinate is unbound and leads to ballistic dissociation, this same state is bound along C–O, and encounters a crossing with the higher lying dissociative state at some distance away from the FC point ($r_{C-O} \approx$ 1.8 Å). This results in an energetic barrier between the minimum of $\tilde{A}$ and the crossing point ($\Delta E_{barrier} \approx 0.45$ eV, compared to an experimental estimate of 0.58 eV[183]) and hence a barrier to dissociation, accounting for the much lower yield of this product in Table 3.5.

Exciting to the S_2 state changes the previous picture entirely. A dramatic change in branching ratios now identifies the production of hydroxyl radicals as the major reaction pathway. Naturally, such a difference caused only by a change in initial excitation energy implies a fairly complex mechanism within the excited state manifold. The result itself is, however, not very surprising given that it was already established that the diabatic $\tilde{B}$ state is dissociative only along the C–O bond, and not along the O–H coordinate. This accounts for the larger proportion of OH dissociation. However, to explain the second most abundant pathway, i.e. O–H photolysis, this qualitative picture, based on the known PESs, must be extended to include the role of the rather strong NAC between the low lying electronic states (especially S_1 - S_2). Through their coupling, population transfer between states can be very efficient and hence H atoms

can be produced via population transfer to the $\tilde{A}$ state followed by the previously discussed direct dissociation of the O–H bond.

Comparing these findings with the experimental results reported in Table 3.1 is not possible since no branching ratio is reported for the CH_3 + OH channel. The focus of that investigation was on the production of atomic versus molecular hydrogen, but the authors do state that C–O cleavage is observed and is thought to be a "not very minor" contributor to the overall signal. Additionally, other experiments investigating the dissociation processes of methanol after excitation to 157 nm conclude that the production of hydroxyl radicals is the result of a fast, direct dissociation. No explicit branching ratios are calculated unfortunately.[182,183] These observations, although not conclusive, are at least consistent with the mechanism presented here, derived from the simulations.

In the final calculation, where the initial wavepacket is instead projected onto the S_3 state, the resulting dynamics are fairly similar to the S_2 result. The predominant reaction channel is again the dissociation of the hydroxyl group. The ratio between this channel and the production of H atoms is now much more in favour of the former reaction though. Intuitively, a possible explanation would be that a more energetic excitation opens up different, higher lying channels leading to C–O bond breaking, while no such channel exists for O–H dissociation. In order to confirm this hypothesis, one has to calculate the state-specific contribution of the dissociation channels. Such an analysis will be presented in the next section.

Finally, a discussion of the CH_3 + O + H channel is given. This three-body dissociation only occurs when the system is excited above the first excited state, and has a low quantum yield. The trajectory plots show that only 2 wavepackets end up in this product channel after S_2 excitation (Figure 3.13b), and only a single one does so for S_3 excitation (Figure 3.13c). Even though this product accounts for such a minor fraction of the photochemistry, it is potentially important in the possible chain reactions following the initial fragmentation of methanol. The production of a reactive triplet O(^3P) atom can lead to H abstraction of neighbouring molecules and hence initiate a complex sequence of chemical reactions.[219] Experimentally, the production of CH_3 + O + H has been observed after excitation at 157 nm (i.e. S_2), and the reaction is suggested to proceed via a two-step process. In the first step, a methoxy radical is produced by O–H bond cleavage, which is then followed by C–O

dissociation caused by the probe laser in a pump-probe experiment.[183]

Analysis of the computed trajectories however, point to a direct three-body dissociation of the molecule. As the C–O–H angle starts to open up beyond 180°, both the C–O and O–H bonds start to lengthen simultaneously to a point where the fragments can be considered fully dissociated. Such a channel is accounted for by a local energy minimum of a high lying state ($\tilde{F}$). This minimum was found by inspection of a two-dimensional surface of that state along the O–H stretching and the C–O–H angle bending modes (see Figure A.2 in Appendix A). This state comes down in energy along both of these coordinates and can therefore be populated after excitation to S_2 or S_3. It is somewhat surprising to see such a channel and it could very well be a result of the electronic structure calculations breaking down, or not being sufficiently sampled in that region. If one assumes, on the other hand, that such a channel does indeed exist, it could correspond to C–O bond breaking where the OH fragment dissociates with high translational and rotational energy, leading to the increased O–H bond length. From velocity map imaging experiments, the OH product produced after excitation above S_1 is in fact seen to be vibrationally excited.[183] Either way, the excited state minimum explains why this seemingly 3-body dissociation product is observed in the simulations, and only accounts for a minor fraction of the photoproducts.

To briefly summarise, the observed branching ratios of the three simulations discussed above can be rationalised in terms of a competition between direct and indirect dissociation, depending on whether a bound or unbound state is initially populated. Not surprisingly, direct dissociation dominates in all three cases, but because the topography of the states differs along the different bond dissociation coordinates, the relative branching ratios can vary significantly based on the excitation energy. More energetic excitations also make higher lying bond dissociation channels accessible, which are discussed next.

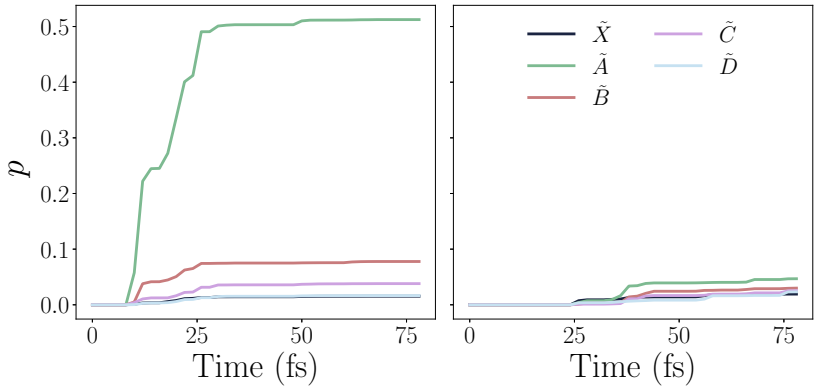
State-specific flux

In order to identify the specific excited state dissociation pathways for each bond breaking coordinate, the state-specific flux for each photoproduct is calculated. Two dividing surfaces were defined, along the O–H and C–O coordinates. The locations were chosen to be roughly twice the equilibrium bond length, i.e. x_d (OH) = 1.93 Å and x_d (CO) = 2.65 Å. As GWPs cross these dividing surfaces, the expectation

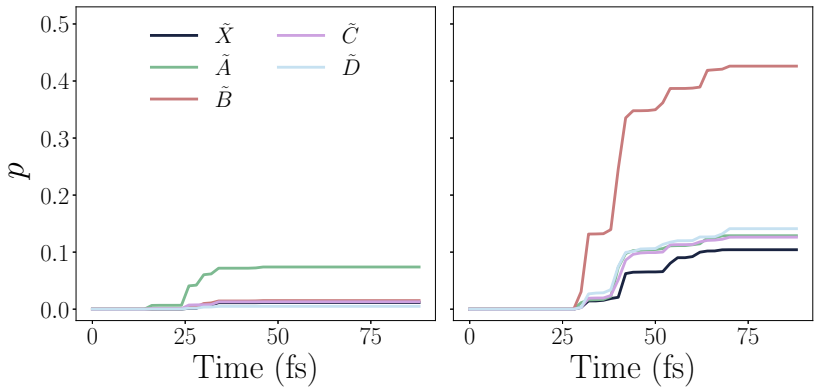
value of the flux operator is calculated for each of the first 5 electronic states separately. The change in integrated flux can then be plotted to reveal the relative contribution of those states to each of the two bond dissociations. The results are presented in Figure 3.14. These reveal, or rather confirm, that when the first excited state of methanol is populated, a ballistic hydroxyl hydrogen dissociation occurs via the unbound $\tilde{A}$ state. This channel accounts for over 50% of the total dissociation. The remaining contributions are seen to be partly due to some higher lying O–H dissociation channels, and partly due to some density going into the C–O channel.

This data can also be used to estimate the branching ratios between the two main products. One simply sums up the total integrated flux of all states along each of the two channels. Although this was already determined from the trajectories and their corresponding GGPs, calculating it again using an entirely different method is valuable since these two approaches, provided the results match, serve as a validation for each other (in addition to comparison with experimental values). The branching ratio after $\tilde{A} \leftarrow \tilde{X}$ excitation is found to be $CH_3O + H : CH_3 + OH = 0.82 : 0.18$, in excellent agreement with the ratio in Table 3.5. In the next simulation, the $\tilde{B}$ state is excited which is bound along the O–H coordinate but not along C–O. This is reflected in the state-specific flux, showing close to 50% dissociation of the C–O bond, via the unbound $\tilde{B}$ state. Notice that the dissociation starts after about 30 fs, whereas in the first simulation the O–H bond started dissociating after only 10 fs. As stated previously, this is accounted for by the small energetic barrier with respect to C–O bond breaking.

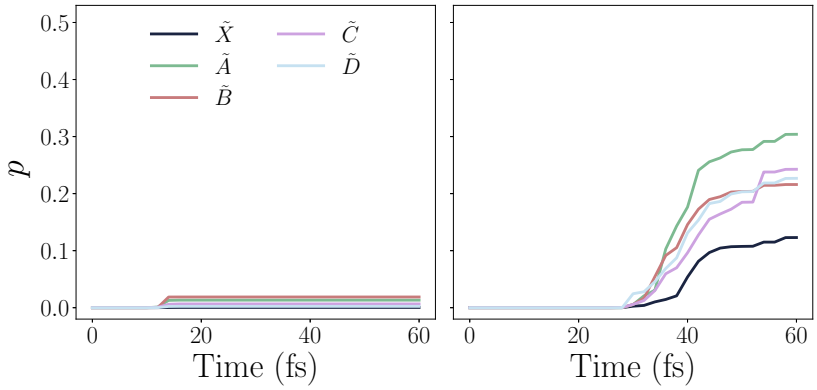
One also notices the non-negligible contributions of the other 4 states towards the formation of the hydroxyl radical. This observation is not surprising considering the energy present in the wavepackets. As previously remarked, the zero-point energy is included in the initial excitation. Looking back at the PES along the C–O coordinate (Figure 3.10), the asymptote defined by the $\tilde{A}$, $\tilde{C}$, and $\tilde{D}$ states is relatively close in energy to the $\tilde{B}$ state. Furthermore, the momentum of the individual wavepackets, plotted in Figure 3.15, suggests that enough energy is present in the system for some density to reach the higher states, which is confirmed by the diabatic state populations in Figure 3.11. About 10% of the total dissociation is accounted for by hydrogen loss via the $\tilde{A}$ state. In this case, the timescale is delayed slightly since internal conversion from $\tilde{B} \rightarrow \tilde{A}$ needs to occur first. Calculating the branching ratio from this data, one



(a) $\tilde{A} \leftarrow \tilde{X}$

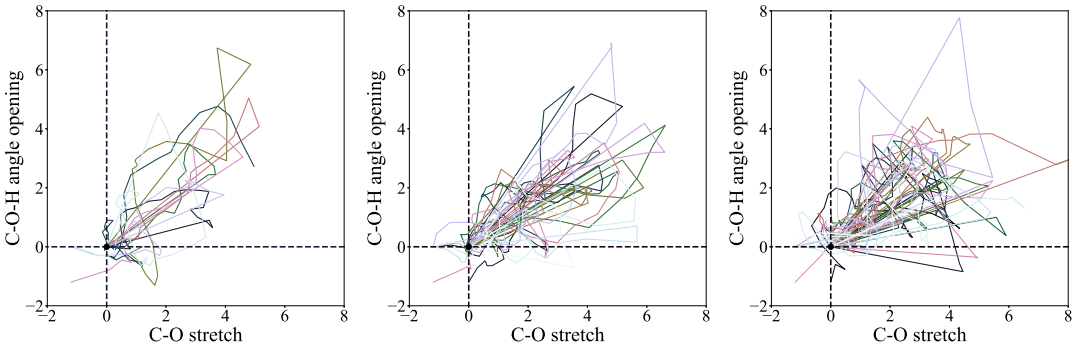


(b) $\tilde{B} \leftarrow \tilde{X}$



(c) $\tilde{C} \leftarrow \tilde{X}$

Figure 3.14: Integrated flux over total simulation time (representing the probability of dissociation) for three initial excitations. Left panels: O–H dissociation. Right panels: C–O dissociation.



(a) $\tilde{A} \leftarrow \tilde{X}$ (b) $\tilde{B} \leftarrow \tilde{X}$ (c) $\tilde{C} \leftarrow \tilde{X}$

Figure 3.15: Individual GWP trajectories plotted in the momentum space spanned by the C–O stretching and C–O–H angle bending coordinates.

finds $CH_3O + H : CH_3 + OH = 0.16 : 0.84$. The deviation from the results in Table 3.5 is now much bigger, although the same qualitative trend is observed. Reasons for this discrepancy include the fact that no distinction can be made between the 2-body dissociations and the observed 3-body dissociation. Additionally, in some trajectories the molecule fell apart due to the artificial accumulation of forces on a particular basis function, leading to runaway wavepackets. These were ignored in the calculation of the branching ratio using trajectories, but cannot be easily removed from the flux-based approach, and hence these unphysical contributions will affect the quantitative results of the latter analysis.

Finally, the dissociation reaction after $\tilde{C} \leftarrow \tilde{X}$ excitation is dominated by C–O bond breaking via the $\tilde{A}$, $\tilde{C}$, and $\tilde{D}$ states, which converge asymptotically to the same product channel. Most of the remaining density goes into the lower energy C–O channel via the $\tilde{B}$ state. Only a small fraction of the density accounts for O–H dissociation, leading to the following branching ratio $CH_3O + H : CH_3 + OH = 0.04 : 0.96$, again in good agreement with Table 3.5.

Although one could go on to compute many more properties of the system, such as the expectation value of a particular bond or angle during a simulation, or the wavefunction density along a bond for example, these would not necessarily add any further insights into the excited state dynamics of methanol. A fairly complete picture is obtained and possible reaction mechanisms are elucidated based on the presented data. The results section will hence be concluded here.

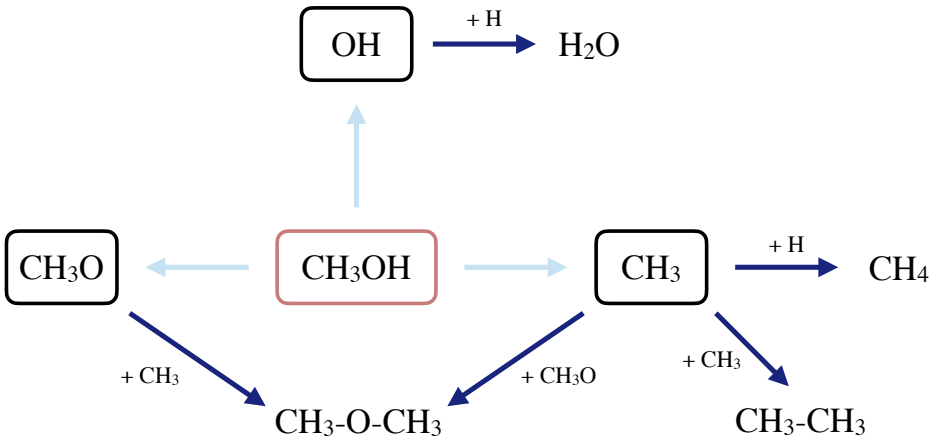


Figure 3.16: First generation reaction scheme after methanol photodissociation. Light blue arrows are photodissociation steps, while dark blue arrows indicate radical reactions.

3.4 Reaction scheme

Before moving on to a final summary and outlook, a brief discussion is dedicated to the implications of the photodissociation reactions of methanol, with respect to the chemistry that may be induced by the initial fragmentation. Since radicals are produced, possible chain reactions are possible when these reactive intermediates encounter other species. However, the scope of this discussion is limited to first generation reactions only, meaning the first step in any possible chain reaction. Additionally, the only species allowed to take part in the reactions are those that are directly generated from the initial photodissociation, i.e. H, CH_3, OH, and CH_3O. These reactions are shown in Figure 3.16 as a chemical network, starting from UV irradiation of CH_3OH. Some important molecules are immediately identified, and it should be obvious that their relative abundance is directly related to the branching ratio of the parent species, i.e. methanol. The formation of water is possible from the reaction of a hydroxyl radical with hydrogen. Methane can also be produced by CH_3 + H. This is relevant in an atmospheric context since methane is a potent greenhouse gas. Interestingly, the growth rate of methane has been fluctuating in recent years, most likely caused by variations in its main atmospheric sources and sinks, both of which are not fully understood.[220,221]

This lack of understanding points to a need for better atmospheric models which can correctly account for the abundances of different species. In fact, methanol itself is the second most abundant (organic) gas in the troposphere, methane being the first,

and it has been pointed out that major discrepancies exist between different chemical transport models with respect to the role of methanol in atmospheric dynamics.[222] In all cases however, it is clear that methanol influences the abundance of many other tropospheric gases.[223]

In a recent perspective article, it was highlighted that to fully understand the chemistry of our atmosphere, computational photochemistry should be used to inform large scale models, which can then be compared to experimental data collected in labs as well as in the field.[224] It is now clear that obtaining accurate branching ratios for important atmospheric species is very desirable in this context. Similarly, detailed knowledge of the photochemistry of molecules abundant in interstellar space is crucial in understanding astrochemical reactions. As pointed out in the introduction, methanol is an important species in these environments. Note however, that in contrast to its atmospheric chemistry, interstellar methanol is found in the condensed phase and therefore the effects of its environment should be included if quantitatively accurate results are to be obtained.

3.5 Conclusions & Outlook

This chapter presents a detailed case study of methanol that explains and demonstrates how photodissociation processes can accurately be modelled using computational techniques, specifically QD simulations. The method in question consists of the propagation of a wavefunction on the electronic excited states of a molecule, in which the wavefunction is decomposed into individual, but coupled, Gaussian basis functions. By using an implementation of this QD method that generates the PESs *on-the-fly*, i.e. the DD-vMCG method, one has an ideal tool to try and understand the complex photochemical and photophysical processes underpinning the light-induced dissociation of small organic molecules. The molecules have to be small because these simulations are very costly as a consequence of the fully "quantum" nature of the representation of the wavefunction. That being said, methanol is small enough that no compromises had to be made to reduce the computational cost of the calculations. The wavepackets were allowed to move in all dimensions (defined by all the normal modes, except 1) and through a manifold of eight coupled electronic states. By allowing this flexibility, the dynamics were not biased in any way towards a particular

reaction channel, and therefore the expectation is to obtain quantitatively accurate quantum yields of the photoproducts. Additionally, it is desirable to be able to derive a detailed mechanism from the simulations.

Both these goals can only be achieved if the underlying potentials on which the dynamics proceed are accurate. The electronic structure method used to obtain these was thus benchmarked against a range of similar approaches, but even then it remains difficult to judge the ultimate quality of the surfaces during a direct dynamics simulation. Ultimately, the most suitable method was found to be RASSCF(12,2+8+2)[1,1] which produced smooth, continuous potentials in agreement with previous studies. The suitability of these potentials was tested by fitting analytical functions to them, along four of the main modes, and performing a grid-based dynamics simulation in order to calculate an absorption spectrum. Although certainly not perfect, the match of the computed spectrum with the experimental one was good enough to be confident in the validity of the RASSCF calculations.

Using the DD-vMCG method, three separate QD simulations were run with the aim of comparing the effect of initially populating different excited states. To summarise the results section, five different analyses of (DD)-vMCG simulations were carried out and discussed: calculation of an absorption spectrum, generation of PESs, diabatic state populations, wavepacket trajectories, and state-specific flux. Each time, this data was considered specifically with respect to the two most important dissociation channels of methanol; the O–H and the C–O bond breaking coordinates. From all three simulations, the branching ratios of the photoproducts were determined using two different methods. The first is based on an inspection of the trajectories followed by individual basis functions, whereas the second is obtained from calculating the state-specific flux of a particular channel over time. Both these calculations are in good agreement with each other and certainly show the same trend, namely that the dominant dissociation pathway after excitation of the first electronic state is changed by exciting to higher lying states. This observation, consistent with experimental studies, is rationalised by looking at the diabatic PESs. When a steeply dissociative state along the O–H bond is directly excited, a dissociation into methoxy radicals and hydrogen is the result. The photolysis of the C–O bond carries a much smaller probability. However, when populating higher lying states this picture is reversed and one generates predominantly hydroxyl radicals. In short, when methanol is excited

above its first excited state, more dissociation channels become accessible for C–O cleavage, compared to O–H cleavage which is the dominant reaction after S_1 excitation.

Nothing is said about ground state reactions but extending the simulations to include these processes is possible. It is however believed that the vast majority of photodissociation events in methanol occur via excited electronic states. A notable exception is the formation of HCOH ("reactive formaldehyde") via S_0 dissociation, following UV excitation to S_1. This product has recently been observed experimentally,[181] and subsequently modelled by trajectory-based dynamics calculations.[187] Although some population is transferred to the ground state in the simulations presented here, these types of reactions were not investigated. Looking back, the excited state dynamics of methanol proved to be an ideal case study to demonstrate the capabilities of high-level QD simulations since a radical change in branching ratio is observed depending on which state is initially populated. Fully explaining such an observation undeniably requires the use of non-adiabatic dynamics methods.

On a more general note, it is important to realise that the QD methods being used here are still under development and there remain some questions about the ultimate accuracy of the results. The level of convergence is one such question, or more precisely, how the level of convergence affects the computed results. Certainly, there will be quantitative discrepancies but the main conclusions are not expected to be affected by not fully converging the dynamics. As such, the DD-vMCG method is shown here to be a powerful and flexible tool to study photodissociation, an infamously difficult challenge. If one is interested in determining precise quantities, the performance of the presented simulations may not be satisfactory, but it should be pointed out that they can be used as an initial starting point for further refinement. For example, the surfaces stored in a database could be fitted analytically and subsequently used in more accurate, grid-based approaches to compute the dynamics (ideally, MCTDH). Indeed, efforts to make this process easier are under way, one of which is discussed in chapter 5. This avenue may be further explored in the future.

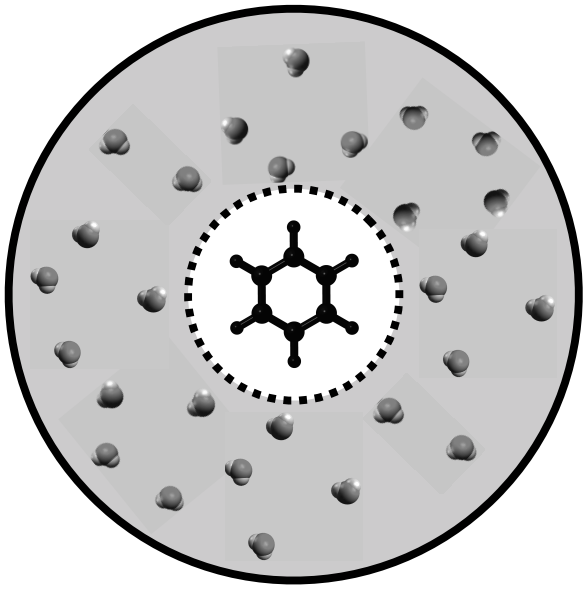
An additional idea to extend this work is to investigate the effects of molecular environments (e.g. solvents) on the predicted branching ratios. This is a difficult challenge but it may be achieved using the newly developed AQUATICS procedure, which combines force fields with QD simulations to allow the modelling of explicit environment interactions with a photo-excited solute molecule. This method and its development

is introduced in the next chapter. Referring back to the introduction of this chapter, methanol is thought to be relevant in an astrochemical context, where it is observed as one of the more abundant components of interstellar ice grains. Interestingly, the photodissociation reactions of methanol in those ices are thought to be quite different from the gas-phase reactions. Modelling a system composed of methanol on an ice surface, and determining the branching ratio in a manner similar to what has been described here, could explain the observed differences. Far from being insignificant, such a model could also help our understanding of the general chemistry related to these ices and potentially contribute to a working theory of how chemical complexity is built up in our universe.



CHAPTER 4

Force Field Quantum Dynamics



4.1 Overview

In this chapter, the development of a procedure to include environment effects in QD simulations is presented. In the first section, a brief review is given of the most popular methods to model environments of molecular systems. The basic features and limitations of those models are considered, as well as how they could be combined with non-adiabatic dynamics to account for environmental effects on fundamental photochemical phenomena such as electronic transfer processes (internal conversion, intersystem crossing). Having reviewed some recent developments in that area, a new approach is formulated that aims to explicitly model a molecular environment during a non-adiabatic quantum dynamics simulation. This approach shall generally be referred to as Force Field Quantum Dynamics (FFQD). An overview of the conceptual idea behind FFQD is given, along with the basic theory, implementation, and testing of the method. Given that this type of calculation is still being actively developed, the presented work should be seen as explorative. As such, the final section will consider possible future directions for the continued development of this approach.

4.2 Modelling the environment

To begin with, the meaning of the term "environment" is clarified. In the context of molecular simulations, the inclusion of an environment refers to *any* form of representation of interactions between a molecule (or molecular fragment) and its surroundings (that are not, in a chemical sense, part of the molecule, or fragment, itself). The most obvious form of a system with an environment, in terrestrial chemistry at least, is a solvated molecule. The environment is then understood to mean every solvent molecule and all its interactions with the solute. Further examples include the modelling of solid state systems, where one can define a unit cell to act as a single molecule, surrounded by an environment of infinite identical cells. An electromagnetic field can also constitute an environment.

It is not hard to imagine that the effects of an environment in general – whether electrostatic, thermal, steric or otherwise – play a very important role in the observed chemistry and physics around us. All biologically relevant chemistry for example, happens in the condensed phase (hence the expression *"water is the solvent of life"*).

It follows therefore that there is great interest in theoretical models that are able to simulate the complex interactions of an environment with a molecular system. This is an exceedingly difficult task since it increases the computational requirements of the problem massively. Nevertheless, much progress has been made in this field and different methodological schemes have emerged. Of specific importance here are the so-called solvation models, which aim to provide a description of the interactions of a solvent with a solvated molecule. They will be briefly discussed and compared to provide context for the newly proposed methodology that is presented in this chapter.

Before looking at the differences between commonly used solvation methods, there are a few fundamental aspects worth mentioning that are shared by all good descriptions of solvent. The first is that a good solvation model should allow for the solute to be described separately. This is advantageous since it allows one to treat the solvated species at any desired level of theory, while the solvent interactions can be represented in an entirely different way. The downside of this approach however, is that it naturally gives rise to a partition of the whole system, separating the solute from the solvent. It will become clear that this can cause issues. A second feature that all good solvation models have in common is that they are, at least in principle, easy to extend to all kinds of different environments/solvents. This of course is desirable simply for the sake of generality. Figure 4.1 illustrates the structure of a generic solvation model, containing the solute (*sol*) and an arbitrary environment (*env*) as two separate "layers" of the combined, solvated system (*sys*). The Hamiltonian of such a system can be formally expressed as

$$\hat{H}^{\mathrm{sys}}(\mathbf{x}_{\mathrm{sol}}, \mathbf{x}_{\mathrm{env}}) = \hat{H}^{\mathrm{sol}}(\mathbf{x}_{\mathrm{sol}}) + \hat{H}^{\mathrm{env}}(\mathbf{x}_{\mathrm{env}}) + \hat{H}^{\mathrm{int}}(\mathbf{x}_{\mathrm{sol}}, \mathbf{x}_{\mathrm{env}}) \quad , \tag{4.1}$$

where the total system Hamiltonian is partitioned into the individual solute Hamiltonian, $\hat{H}^{\mathrm{sol}}$, and the collective environment is described by $\hat{H}^{\mathrm{env}}$. The interactions between solute and solvent is then given by $\hat{H}^{\mathrm{int}}$, dependent on both the solute and environment coordinates ($\{\mathbf{x}_{\mathrm{sol}}\}$ and $\{\mathbf{x}_{\mathrm{env}}\}$, respectively).

Models of this kind, which exhibit both of the aforementioned properties, will now be described, grouped in two categories. The categories arise based on a dichotomy defined by whether or not the solvent is *explicitly* simulated (meaning atomistically) or if the interactions are *implicitly* generated.

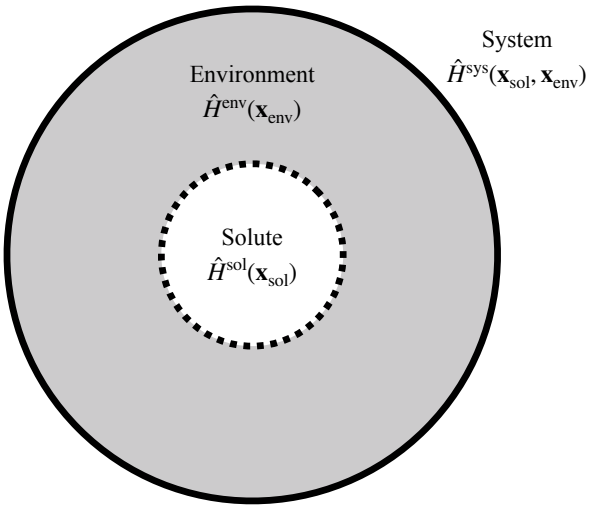


Figure 4.1: A general solvation model, containing the solute and the environment as separate "layers" of the model, combining to form the overall solvated system.

4.2.1 Implicit solvation models

Solvation models which do not include an atomistic description of the environment are termed implicit models. These have been, and still are, very popular approaches thanks to their simplicity. In particular, the Polarisable Continuum Model (PCM) has been around for over half a century and is still one of the most widely used methods to describe solvent interactions, with varying degrees of accuracy.[225,226] The basic idea behind the PCM (and all other continuum models for that matter) is to ignore the configurational degrees of freedom of the solvent and instead reduce the solvent interaction to a purely electrostatic potential, defined by the charge distribution of the solute and the dielectric constant of the solvent, ε. The charge distribution is assumed to be contained entirely in a cavity (C), defined by the molecular shape of the solute (typically based on the van der Waals radii of the constituent atoms). By additionally assuming the solvent to be isotropic and homogeneous, the following condition is applied to the system:

$$\varepsilon(\mathbf{x}_{\text{sys}}) = \begin{cases} 1 \ , \mathbf{x}_{\text{sys}} \in C \\ \varepsilon \ , \mathbf{x}_{\text{sys}} \notin C \end{cases} \quad , \tag{4.2}$$

where $\mathbf{x}_{\text{sys}}$ is the coordinate space of the system. The interaction between the solute and the environment is now simply an electrostatic potential that combines both the solute potential and the contribution of the charge distribution on the surface of the cavity (i.e. the area where the environment interacts with the solute).

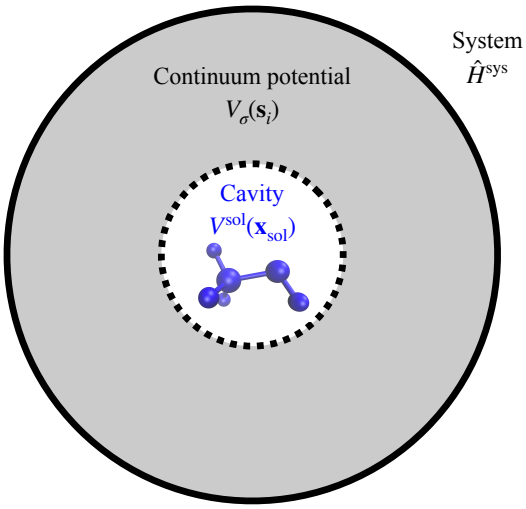


Figure 4.2: The PCM solvation model, showing the molecular cavity surrounded by a continuum potential.

In practice, the cavity surface consists of a number of individual points and hence the charge distribution is also discretised into a set of point charges, q. The potential operator describing the interaction between the solute and solvent, $\hat{V}^{\text{env}}$ can thus be written as

$$\hat{V}^{\text{env}} = \sum_i q_i(\mathbf{s}_i)\hat{V}_\sigma(\mathbf{s}_i) \quad , \tag{4.3}$$

where $\hat{V}_\sigma$ is the potential operator calculated at each point, i, on the surface, $\mathbf{s}_i$ This scheme, like many solvation models, is designed to be coupled to the QM description of the solute molecule. In effect, this is done by constructing an effective Hamiltonian describing the entire system, $\hat{H}^{\text{sys}}$, composed of the solute-only Hamiltonian, $\hat{H}^{\text{sol}}$, and the electrostatic contribution due to the solvent, as described above:

$$\hat{H}^{\text{sys}} = \hat{H}^{\text{sol}} + \hat{V}^{\text{env}} \quad . \tag{4.4}$$

Schematically, the PCM scheme is summarised in Figure 4.2.

The advantages, as well as the main drawbacks, are immediately clear from this brief description. On the one hand, the fact that no explicit solvent DoFs are considered makes the PCM very efficient, only adding a small computational overhead to the quantum chemistry calculation performed on the solute (note: PCM is also compatible with the computation of electronically excited states[227]). Furthermore, describing different solvents is straightforward since their identity is entirely determined by a single variable; the dielectric constant. On the other hand, and this is true for any implicit

solvation model, direct solute-solvent interactions (e.g. hydrogen bonding) can never be described. These short-range direct interactions may not always play an important role but in the context of molecular dynamics, it has been shown that the particular configuration of the solvent, even at long ranges, can significantly impact the outcome of the dynamics simulations. This is referred to as the "static disorder effect",[228–230] and is of course also not captured in continuum models.

4.2.2 Explicit solvation models

Given the shortcomings of implicit solvation models, namely the lack of any solvent DoFs, an obvious alternative is to include an atomistic description of the solvent in the model. These so-called explicit solvation models also have a long history of development, starting in the 1970's with the work of Karplus, Honig, Warshel and Levitt.[231–233] This early work exploited the natural partition that arises by defining, separately, the solute and the environment. Different levels of theory can then be combined to model the whole system. The solute, provided it is small enough, can be modelled with QM methods, whereas the solvent and its many DoFs can be explicitly included if treated classically. This means that the intra- and intermolecular solvent interactions, as well as the solute-solvent interactions, are defined by Force Fields (FF), containing both bonded and non-bonded terms. The properties of the system are then governed by Molecular Mechanics (MM). These two regimes can be combined in a formalism referred to as QM/MM, illustrated in Figure 4.3.

The total energy of a system set-up in this way is obtained by simply summing up the QM and the MM components, followed by a term describing the interaction between the two:

$$E^{\text{sys}}_{QM/MM} = E^{\text{sol}}_{QM} + E^{\text{env}}_{MM} + E^{\text{int}}_{QM-MM} \quad . \tag{4.5}$$

The coupling term, E^{int}_{QM-MM}, includes non-bonded interactions (electrostatic and van der Waals) as well as any bonded interactions in the case of covalent bonds crossing the QM/MM barrier. Accurately computing these contributions to the total energy is often not straightforward. It is possible to remove the need for this term by adopting a so-called subtractive scheme, as opposed to the additive approach of Equation 4.5. The total energy is now obtained by the QM energy of the solute plus the MM energy of the entire system, minus the result of a third calculation which computes the MM

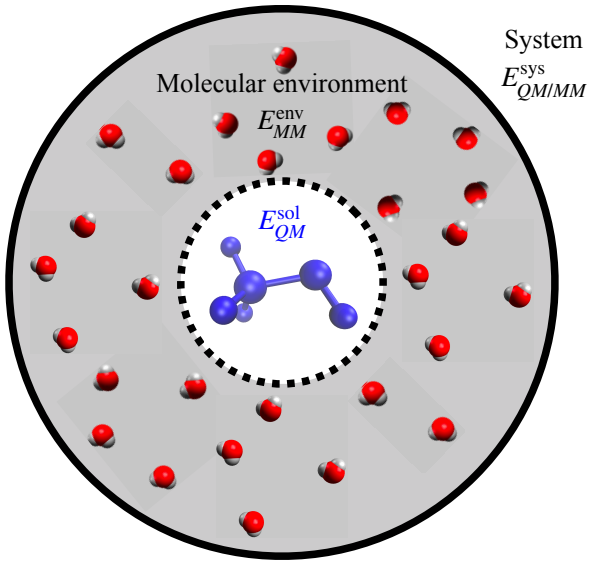


Figure 4.3: The QM/MM solvation model. The individual solvent molecules are described using MM, whereas the central molecule is a solute modelled using a QM level of theory.

energy of the solute (to avoid double counting):

$$E^{sys}_{QM/MM} = E^{sol}_{QM} + E^{sys}_{MM} - E^{sol}_{MM} \quad . \tag{4.6}$$

This way of expressing the energy forms the basis of a slightly different explicit solvation model; the ONIOM method.[234,235] Like QM/MM, it also partitions the system in layers but one is not limited to only 2 layers any more (although the 2-layer QM/MM model can also be realised using ONIOM). In principle, the ONIOM approach can be used to construct an n-layered system where each layer uses a particular level of theory to describe the whole system up to that layer. In practice however, one typically does not go beyond three layers (referred to as ONIOM3). These layers generally define a *high*, *medium* and *low* level of theory, which are applied, respectively, to a small model sub-system (sol), an intermediate sub-system (env') and the large "real" system (env). This partition is illustrated in Figure 4.4. As stated, ONIOM uses a subtractive scheme to evaluate the total energy:

$$E^{sys}_{ONIOM3} = E^{sol}_{high} + E^{env'}_{medium} - E^{sol}_{medium} + E^{env}_{low} - E^{env'}_{low} \quad . \tag{4.7}$$

There are of course many ways in which one can define each layer but the most widely used schemes are either 3-layer systems which use a high-level QM method, followed by a lower level QM method, and finally the last layer uses MM (written

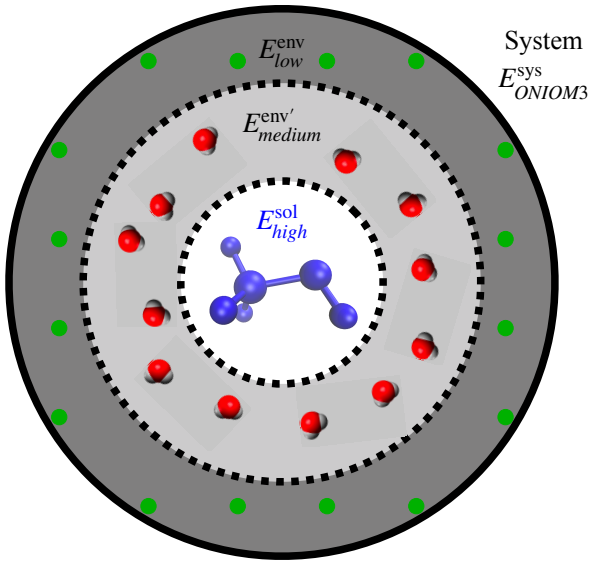


Figure 4.4: The ONIOM solvation model. The central molecule (sol) is described with a high-level theory, the intermediate layer (env') with a lower level and the largest layer (env) uses the lowest level of theory. The green dots represent an approximate way to model a molecular environment, e.g. point charges.

as ONIOM3(QM,QM',MM)). Another popular alternative are models where all three layers are described at QM level.

Layered approaches have allowed for the simulation of accurate molecular dynamics of very large systems, but of course suffer from some drawbacks as well. The most obvious one is that the user has to define the sub-systems (layers) of the model, and this choice is not always straightforward. In the case of covalently bonded systems, one runs into problems when a bond crosses a boundary between layers. Furthermore, the polarisation of the environment on the wavefunction of the lowest layer is not easily included. Different electronic embedding schemes have been devised addressing this issue, each with their own advantages and shortcomings.

4.2.3 Microsolvation

Finally, one more solvation approach will be briefly discussed in this section; micro-solvation.[236,237] It is a form of explicit solvation that does not use layers, but instead treats the entire system at the same level of theory. If one wants to use accurate QM methods, this of course means that only a very limited number of solvent molecules can be included (typically between 1 and 4). Nevertheless, this approach can be used to overcome a general weakness inherent in layered methods, namely that they cannot

handle an atom (or a molecular fragment) moving from one layer to another. If one is interested in studying proton transfer dynamics for example, this may be problematic. By creating a microsolvated system, carefully choosing which solvent molecules should be included, the potential reaction dynamics between a solute and its environment can be simulated. The model can then be improved by placing the microsolvated system in a cavity and surrounding it by a PCM for example, to account for some of the long range electrostatic interactions of the solvent. This method is potentially very well suited to the description of direct solute-solvent interactions but it does not have the property of being easily adaptable to different kinds of environment. If one wants to study the interactions of a solute with different solvents, it is necessary to construct an entirely new model every time, in which the particular solvent molecules are placed at sensible positions. The optimal solvent locations can usually not be determined *a priori* and thus often requires additional calculations.

4.2.4 Quantum dynamics in solution

It is clear from the previous discussion that describing environment effects is by no means straightforward. If one wants to study time-dependent properties of solvated systems, the task becomes more difficult still, and if one is interested in studying ultrafast, non-adiabatic processes, the available tools become very scarce. Before presenting a novel approach to achieve this goal, some comments are given on the specific difficulties one encounters, and on some existing and ongoing efforts that address this challenge. For more detail, a fairly comprehensive review of such approaches and associated difficulties is given in ref. [238].

Including the effect of an environment in non-adiabatic molecular dynamics simulations remains a formidable challenge, in part due to the inevitable increase in system size, but also due to the requirement of describing subtle effects such as the solvent response to electronic excitation of the solute. There are several factors to consider with respect to this last point, most notably the polarisation of both solute and solvent, and the structural rearrangement of solvent molecules (both local changes in the first solvation shell of the solute, and collective solvent rearrangements). Making matters worse, these effects occur on different length and time scales. Clearly, a complete description could not rely on implicit models since they are necessarily parametrised by some macroscopic properties, and do not contain a structural representation of the

solvent. Despite these shortcomings, implicit models remain a popular choice. To give an example, a traditional way of adding solvent effects in QD is to represent the environment Hamiltonian with a set of harmonic oscillators, referred to as "bath-modes", which couple to the solute and are parametrised from the spectral density obtained from classical MD simulations for example.[239] In fact, the *G*-MCTDH method was first developed with that application in mind.[133] Such an approach is effective when both the solute and the solvent undergo mostly harmonic motion, but breaks down as this approximation becomes less valid.

Going beyond implicit models and harmonic approximations, more accurate dynamics can potentially be obtained by propagating nuclei on a PES generated from single-point calculations using one of the explicit solvation methods described previously. This phrasing naturally brings to mind direct dynamics methods, where the propagation progresses not on pre-computed surfaces, but on PESs generated *on-the-fly* by stationary calculations at relevant points in the configuration space. Here, the challenge boils down to being able to perform accurate calculations of the energy and gradients of the system, and account for non-adiabatic couplings between the excited states of the solute. These properties are available from QM/MM calculations for example, and indeed, there are quite a few examples of classical trajectory dynamics (mostly TSH) which employ the QM/MM framework to study excited state dynamics in solvated molecules.[240–245] Alternative approaches include employing very efficient (but approximate) electronic structure methods, able to treat many explicit solvent molecules and hence doing away with QM/MM by employing a microsolvation approach, but with significantly more solvent molecules than are typically possible.[246]

4.3 FFQD outline

This section introduces the main ideas behind the newly proposed methodology known as Force Field Quantum Dynamics (FFQD). The aim of the development of this approach is simple; to provide the possibility of performing non-adiabatic QD simulations including explicit environment interactions. The basic principle of the method is nearly as simple as its aim, namely to combine the desirable features of both classical MD and QD methods by using FFs to calculate molecular properties, and the coupled vMCG EoMs to propagate the wavefunction. As discussed in the introductory chapter, MD

and QD lie at opposite ends of a spectrum in terms of the size of systems that are accessible and the ultimate accuracy of the results. If one succeeds in merging the computational efficiency of classical dynamics algorithms with the details accessible to quantum dynamics calculations, a new approach could be designed that provides the desired accuracy at a low enough cost that it can be used on large systems, such as solvated molecules. This of course is the holy grail of computational modelling in general; finding a cheap yet accurate way of simulating a particular system. The danger however, is that rather than an accurate and cheap method, one is left with a computationally expensive procedure that does not provide any detailed insights into the processes being modelled. In order to avoid the latter scenario, it is crucial that enough testing and thorough benchmarking is carried out using the proposed method. Such an evaluation shall be provided in this chapter.

In terms of implementation, a general outline is provided here and later expanded upon in section 4.4. FFQD uses the fully coupled, variational EoMs of the vMCG method (introduced in chapter 2, section 2.3.6) to propagate a wavefunction that includes, explicitly, a large number of environment DoFs. In order to make the simulations computationally feasible, the calculation of the required static properties (energies, gradients, Hessians, NACs) is carried out by the classical MD engine GROMACS[247] which performs simple MM calculations. An interface is created between GROMACS and QUANTICS such that these MM calculations can be used to build up the PES in a typical *on-the-fly* manner. A number of GWPs is then propagated on these surfaces as per the vMCG formalism. In effect, FFQD is simply a DD-vMCG simulation that uses MM rather than QM to perform the electronic structure calculations. However, some level of quantum-mechanical accuracy is retained in the treatment of the solute molecule by generating the intramolecular parameters (used in the FF) from QC calculations. A vibronic coupling model is also set up to determine the NACs between the states of interest, which of course are not available from classical MM calculations in GROMACS.

An important point worth highlighting is that the properties of the system are obtained from one single MM calculation, as opposed to using a layered approach like those outlined previously. The environment and the photo-excited species are thus both characterised by their respective FFs, the interaction of which is captured by the MM computations. An important consequence of this approach is that FFQD

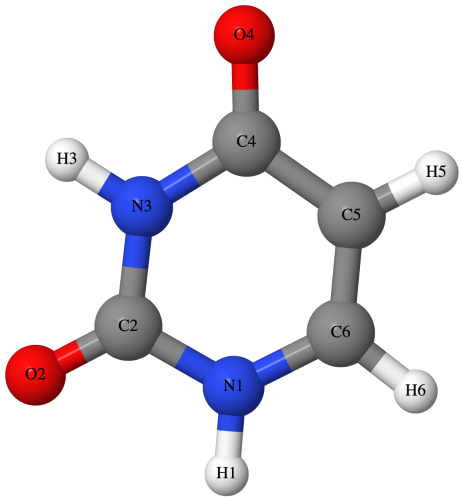


Figure 4.5: Equilibrium structure of Uracil, with atomic labels.

simulations are not considered to be open quantum systems since there is no dissipation from a system to an environment, in the sense that the environment is considered part of a closed system in which the norm and total energy must be conserved.

All the work done on FFQD, from the initial conceptualisation to the implementation and testing of the procedure, has been the result of a collaboration between UCL in London and the CNR in Pisa. This exchange program was funded by the Royal Society and resulted in the publication of a first contribution, detailing the work carried out so far.[218] The basic idea behind the proposed approached is explained and benchmark calculations show that the methodology, and its implementation, is sensible and can in principle achieve the goal of simulating the effects of an explicit environment on the ultrafast non-adiabatic QD of a solute molecule. This was a *proof-of-principle* of FFQD since at the time of writing the paper, no solvent molecules were included in the calculations. The contents of that work will be detailed in the following sections of this chapter, and additionally the results of some initial calculations that do actually include solvent molecules are presented.

4.3.1 Uracil as a benchmark

The performance of FFQD was evaluated in a study where the amino acid Uracil (referred to as Ura) is used as a test system. Ura (see Figure 4.5 for chemical structure) is one one of the 5 nucleobases in RNA and is closely related to thymine, which is Ura's replacement in DNA. Being one of the main components of the genetic code, Ura's

properties are of scientific and commercial interest. More specifically, the study of the effects of UV radiation on DNA, causing degradation and hence many harmful processes in living organisms, has been an active area of research for time-resolved spectroscopic studies, as well as computational work on the ultrafast excited state dynamics of the nucleobases.[248–251]

At present, there are still some open questions about various aspects of the photophysics of nucleobases. One such question is the role played by spectroscopically dark states during the excited state decay of Ura and other similar molecules. What is known is that the lowest energy bright state of Ura is characterised by a transition from a bonding π orbital to a π^* orbital (hereafter referred to as the $\pi\pi^*$ state). This state can be populated by UV light, causing an allowed electronic excitation. At this point, different non-radiative de-excitation pathways are open such as relaxation back to the ground state via a CoIn. If other electronic states are close in energy to the $\pi\pi^*$ state, internal conversion leading to population transfer between these states is also possible. Indeed, the presence of a low-lying dark state is unambiguously predicted by different electronic structure methods. This state is characterised by a transition from a non-bonding lone pair orbital of the carbonyl oxygens to the aforementioned π^* orbital (the $n\pi^*$ state) and lies close in energy to the $\pi\pi^*$ state (within 0.5 eV). Previous experimental and theoretical work seems to indicate that population transfer between the $\pi\pi^*$ and $n\pi^*$ states is not only possible but can be a major deactivation pathway in pyrimidines (such as Ura). Interestingly, the quantum yield of the population transfer is dependent on whether or not solvent is present, as well as its polarity. For Ura, it has been shown (both by continuum and explicit solvation) that the presence of a polar solvent limits the amount of $\pi\pi^* \rightarrow n\pi^*$ transfer because, compared to the gas-phase, the $n\pi^*$ state is destabilised.[252,253]

In summary, it is now fairly clear that the population transfer dynamics of Ura are affected by solvent effects, although one could still argue about the exact mechanisms. Whether or not direct solute-solvent interactions are important, compared to electrostatic effects for example, is not very clear. Nevertheless, this system is ideally suited to be used as a benchmark since there is a consensus about which states play a role in the population decay upon excitation to the first bright state. The observable of interest for a QD study of Ura is therefore obvious; the diabatic state populations of the ground state, and both the $\pi\pi^*$ and $n\pi^*$ states. However, in order to keep things

simple during the initial development of the FFQD method, a 2-state system of Ura was used where the ground state is ignored. This limitation may be explored in future work once the method is more fully developed.

4.4 Theory & Methodology

4.4.1 Building a QMD-FF

As hinted at in the introduction, classical FFs are used to perform MM calculations which provide the necessary data to obtain PESs. However, since the aim is to simulate an electronic excitation of a solute molecule, one cannot rely on the standard ground-state bonding and non-bonding interactions included in widely available FFs. Hence, a custom procedure is employed for each individual electronic state of a given solute molecule whereby the intra- and inter-molecular parameters are derived directly from quantum chemical calculations, yielding a quantum-mechanically derived FF (QMD-FF). Since the parametrisation of the QMD-FFs was done by a collaborator in Italy (G. Prampolini), only a brief overview of the procedure will be given. The original literature can be consulted for the full details.[254–256]

The JOYCE protocol developed in Pisa[256] automates the construction of a QMD-FF for each individual electronic state. Starting from the optimised geometry of a particular state, a QC calculation is performed to obtain the energy and its first and second derivatives. The calculation can be repeated at different geometries along some coordinate to obtain an energy profile of, for example, a torsional mode. This information is used to optimise the parameters that define the intramolecular FF. It is useful to express the DoFs in a set of redundant internal coordinates $\{\mathbf{x}_{\mathrm{RIC}}\}$, where $N_{\mathrm{RIC}} > 3N - 6$. Then, the intramolecular potential of an electronic state, S, is expressed as

$$\begin{aligned} E^{\mathrm{sol},S}_{QMD-FF}(\mathbf{x}_{\mathrm{RIC}}) = \; & E^{\mathrm{sol},S}_{stretch}(\mathbf{x}_{\mathrm{RIC}}) + E^{\mathrm{sol},S}_{bend}(\mathbf{x}_{\mathrm{RIC}}) + E^{\mathrm{sol},S}_{tors}(\mathbf{x}_{\mathrm{RIC}}) \\ & + E^{\mathrm{sol},S}_{nb}(\mathbf{x}_{\mathrm{RIC}}) + E^{\mathrm{sol},S}_{coup}(\mathbf{x}_{\mathrm{RIC}}) \quad . \end{aligned} \tag{4.8}$$

The first three terms on the right hand side are the bond stretching, the angle bending, and the dihedral torsion energies, respectively. These are all calculated using harmonic expressions, based on equilibrium values taken directly from the QC computations. The last two terms account for the interactions between two atoms not directly bonded,

and the coupling between two or more RICs, respectively. The JOYCE procedure then optimises the terms contributing to the expression of $E^{\mathrm{sol},S}_{QMD-FF}$ by minimising a merit function which takes into account certain QC and FF descriptors.

Importantly, since the procedure is always performed starting at the optimised geometry of a state S, the character of the state is assumed to be pure at this point. Hence, the QMD-FFs can be said to be diabatic and since vMCG dynamics are also run in the diabatic picture, the need for a diabatisation scheme is avoided. Additionally, the diabatic nature of the QMD-FFs means that the couplings between states can be added as a separate component, to allow non-adiabatic transitions during the simulations. These NACs can be computed by, for example, building a vibronic coupling model (see chapter 2, section 2.3.3) based on the same level of theory as the QC calculations used to parametrise the QMD-FF. Finally, one more feature of this procedure is highlighted related to anharmonicity. Even though the terms included in the definition of the total energy are based on harmonic potentials, this does not eliminate the possibility of anharmonic PESs being generated during the propagation. For rigid molecules, this is thanks to the redundancies of the internal coordinates used, and for more flexible molecules one can explicitly include anharmonic terms when building a QMD-FF.

In terms of environment effects in this framework, they are straightforwardly obtained through the interaction of the QMD-FF and any readily available FF in GROMACS. The energy of the system is given by the sum of the solute interactions and the intermolecular solute-solvent and solvent-solvent interactions:

$$E^{\mathrm{sys},S}_{QMD-FF}(\mathbf{x}_{\mathrm{sol}},\mathbf{x}_{\mathrm{env}}) = E^{\mathrm{sol},S}_{QMD-FF}(\mathbf{x}_{\mathrm{sol}}) + E^{\mathrm{sol-env},S}_{QMD-FF}(\mathbf{x}_{\mathrm{sol}},\mathbf{x}_{\mathrm{env}}) + E^{\mathrm{env-env}}_{FF}(\mathbf{x}_{\mathrm{env}}) \quad . \tag{4.9}$$

The last two intermolecular terms consist of the familiar 12-6 Lennard-Jones potentials and Coulomb interactions, calculated pairwise between any two molecules A and B:

$$E^{inter}_{AB} = \left(4\varepsilon_{AB}\left[\left(\frac{\sigma_{AB}}{r_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}}\right)^{6}\right]\right) + \left(\frac{q_A q_B}{4\pi\varepsilon_0 r_{AB}}\right) \quad . \tag{4.10}$$

As usual, r_{AB} is the internuclear distance between two nuclei A and B, and q is their charge. The depth of the potential well is represented by ε_{AB}, ε_0 is the (constant) permittivity of free space, and σ_{AB} corresponds to the internuclear distance at which the potential between the nuclei is zero (this is equivalent to the van der Waals radius).

4.4.2 Hierarchical approximations of (DD)-vMCG

Since FFQD is generally aimed at describing large systems, it is useful to examine ways in which the computational cost of propagating a high-dimensional wavefunction could be reduced. This section shows how the vMCG ansatz can be adapted for this goal using two distinct strategies. The first relates to the construction of the nuclear basis in terms of how the GWPs are parameterised, and the second involves certain approximations being introduced in the EoMs.

Starting with the GWPs, the idea of partitioning is introduced. Rather than a basis set composed of full-dimensional GWPs, G_i, the wavepackets are expressed as a product of lower dimensional GWPs,

$$G_i(\mathbf{x},t) = \prod_{n}^{N_p} G_{j_n}^{(n)}(\mathbf{x}_n,t) \quad , \tag{4.11}$$

where j_n is the index of the wavepacket in partition n, N_p is the number of partitions, and $\{\mathbf{x}\} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N_p}\}$ where each $\{\mathbf{x}_n\}$ contains k number of normal modes $\{q_1, q_2, ..., q_k\}$. Note that a direct product is taken between all sets of functions in all partitions, and hence a full set of configurations is generated. Plugging Equation 4.11 into the vMCG wavefunction gives the following ansatz,

$$\Psi(\mathbf{x},t) = \sum_{j_1 \cdots j_{N_p}} A_{j_1 \ldots j_{N_p}}(t) G_{j_1}^{(1)}(\mathbf{x}_1,t) \cdots G_{j_{N_p}}^{(N_p)}(\mathbf{x}_{N_p},t) \quad . \tag{4.12}$$

This splitting of the wavefunction into different sets of basis functions is equivalent to the G-MCTDH method, the only difference being that in Equation 4.12 all basis functions are GWPs, whereas the G-MCTDH ansatz may use both SPFs and Gaussian functions. Since the partition only relates to the basis functions, the EoMs for the expansion coefficients (chapter 2, Equation 2.62b) are unchanged and simply require replacing the single set, G_i, with the product in Equation 4.11. As for the evolution of the Gaussian parameters themselves, the EoM is repeated here (see chapter 2, Equation 2.65 for original description) for clarity since it will be discussed in detail in what follows (note the use of the "CX" formalism):

$$i\dot{\mathbf{\Lambda}} = \mathbf{X} + \mathbf{C}^{-1}\mathbf{Y}_R \quad . \tag{4.13}$$

When using a partitioned basis set, a difference does arise here related to the coupling between the GWPs. As opposed to direct variational coupling (through the second

term in Equation 4.13), the partitioned GWPs are linked by mean field operators introduced in the Hamiltonian matrix of $\mathbf{Y}$. The exact implementation of these operators will not be further discussed here, the working equations are derived in the original G-MCTDH paper[133] and remain unchanged when applied to the vMCG ansatz.

To see how a partitioned basis set leads to a reduction in computational effort, Equation 4.13 is analysed. Most of the effort in solving this differential equation goes into inverting the C matrix. The dimensions of this square matrix are $(j \times N)^2$ with j = nr. of GWPs and N = nr. of DoFs. Thus, by splitting up the GWPs into sets of lower dimensional functions (reducing N, and potentially j), the original matrix is split up into smaller matrices which can be more quickly inverted. A caveat here is the extra cost involved in building the mean field operators. Overall though, a well chosen partition results in a net increase in efficiency, especially as the size of the system increases.

Now, the second way to reduce computational effort is discussed. This involves introducing certain classical approximations in the EoMs of the Gaussian functions of vMCG. A useful property of Equation 4.13 is exploited, namely that it separates classical terms from higher order coupling terms. So, if the higher order terms are ignored, the EoM for the linear GWP parameter, ξ, becomes simply

$$i\dot{\boldsymbol{\xi}} = \mathbf{X} \quad , \tag{4.14}$$

which can be shown to further expand into

$$i\dot{\boldsymbol{\xi}} = -i2\boldsymbol{\zeta}\dot{\mathbf{q}} - \dot{\mathbf{p}} \quad , \tag{4.15}$$

where q and p are the position (given by the centre of the Gaussian) and momentum, respectively. Note that, since the widths, ζ, are kept fixed for the frozen Gaussians used in vMCG, the actual parameters being propagated boil down to

$$\boldsymbol{\xi} = \mathbf{q} + i\mathbf{p} \quad . \tag{4.16}$$

Finally, the EoMs of the position and momentum can be written down classically for a DoF α since all coupling terms are ignored:

$$\dot{q}_\alpha = \frac{p_\alpha}{M} \tag{4.17a}$$

$$\dot{p}_\alpha = -\frac{\partial V(x_\alpha)}{\partial x_\alpha} \quad . \tag{4.17b}$$

Notice that these expressions are truncated versions of the ones given in chapter 2, Equation 2.66. They can be plugged into Equation 4.15, in which case the GWPs will follow classical trajectories and the method is referred to as clMCG. Note however, that nothing has been done to the fully variational character of the expansion coefficients of the vMCG ansatz. Therefore, the wavefunction is still said to be quantum mechanical but since the trajectories followed by the classical basis functions are not variationally optimised, convergence will be slower and certain purely quantum-mechanical processes might not be properly described (e.g. tunnelling). This classical approximation is the first step down a hierarchy going from most exact to more and more approximate descriptions of the moving wavepackets. The clMCG formalism strongly resembles other methods which also employ classical trajectories, such as CCS, with which a detailed comparison is available.[257]

Staying in this classical trajectory framework, a further approximation can be introduced. Evaluating both the on- and off-diagonal Hamiltonian elements ($\langle G_i| \hat{H} |G_j\rangle$, for $i = j$ and $i \neq j$) is required to propagate the coefficients, and constitutes most of the effort. Therefore, any acceleration of this step could reduce the overall cost of the calculation. An analytical evaluation of the Hamiltonian is thus desirable. This is made possible by firstly expressing the kinetic energy operator in a sum of one-dimensional terms, which can easily be done if normal mode coordinates, q, are used:

$$\hat{T}^N(\mathbf{Q}) = \sum_{\alpha} -\frac{1}{2}\frac{\partial^2}{\partial q_\alpha^2} \quad . \tag{4.18}$$

This is the frequency scaled equivalent of the operator used in the LVC model (chapter 2, Equation 2.47).

Secondly, a *saddle-point approximation*[258] can be employed to obtain analytical expressions for the Gaussian integrals over the potential. This technique is commonly used in semi-classical Gaussian-based trajectory methods such as FMS and AIMS, and is thought to be valid because the basis functions are localised.[141,259,260] Taking the approximation to zero$^{\text{th}}$ order, the following simplification is made:

$$\langle G_i| \hat{H}^{el} |G_j\rangle \approx V^{el}(x_0) \langle G_i|G_j\rangle \quad , \tag{4.19}$$

where x_0 is defined by the centre of the product of two GWPs, i and j. Notice that only the zero$^{\text{th}}$ order term of the LHA, i.e. the potential energy at x_0, is used. This expression can be evaluated analytically since the integrals are reduced to Gaussian

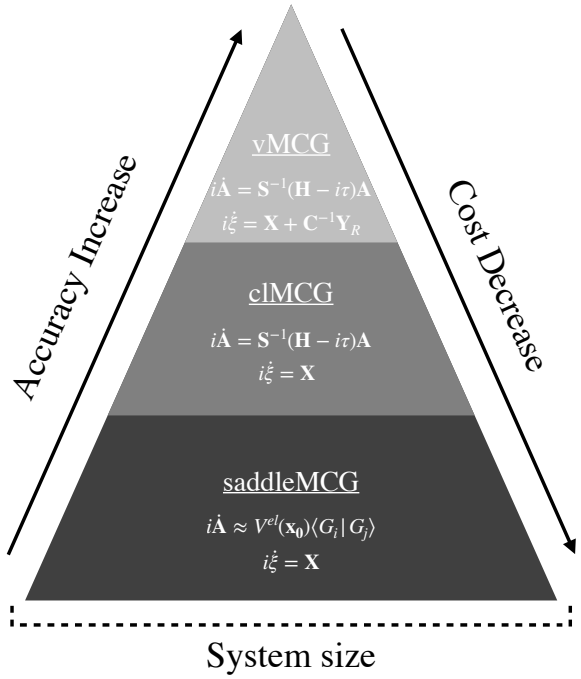


Figure 4.6: Summary of the hierarchy of approximations of the vMCG ansatz, and their EoMs.

overlaps, and are thus independent of the underlying PES. It is pointed out that an intermediary approximation could also be done by including the first order term of the potential (i.e. gradients), although this is not expected to improve the results and comes at a significantly higher cost. As it stands, the use of this approximation in vMCG calculations, referred to as saddleMCG, is expected to be particularly useful when dealing with large systems since the the second-order terms of the LHA, which are now ignored, grow quadratically with size.

Pictorially, the levels of approximations discussed above are represented in Figure 4.6. These, together with the partitioning of the nuclear basis set discussed at the beginning of this section, allow for a flexible choice of parameters with which to define the propagation of a system and hence an optimal balance between cost and accuracy can in principle be found for a given system size (represented in the Figure by the size of each layer). A very powerful aspect of these approaches emerges when the basis set partitioning is combined with the hierarchy of approximations, where different sets of GWPs are propagated according to different EoMs. This allows for hybrid methods to be used, giving even more flexibility in the cost-accuracy trade-off.

To give a concrete example, consider a 2-dimensional system with a partitioned

basis into two sets of 1-dimensional GWPs, a and b, that each consist of 2 wavepackets. The wavefunction can then be represented by writing out the ansatz of Equation 4.12 in full matrix form, with $\{\mathbf{x}\} = \{x_a, x_b\}$, $N_p = 2$ and 2 GWPs in each partition:

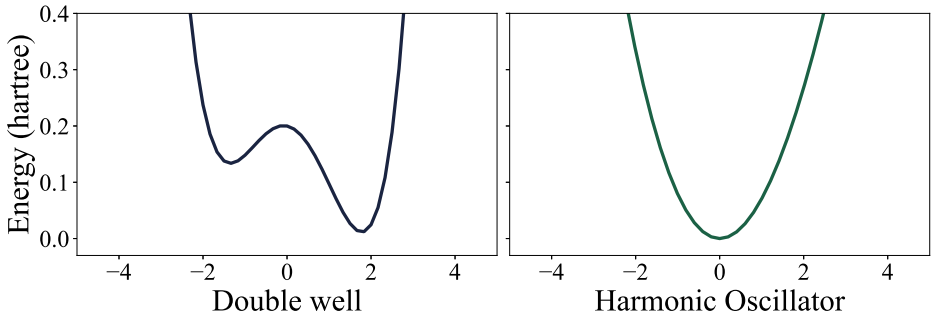
$$\Psi(\mathbf{x}, t) = \begin{bmatrix} G_1^{(a)}(x_a, t) \\ G_2^{(a)}(x_a, t) \end{bmatrix} \times \begin{bmatrix} A_{1,1}(t) & A_{1,2}(t) \\ A_{2,1}(t) & A_{2,2}(t) \end{bmatrix} \times \begin{bmatrix} G_1^{(b)}(x_b, t) & G_2^{(b)}(x_b, t) \end{bmatrix} \tag{4.20}$$

The EoMs for the two sets of wavepackets, each describing one particular coordinate, can then be chosen to be different. If one is mostly interested in the x_a coordinate for instance, the fully variational vMCG EoMs may be used for the GWPs in this set, while x_b is only described by wavepackets moving according to saddleMCG.

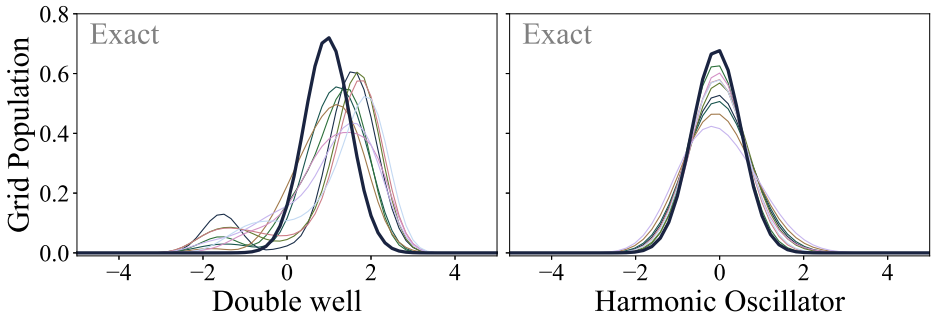
The simple 2-mode model is now used to perform a straightforward analysis of the performance of this approach. A system was set up mimicking this model, where x_a is defined as an asymmetric double-well potential and x_b is a harmonic oscillator potential (see Figure 4.7a). Three calculations were compared: an exact wavepacket propagation serving as a reference result, and two hybrid vMCG + saddleMCG calculations where x_a = vMCG and x_b = saddleMCG and vice versa, with 2 GWPs along each mode. The results are summarised in Figures 4.7b - 4.7d, where the grid populations of the Gaussian functions are plotted, with different coloured lines representing the population at different points in time (from 0-100 fs at 10 fs intervals, going from darker to lighter colours). Each simulation is initiated with the GWPs having zero initial momentum and initial positions of $x_a = 1.0$ and $x_b = -0.1$.

The expected behaviour of this system, confirmed by the exact propagation, is that some population is transferred from the right well (where the wavepackets start) to the left one. In the harmonic potential, the population remains centered around zero throughout the simulation. It is easy to see that in the hybrid calculations the mode described by coupled vMCG wavepackets recreates the exact result fairly well, both in the double well and the harmonic oscillator. In fact, as more wavepackets are added, it should converge exactly on that result. On the other hand, the wavepackets propagated using the saddleMCG method deviate significantly. When used to propagate GWPs in the harmonic well, much wider oscillations are observed, whereas in the double-well potential, no population is transferred to the other well because of the lack of coupling between the functions.

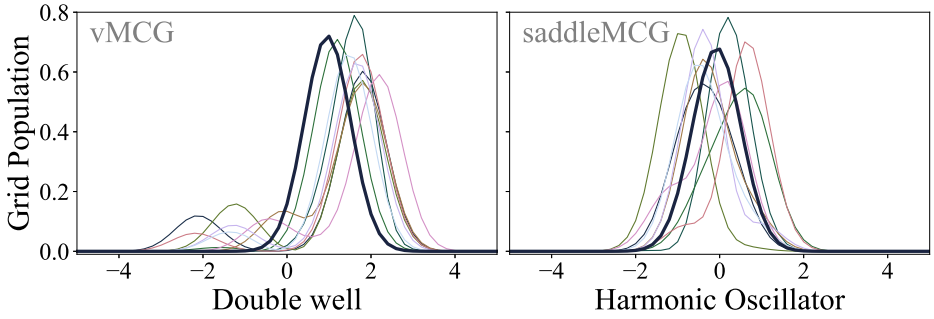
This result matches ones intuition since the correct description of the dynamics in a double well is expected to require a quantum-mechanical treatment of the nuclei in



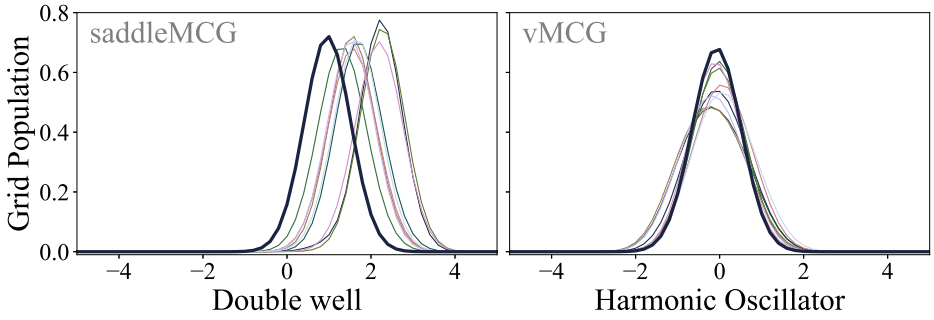
(a) Potential surfaces



(b) Exact propagation



(c) vMCG + saddleMCG



(d) saddleMCG + vMCG

Figure 4.7: (a) potential energy surfaces of the 2 modes included in the model system. (b) time-dependent grid populations (bold line: $t = 0$) of an exact wavepacket propagation, used as a reference for the hybrid calculations (c) and (d) in which one mode is described by 2 fully coupled GWPs, as per the vMCG method, and the other mode uses the saddleMCG approximation.

order to include tunnelling effects. Hence, the point of this example is to show that it is possible to make an *a priori* judgement, based on intuition, about which modes of a particular system should be described in a more quantum-mechanically accurate way (i.e. using normal vMCG) and which coordinates could be approximated using classical trajectories, with or without additionally using the saddle-point approximation. In more realistic, high-dimensional cases, this is of course much easier said than done, but this simple example nicely illustrates the idea of using hybrid calculations.

Summarising this section, a number of proposed approaches are put forward that should be able to reduce the overall computational cost of the rather expensive vMCG method. Eliminating the high-order terms in the EoM of the wavepacket parameters, which account for the coupling between wavepackets, leads to a method where classical trajectories are followed (clMCG). Further reducing the number of high-order terms – this time in the evaluation of the integrals – by means of a saddle-point approximation leads to the saddleMCG variation. Combining these methods with a partitioned nuclear basis set allows for hybrid calculations to be run, where any combination of EoMs is possible. This results in a hierarchy of approximations aimed at speeding up the propagation of very high dimensional wavefunctions in order to open up routes for the simulation of non-adiabatic dynamics of explicitly solvated systems. The implementation and validity of this proposed approach is demonstrated later on. First, the details of the workflow itself are presented.

4.4.3 Workflow - AQUATICS procedure

The workflow behind the FFQD method used here is referred to as the AQUATICS procedure (*"A Quantum dynamics Algorithm That Includes Complex Solvation"*). An outline is given of the steps required to set up the system, run the dynamics, and analyse the results. The process is summarised in Figure 4.8, in which three main stages are identified.

I. parametrisation of wavepackets:

The GWPs used in vMCG dynamics are parametrised based on the initial state (position and momentum) of the system at $t = 0$. The initial potential is given by an LHA based on the electronic energies, nuclear gradients, and Hessians of the entire system, obtained from a single-point MM calculation using the

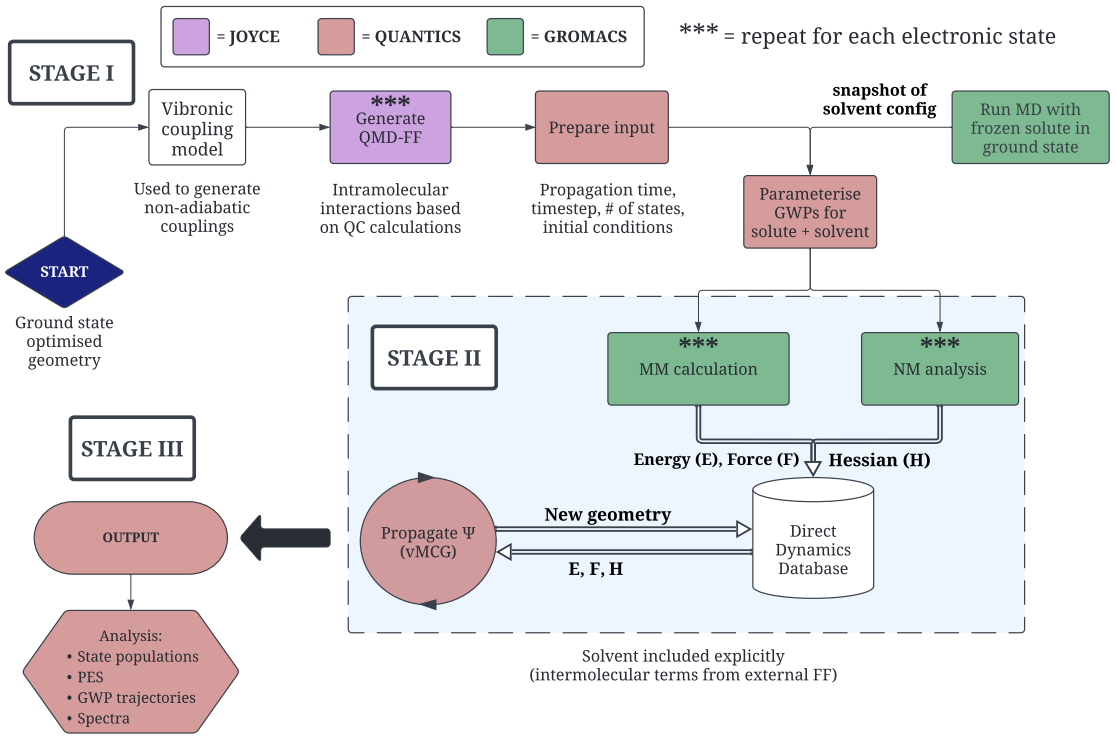


Figure 4.8: Flowchart of the AQUATICS procedure.

QMD-FFs. Additionally, a vibronic coupling model is needed to determine the NACs between the states. These couplings are not updated during the dynamics which is arguably one of the biggest limitations of the method. The vibronic coupling model is also used to obtain corrections for the initial vertical excitation. Since the total energy calculated by GROMACS corresponds to the excited state minimum, the vertical excitation energy out of the ground state minimum must be adjusted.

To describe the solvent, one can use standard FFs included in GROMACS (e.g. AMBER, OPLS, CHARMM...). A traditional MD simulation of the solvated system is run at the desired temperature in a canonical ensemble (NVT), but keeping the solute frozen. From this classical trajectory, a snapshot is extracted to provide the initial positions and momenta of the solvent atoms, which are then used to parametrise a number of solvent GWPs. It is of course necessary to sample more than one solvent geometry so one will need to perform several FFQD simulations, each time with a different snapshot extracted from the same MD

trajectory. Upon completion of all these individual simulations, any properties of interest must then be properly weighted (see stage III. below). Finally, it is noted that a benefit of having to run separate calculations for each initial solvent configuration is that they can be launched simultaneously, hence giving the method a kind of parallelisation.

II. Propagation of Ψ:

The propagation of FFQD works in exactly the same way as standard DD-vMCG simulations (see chapter 2, section 2.3.7). The GWPs are driven by the underlying potentials, defined by the output of single-point calculations stored in a database. Initially, the potentials are expanded around the FC point using the LHA, after which they are gradually updated as more and more points are added. There are a few important things to note however. In FFQD, the GWPs are partitioned into solute and solvent wavepackets. This has implications for the coupling between them, i.e. adding the requirement of mean-field operators. This was explained in detail in section 4.4.2. A further advantage of separating solute and solvent GWPs is that one can use standard normal mode coordinates to propagate the solute DoFs, while keeping the solvent in cartesian coordinates.

III. Analysis of results:

In the final stage of AQUATICS, the output of a FFQD simulation is analysed. Again, since the propagation is equivalent to a standard DD-vMCG calculation, so is the analysis. A full-dimensional wavefunction is obtained for the entire system, and hence any observable of interest can in principle be calculated. One of the most useful quantities that will be used throughout this chapter is the change in diabatic state population of the solute. Also of interest in the context of this benchmarking study is the inspection of the diabatic PES obtained from the *on-the-fly* MM calculations.

Although extracting useful data is typically straightforward, there is an important caveat related to the calculation of expectation values that ought to be discussed. A system surrounded by an environment is subject to quantum decoherence, causing some of its properties to quickly diverge. In FFQD, this decoherence manifests itself as the different initial conditions of different sampled solvent configurations. Hence, the total system should be regarded as an

incoherent superposition of those individual instances, whose wavefunctions deconstructively interfere.[261–263] Therefore, in order to properly account for the multitude of microstates of the environment, one has to average any calculated property over all the individual simulations. The most suitable strategy to do this has not yet been investigated, but since the system is thermalised and the snapshots are taken from a canonical ensemble, a Boltzmann distribution can be used to provide suitable weights for each simulation. Every property obtained from any one individual simulations is then simply multiplied by a normalised Boltzmann factor, w^{Bz}, given by

$$w_i^{Bz} = \frac{e^{-\frac{E_i}{RT}}}{\sum_n e^{-\frac{E_n}{RT}}} \quad , \tag{4.21}$$

where T is the temperature, E_i is the total energy of the system in configuration i, calculated in J mol^{-1}, hence why the ideal gas constant, R, is used. The normalisation is achieved through division of the weights by the sum of all n configurations. The averaged expectation value of a calculated property P is thus calculated as a weighted sum of the n individual expectation values of that property p_n:

$$\langle P \rangle = \sum_n w_n^{Bz} p_n \quad . \tag{4.22}$$

A more in-depth description of practical details related to the AQUATICS procedure can be found in the documentation, which includes a tutorial outlining the series of steps taken to perform the Ura simulations, presented hereafter. An initial draft of this documentation is included in Appendix C.

4.5 Uracil benchmark

As stated, Ura was chosen as a test system to evaluate the performance of AQUATICS. All simulations presented here are performed using the full-dimensional (i.e. 30 normal modes), 2-state model of Ura. The parameters used in the QMD-FFs constructed for the $\pi\pi^*$ and $n\pi^*$ states can be found in the original paper.[218] The optimised geometry and a description of the 30 normal modes of Ura are given in Appendix B (Tables B.1 and B.2). The vibronic coupling model used to obtain the NACs is also detailed there (section B.2).

Many different simulations are presented, all starting from an excitation to the bright $\pi\pi^*$ state. Comparisons between the performance of each calculation are made using the diabatic state population decay of the $\pi\pi^*$ state during the first 100 fs. The level of convergence and the time required to run the simulations are both used to define performance. It should be noted that no direct comparison to experiments will be made. It has in fact been argued that a detailed comparison to experiment would be rather pointless because of the fact that the correct shape of the excitation pulse used in experiments is not modelled.[252] Either way, the ultimate accuracy of the results is not of much relevance here, rather the comparative performance of different calculations with respect to a benchmark theoretical result shall be the focus.

This study is structured in such a way that different features of the AQUATICS procedure can be evaluated independently of each other. The first set of results aims to demonstrate the performance of the Gaussian basis set used in the vMCG ansatz, and is therefore compared to a ML-MCTDH calculation. These two methods are known to converge to the same result, provided that 1) the LHA is used, and 2) the same Hamiltonian of at most second order is used.[152] Additionally, a converged *G*-MCTDH calculation was carried out to get an idea of the size of the Gaussian basis required for convergence.

Having obtained these first benchmark results, the effect of using different kinds of partitions in the nuclear basis was tested. Then, once the optimal partition was identified, the performance of different classical approximations was compared. Finally, the actual FFQD procedure was evaluated by running *on-the-fly* simulations employing the QMD-FFs. Before any QD calculations were run however, the validity of the QMD-FF was checked for each state. First, the excited state geometries were optimised using MM calculations based on the Hessian derived from the QMD-FFs. These structures were then compared with corresponding structures predicted by regular QC optimisations. As a second point of comparison, the vibrational frequencies were also computed at MM and QC levels. The results of both these tests indicate that the FFs do indeed reproduce these properties, with only minor deviations from the QC results. The full description of these results is given in the original paper.[218]

4.5.1 Computational detail

Before presenting the results of the benchmark, the different calculations that were used are briefly explained, starting with a note on nomenclature. The following convention is introduced to label different calculations:
$M//(T_1 f_1/N_1, T_2 f_2/N_2, ..., T_n f_n/N_n)$, where

- M = QD method (vMCG or DD-vMCG)
- T = Type of GWP trajectory (V = variational, C = classical, S = classical and using the saddle point approximation)
- f = Number of DoFs (normal modes)
- N = Size of the Gaussian basis (nr. of GWPs)
- n = Number of sets of combined DoFs in a partitioned calculation.

ML-(G)-MCTDH and vMCG with partitions

Both MCTDH and vMCG require a pre-defined Hamiltonian. In this case, a suitable form for this operator can be derived from the QMD-FFs by generating the first order diagonal and off-diagonal terms (κ and λ in chapter 2, Equations 2.48 and 2.49, respectively), as well as second order bi-linear coupling terms between two normal modes on a particular state. This operator is in effect a quadratic vibronic coupling Hamiltonian, well suited for use in both QD methods. In order to ensure a proper comparison, the Hamiltonian used in the two calculations is of course kept identical. Since MCTDH has well defined convergence criteria, it can be used as a way to assess the level of convergence of vMCG simulations, for which this is harder to formerly define. However, the two methods use different nuclear basis functions, which is why the ML-MCTDH calculation was repeated with all primitive basis functions replaced by GWPs (i.e. ML-G-MCTDH). This provides a measure for the Gaussian basis size required for convergence. In that respect, it is important to consider the scaling of the method with respect to basis size since this will ultimately determine how large of a basis can realistically be achieved, and consequently whether or not convergence is possible.

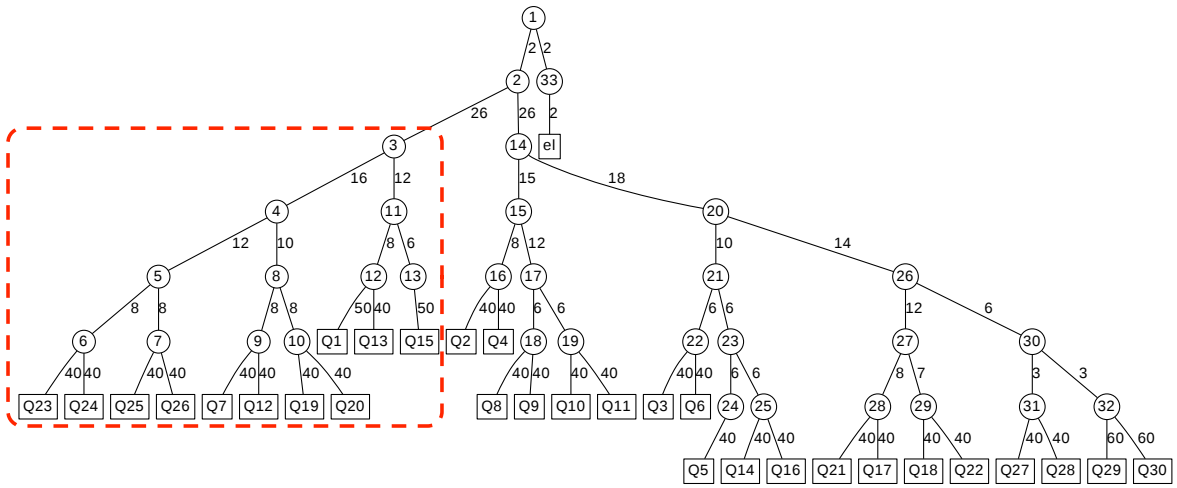


Figure 4.9: Multi-layer tree structure for the ML-MCTDH dynamics of Uracil on the vibronic coupling Hamiltonian. The numbers on the last layer correspond to the number of primitive harmonic oscillator DVR functions used. The red box indicates a branch containing the 10 system modes, and mode Q_{26}.

ML-MCTDH provides a way to reduce computational effort by partitioning the dimensions of the system in an appropriate way, effectively contracting the wavefunction representation. Similarly, this can be achieved in vMCG calculations. It is not trivial however, to choose how to split up the dimensions of a system. Some general rules of thumb exist though. For example, grouping together modes that are strongly coupled is desirable since the correlation between them can be treated in a more direct way. Some key modes that need to be grouped together are termed "coupling modes" and are characterised by a large value of the aforementioned λ parameter in the vibronic coupling model. Additionally, modes in which an excited state has a large nuclear gradient at the FC point should be grouped together. These are the so-called "tuning modes" and have a large κ value. The remaining modes, referred to as "spectator modes", are assumed to be the least important during the ultrafast dynamics following the initial excitation, and can therefore be partitioned together and be described by fewer basis functions than the more important modes.

This procedure is exemplified in the multi-layer tree used for the (G)-MCTDH calculations, based on the vibronic coupling model parameters derived for Ura (see Table B.3 in Appendix B for all parameters). The tree structure is shown in Figure 4.9. At the top, the tree shows a partition into two initial branches, one containing 11 modes (within the red box), and the other the remaining 19. Out of those 11

modes, the 10 most important ones were identified based on the strength of their κ or λ values, and were the following: $Q_1, Q_7, Q_{12}, Q_{13}, Q_{15}, Q_{19}, Q_{20}, Q_{23}, Q_{24}, Q_{25}$. These shall be referred to as the system modes, whereas the remaining 20 spectator modes are labelled as bath modes. Aside from this partition, the rest of the structure of the tree was optimised mainly through trial and error. Ultimately, fully converged results were obtained in less than 13h.

Assuming this multi-layer tree to be close to optimal, the same partitioning was also adopted in the vMCG calculations. To demonstrate the advantage of effective partitioning, three vMCG simulations were carried out with different sets of GWPs, and compared to the fully converged ML-MCTDH result:

- vMCG//(V30/61): No partition, 61 30-dimensional GWPs
- vMCG//(V10/61,V10/61,V10/61): Partitioned into 3 sets of 61 10-dimensional GWPs
- vMCG//(V10/61,V20/41): Partitioned into 2 sets of 61 10-dimensional and 41 20-dimensional GWPs.

Additionally, to show that the 10 system modes do indeed dominate the population transfer dynamics in Ura, a vMCG calculation was run where only those modes were included, and the 20 bath modes ignored:

- vMCG//(V10/61): No partition, 61 10-dimensional GWPs.

Classical approximations

Having identified a good partition that allowed efficient vMCG simulations of the population transfer in Ura, the effect of introducing classical approximations in the Gaussian EoMs was studied. The nature of these approximations is explained in section 4.4.2. Firstly, two calculations were run, one in which all GWPs followed classical trajectories and another in which, additionally, the saddle point approximation was employed for evaluating the integrals:

- vMCG//(C10/61,C20/41): All GWPs propagated on classical trajectories
- vMCG//(S10/61,S20/41): All GWPs propagated on classical trajectories and using the saddle point approximation.

Note that these calculations use exactly the same partition as vMCG//(V10/61,V20/41), mentioned earlier. Next, the very useful feature of the flexible Gaussian ansatz is exploited; the fact that two sets of GWPs can each be driven by different EoMs. Hence, another two calculations were carried out in which the 10 system modes are propagated using the fully variational ansatz, but the GWPs parametrised for the 20 bath modes obey the classical and saddle-point EoMs, respectively:

- vMCG//(V10/61,C20/41): A set of 10-D GWPs following variational trajectories and a set of 20-D GWPs following classical trajectories
- vMCG//(V10/61,S20/41): A set of 10-D GWPs following variational trajectories and a set of 20-D GWPs following classical trajectories and employing the saddle point approximation.

Direct Dynamics

Having performed all previously mentioned simulations using a model Hamiltonian, the *on-the-fly* procedure could finally be tested. A first trivial test was done to make sure the newly implemented interface between QUANTICS and GROMACS was working properly. A comparison was made between the vMCG//(V10/61,V20/41) result and a DD-vMCG calculation where the potentials are not updated, i.e. there is only one point, at the FC geometry, in the database and no further points are added during the propagation. Since the model Hamiltonian of the vMCG simulations is parametrised from this single point database based on the QMD-FFs, the results should be identical. Having confirmed this, the effect of updating the potential was evaluated through a comparison of a fully variational direct dynamics calculation to the corresponding vMCG result:

- DD-vMCG//(V10/61,V20/41).

Importantly, this calculation is not expected to converge to the vMCG result since the energies, nuclear gradients and state-specific Hessian matrices are updated *on-the-fly*, allowing the potentials to deviate from harmonicity. This is achieved through the use of redundant internal, curvilinear coordinates in the QMD-FF. In the end, just under 10 000 records were stored in the direct dynamics database. The Shepard interpolation between those points defines the global potential energy landscape of Ura.

Table 4.1: Computational effort of the reported Quantum Dynamics simulations in terms of CPU time (Dell PowerEdge R450 Server with Intel Xeon Silver 4316 2.3GHz 20Core CPUs and 384GB RAM).

Calculation type	**CPU time (hours)**
ML-MCTDH	12.7
ML-*G*-MCTDH	12.5
vMCG calculations (vMCG//):	
(V30)	494
(V10,V10,V10)	>1000
(V10,V20)	30
(V10)	4.5
vMCG with classical approximation:	
(C10,C20)	11
(S10,S20)	10
(V10,C20)	18
(V10,S20)	16
Direct Dynamics calculations (DD-vMCG//):	
(V10,V20)-No Update	65
(V10,V20)	140
(C10,C20)	57
(S10,S20)	24
(V10,C20)	88
(V10,S20)	36

The same database was subsequently used in all later DD-vMCG simulations, without further updating to ensure proper comparison between the results. The remaining calculations are the corresponding direct dynamics variations employing the series of classical approximations from the previous section, i.e.

- DD-vMCG//(C10/61,C20/41)
- DD-vMCG//(S10/61,S20/41)
- DD-vMCG//(V10/61,C20/41)
- DD-vMCG//(V10/61,S20/41).

4.5.2 vMCG benchmark

The results of the aforementioned vMCG calculations are presented here, in terms of the population decay of the initially excited $\pi\pi^*$ state. As a convention, a dashed

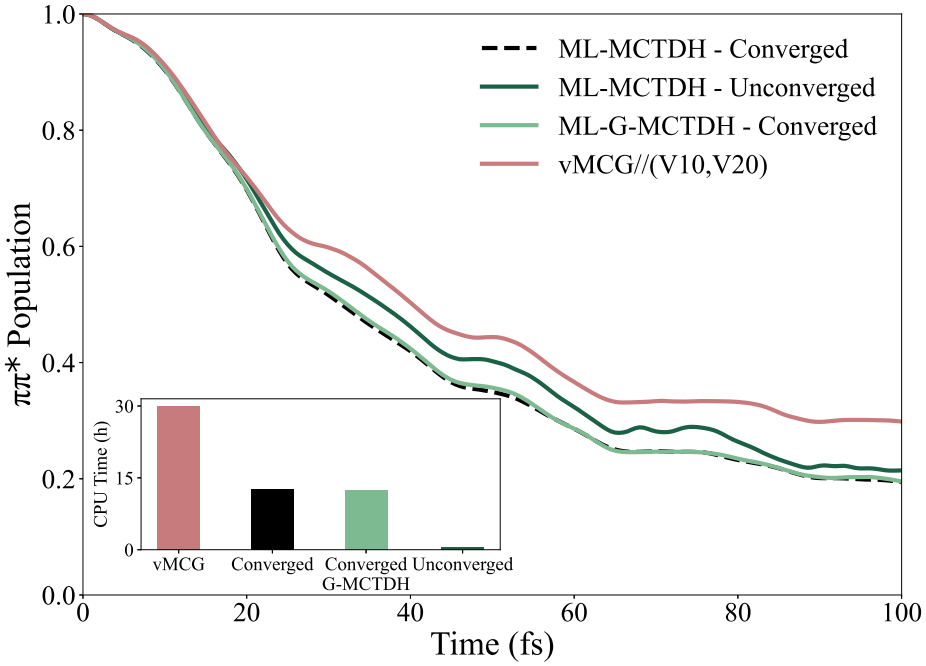


Figure 4.10: Calculated population transfer in Ura: ML-MCTDH, converged (black) and unconverged (dark green), ML-*G*-MCTDH, converged (light green) and vMCG (red). The computational effort is compared in the inset.

black line in the plots shall indicate the benchmark result, to which all others in the same plot should be compared. The total time taken to perform all simulations is summarised in Table 4.1, which will be referred to throughout the discussion of the results. The next few sections will focus on different sets of calculations in order to separately discuss important points, starting with the level of convergence.

Convergence

Since vMCG does not have well defined convergence criteria, it is useful to first obtain fully converged ML-MCTDH and ML-*G*-MCTDH results as an initial measure of convergence. This result is plotted in Figure 4.10, along with an unconverged ML-MCTDH calculation and the vMCG//(V10/61,V20/41) result using the system-bath mode partitioning. This plot provides three important insights

1. Comparison of the computational effort of the converged and unconverged MCTDH results reveals that achieving full convergence requires significantly more resources than nearly converged results.

2. The further a calculation is from convergence, the slower the population transfer.

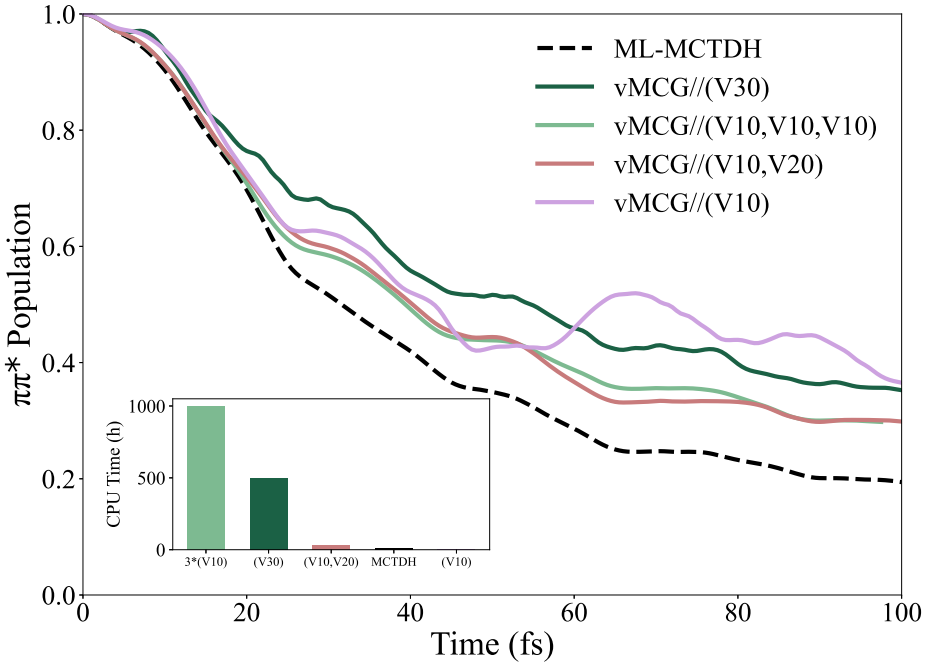


Figure 4.11: Calculated population transfer in Ura: ML-MCTDH benchmark (black, dashed), vMCG with 61 GWPs (dark green), vMCG with three sets of GWPs (ten modes and 61 GWPs in each) (light green), vMCG with two sets of GWPs (ten modes with 61 GWPs and 20 modes with 41 GWPs) (red), and vMCG with only the ten most important normal modes included (purple).

3. vMCG is not converged and, looking at the grid size necessary to converge the ML-*G*-MCTDH calculation (two orders of magnitude larger than the vMCG result), reaching convergence by enlarging the basis seems unfeasible.

With respect to the third point, it is worth noting that although Gaussian based approaches are typically harder to converge, they scale better than MCTDH with respect to memory, and are hence more applicable to larger systems, which is of course the ultimate application of FFQD.

Effect of partitioning

Having established a reference, one can now properly compare the performance of different vMCG calculations. The first set of results, plotted in Figure 4.11, shows that partitioning of the Gaussian basis can lead to improved convergence with fewer basis functions. Using the insights from Figure 4.10, one sees that the calculation with the fewest basis functions (i.e. vMCG//(V30/61)) is furthest from convergence. The two partitioned calculations give essentially the same result, despite the fact that their basis size, N, differs by a factor of 90 ($N_{(V10/61,V10/61,V10/61)} = 61{*}61{*}61 \approx 227,000$ and

$N_{(V10/61,V20/41)} = 61 * 41 \approx 2,500$). This further emphasises the 3$^{\text{rd}}$ point, namely that achieving converged vMCG results by simply increasing the basis size will not work in practice. Although this is known to be a problem in methods using Gaussians, it is particularly apparent in the case of Ura since there is no single dominant coordinate leading to a strong conical intersection. Rather, a number of modes contribute to the overall population transfer, each accounting for relatively weak non-adiabatic coupling between the two states (these are the 5 coupling modes with $\lambda > 0.01$, in Table B.3). Properly describing this weak coupling in multiple dimensions requires a large number of functions. This observation also explains the 2$^{\text{nd}}$ insight, obtained from Figure 4.10, that slower population transfer is a result of an incomplete basis that is not able to account for all the weak coupling effects.

Looking now at the purple curve in Figure 4.11, showing the results of a simulation where only the 10 system modes are included (all other modes are frozen), the choice of the system-bath partition is validated. At early times, the dynamics are indeed dominated by only those 10 modes. After about 50 fs, the results start to deviate and the bath modes start to play a bigger role. Again, this points to a situation where many different modes contribute to the overall dynamics in subtle ways, the implication being that it is necessary to include many, if not all, DoFs in the simulation.

Analysing the results now in terms of their respective computational effort, the advantages of partitioning become clear, but it is also apparent that one needs to take care when deciding how to split up the system. A suitable balance needs to be found between the dimensions of the basis functions and the number of partitions used. On the one hand, not having partitions saves effort because no mean-field operators need to be generated. The downside is that the effort required to propagate high dimensional basis functions may outweigh this saved effort. On the other hand, reducing the dimensions of the functions by having many partitions may not be efficient due to the mean fields required. This trade-off is clearly exemplified by Table 4.1, showing how the cleverly partitioned vMCG//(V10,V20) is much faster than both the vMCG//(V30) and the vMCG//(V10,V10,V10) calculations. Another rather trivial way to reduce the cost of a calculation is to simply reduce the dimensions of the entire system. The vMCG//(V10) calculation is by far the fastest but as discussed above, lacks some detail at later times due to the subtle effects of the ignored bath modes.

Ultimately, it is concluded that these results show that using QMD-FFs to parametrise

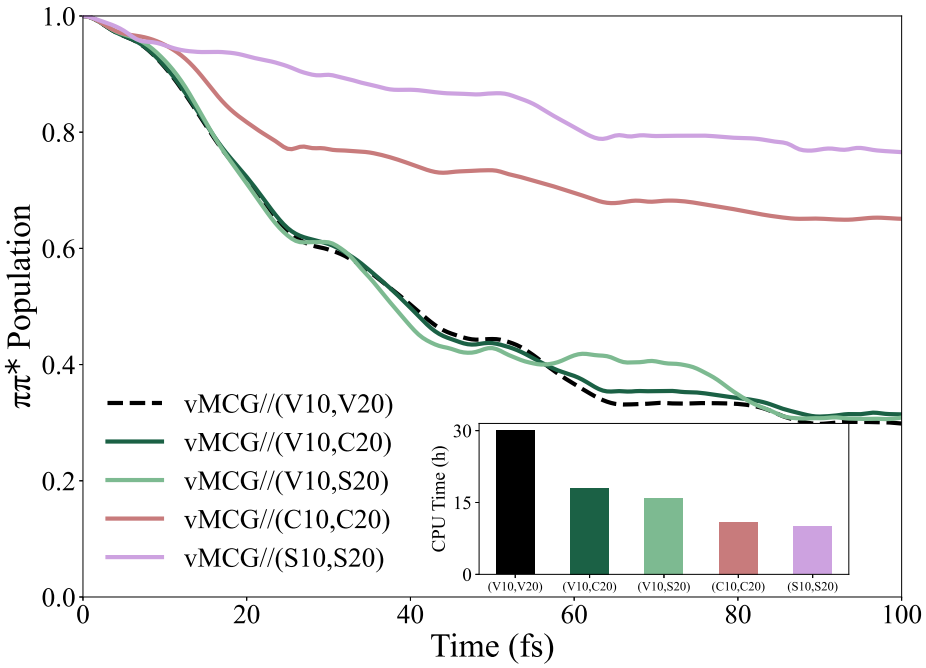


Figure 4.12: Calculated population transfer in Ura: fully variational vMCG benchmark (black, dashed), hybrid vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical ones (dark green), hybrid vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical + saddle point ones (light green), vMCG with all classical trajectories (red), vMCG with all classical trajectories and using the saddle point approximation (purple).

Gaussian wavepackets in the vMCG ansatz is a suitable way to describe the population transfer dynamics of Ura. Equivalent ML-MCTDH results are reproduced to a reasonable degree, despite the calculations not being fully converged. With a well chosen partition of the modes of the system into separate sets of GWPs, the computational cost of the simulations can be brought down significantly.

Effect of classical approximations

To further reduce the cost of the QD calculations, a number of simulations are compared where different levels of a hierarchy of classical approximations are explored. Presented in Figure 4.12, all of these calculations use the same system-bath partitioning as in vMCG//(V10,V20) which, being at the top of the hierarchy, is now the reference calculation to which the others should be compared. Clearly, propagating all GWPs on classical trajectories results in a large deviation from the expected result, regardless of whether the saddle point approximation is used. Much better results are obtained when the trajectories along the 10 system modes are variationally coupled,

and the bath modes are described classically. Importantly, the computational cost of the hybrid calculations is nearly halved compared to the full variational result. It can thus be concluded from these findings that using classical approximations is effective in reducing the overall cost of a vMCG calculation, but the accuracy may suffer significantly. Therefore, employing a hybrid scheme could offer an acceptable compromise between speed and accuracy, provided the partition is able to separate the more important modes from the modes that aren't heavily involved in the simulated dynamics. Thinking about extending this idea to solvated systems, a natural partition arises between solute and solvent, so one can imagine describing the small solute molecule with regular vMCG wavepackets, while the solvent is propagated on classical trajectories. This strategy shall indeed be used to obtain the preliminary results in section 4.6.

Comparison with TSH

To conclude this section, vMCG is compared to an equivalent calculation using TSH dynamics, still employing the same operator (performed in QUANTICS using the Zagreb surface hopping program, implemented by M. Sapunar and N. Došlić[264]). Two surface hopping simulations were performed, one with double the number of trajectories of the other (Figure 4.13). Firstly, notice the expected drop in computational cost compared to vMCG. A caveat here is that the simulations were run serially, as opposed to propagating all trajectories at the same time, in which case there should be no difference between the two TSH results and the calculation would finish in the time it takes to complete a single trajectory (< 2 minutes).

In terms of the quality of the description of the population transfer process, an interesting observation is made in that the population oscillates back and forth between the two states. In that regard, a completely wrong answer is obtained, but one could argue that by taking an average, the trend of the decreasing $\pi\pi^*$ population matches the vMCG result, and the final values after 100 fs are nearly identical. It is also noted that increasing the number of trajectories only makes the plot smoother, but does not change the observed behaviour.

One explanation for the oscillations is that they are caused by vibrational motion along one of the normal modes. Investigating this further, it was found that the the periodic motion of the $C_2 - O_2$ bond (mode Q_{26}) matched the period of the oscillations in the population transfer the best. Other coordinates probably also contribute to this

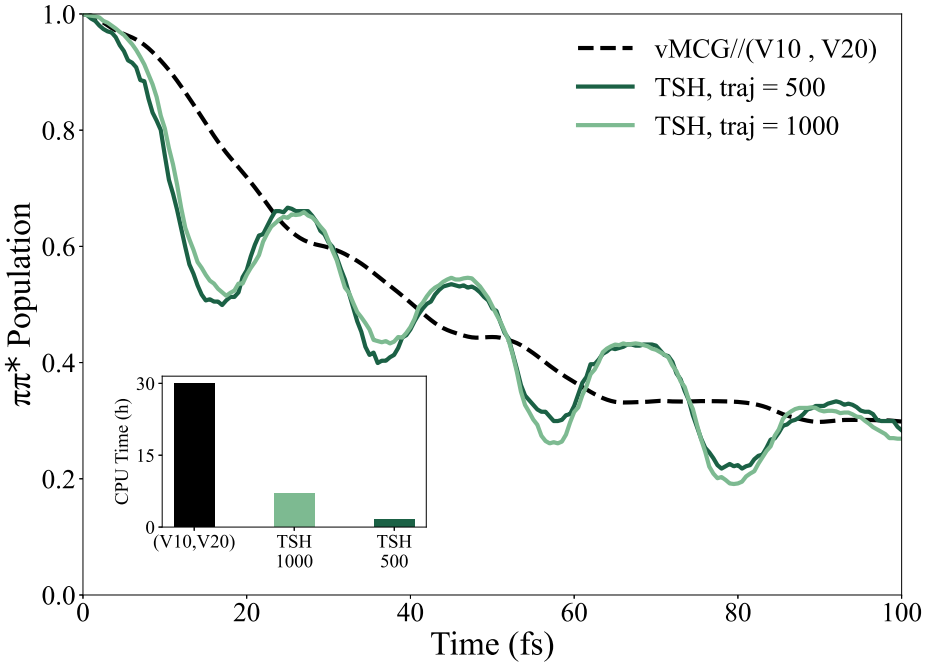


Figure 4.13: Calculated population transfer in Ura: vMCG benchmark (black, dashed), TSH with 500 trajectories (dark green), and TSH with 1000 trajectories (light green).

behaviour since there are other modes which also closely match the oscillations. In any case, it seems clear that surface hopping is unable to capture the decoherence of the nuclear wavepackets, thus predicting a regular crossing and re-crossing of the two states. Ultimately, one may conclude that a somewhat correct qualitative result is obtained, despite the wrong physics being simulated.

4.5.3 Direct dynamics benchmark

Finally, having thoroughly tested the performance of the QMD-FFs, the use of partitioning, and the use of classical approximations, the *on-the-fly* procedure could be evaluated. In contrast with the previously presented calculations, there is no benchmark result available that should be reproduced by the direct dynamics simulations. Since the potentials will be updated during the simulations, the population transfer dynamics are expected to deviate from the previously obtained results, which were run on purely harmonic potentials, parametrised at the FC point. As such, only relative comparisons can be made, and properties such as the level of convergence cannot be properly evaluated any more.

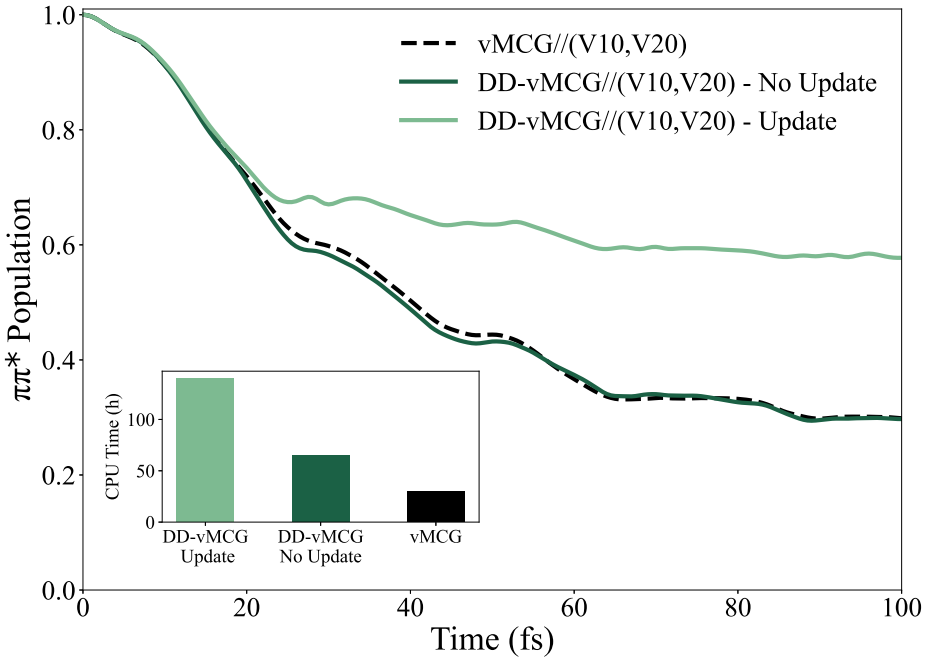


Figure 4.14: Calculated population transfer in Ura: vMCG benchmark (black, dashed), DD-vMCG with a single-point database, and no updating (dark green), and DD-vMCG with *on-the-fly* updating of the database (light green).

Effect of updating potentials

Firstly, a comparison is made between two DD-vMCG calculations, one where the direct dynamics database only has a single point (the FC point), and one where it is continually updated during the propagation, as per the usual *on-the-fly* procedure. The results are plotted in Figure 4.14, along with the vMCG//(V10,V20) result which is confirmed to be equivalent to the direct dynamics calculation without updating. When the PESs are updated during the propagation, essentially allowing the potentials to deviate from harmonicity, the effect on the population transfer dynamics is clearly seen. After about 20 fs of rapid decay, similar to the vMCG result, the rate of population transfer quickly decreases and continues to decay more gradually. To understand this observation, some selected cuts of the high-dimensional potential energy landscape are provided in Figure 4.15. Looking first at the bottom panel of the Figure, one notices how the surfaces along two of the non-totally symmetric modes go from being harmonic to a more quartic shape, effectively raising the energy of the intersection points between the two states. Similarly, the Minimum Energy Conical Intersection (MECI) is calculated, and found to increase in energy when the poten-

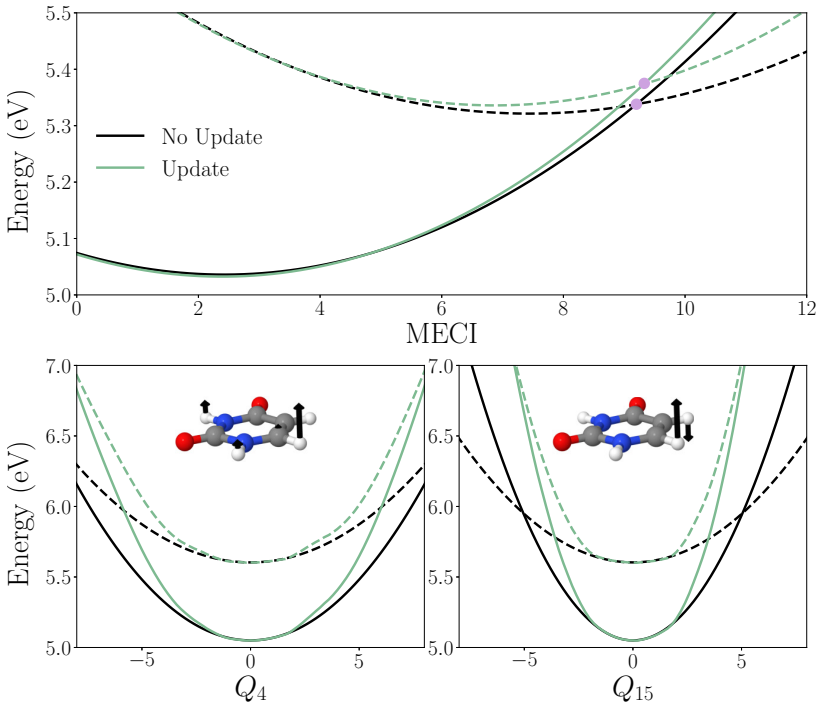


Figure 4.15: Potentials of the $n\pi^*$ (solid) and $\pi\pi^*$ (dashed) states of Ura after DD-vMCG simulations, both with and without on-the-fly updating. FC point located at $x = 0$. *Top*: cut showing the MECI (marked by a purple dot). *Bottom*: non-totally symmetric normal modes 4 and 15. Black arrows show the nuclear motion associated with the modes.

tials are updated. This MECI is characterised by a cut along multiple normal modes, with the highest contributions coming from the in-plane N_1–H_1 and $C_{5/6}$–$H_{5/6}$ wagging modes, the ring breathing mode and the C_5–C_6 stretching mode. Looking at the location of the MECI, one also notices it is relatively far from the FC point, meaning that it will take some time before any wavepackets are able to reach it. This is observed when carefully looking at the population decay during the first 10 fs, which is slower than during the next 10-20 fs. The drastic decrease in the rate of decay seen after the first 20 fs in the updated DD-vMCG calculation is simply explained by the higher energy intersections of the states.

Comparing the CPU times of the updated DD-vMCG calculation with the vMCG one, an almost 5-fold increase is seen for the direct dynamics. This is attributed mainly to the use of the Shepard interpolation between database points. However, as the size of the system increases, it is expected that this contribution to the total cost will become insignificant compared to the cost of the wavefunction propagation itself. At this point, the overall computational effort should become comparable to vMCG.

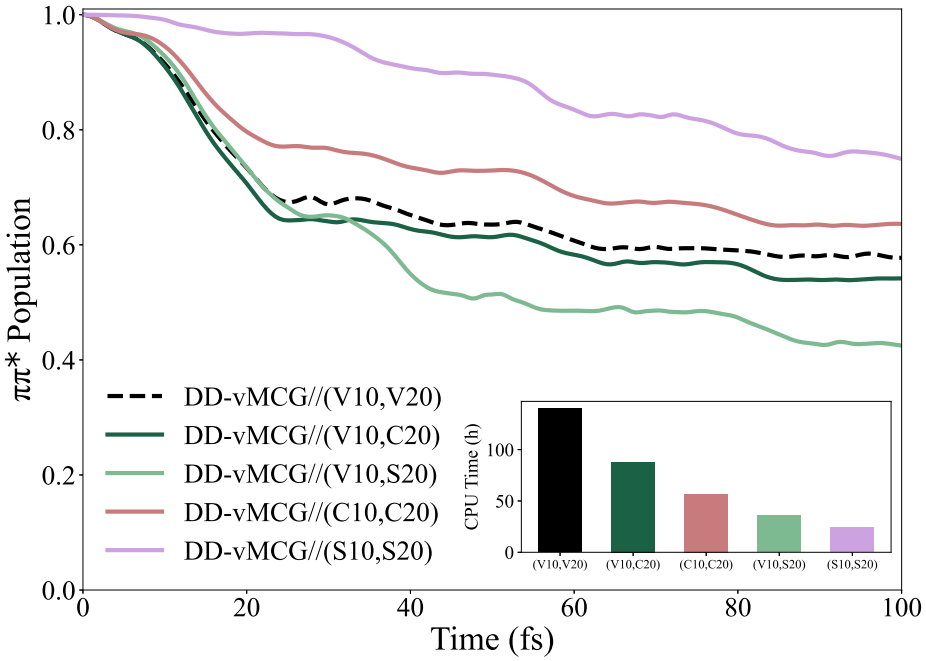


Figure 4.16: Calculated population transfer in Ura: fully variational DD-vMCG benchmark (black, dashed), hybrid DD-vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical ones (dark green), hybrid DD-vMCG with 61 GWPs following variational trajectories and 41 GWPs following classical + saddle point ones (light green), DD-vMCG with all classical trajectories (red), DD-vMCG with all classical trajectories and using the saddle point approximation (purple).

Effect of classical approximations

Just as with the vMCG calculations, different levels of classical approximations are used in the direct dynamics simulations. The results are plotted in Figure 4.16, and show a similar trend to the one seen in Figure 4.12. When all trajectories are classical, the performance is rather poor, especially if the saddle point approximation is also used. If, on the other hand, some important modes are described with variational trajectories, the results improve. Interestingly, the purely classical calculations in the direct dynamics case are closer to the benchmark than the corresponding vMCG ones, while the opposite is true for the hybrid DD-vMCG calculations, which don't reproduce the benchmark as well as their vMCG equivalents. Nevertheless, the description of the population transfer is fairly good, at least qualitatively, and when looking at the rapid decrease in computational cost, the use of more approximate calculations may be justified by the trade-off between speed and accuracy, depending on the system. As a final note, the fact that the hybrid calculations overestimate the population decay is a sign that the 20 bath modes are partly responsible for the slower rate compared

to the non-updated calculations, once again providing a reason to include them in the simulations.

To summarise, the application of FFQD in *on-the-fly* DD-vMCG calculations is presented and its validity in describing the population transfer of a 2-state model of Ura is demonstrated. While it is difficult to say whether or not the various conclusions will hold when applying the method to larger, more complex systems, it is comforting that all the results obtained can be rationalised both in terms of the chemical/physical processes they represent, and in terms of their computational performance. This ultimately increases confidence in the use of FFQD to model solvated systems, especially if a separation of important coordinates can be identified and hybrid calculations can be run, providing a good compromise between cost and accuracy.

4.6 Uracil in water

Following the publication of the original proof-of-concept,[218] the machinery to include actual solvent molecules in the simulations has been put in place. Some very preliminary results are presented here, but a lot of testing and benchmarking still needs to be carried out before the quality and efficiency of the simulations can properly be assessed. Therefore, the point of this section is mainly to show how a solvent is added in the AQUATICS procedure, and how this affects the scaling of the method, and what effects it might have on the population transfer processes in Ura.

4.6.1 Set up

A natural place to begin this discussion is with a description of how the solvent is actually included in the definition of the system. A description is provided of the three distinct steps that were taken to solvate Ura in water. For the sake of generality, some practical details are omitted, but these can be found in the aforementioned AQUATICS guide (Appendix C), in combination with the GROMACS and QUANTICS documentations. It ought to be emphasised that even though the present discussion deals with a specific solvent, in principle, any arbitrary environment defined in GROMACS can be added using the same general procedure.

1. Equilibrated MD simulation with frozen solute:

   A 1 ns classical MD trajectory is simulated in a canonical ensemble with T =

298 K using the ground-state QMD-FF to characterise the solute. The solvent is described using the TIP3P water model[265] included in the AMBER03 FF.[266] To insure consistent initial conditions for the solute, its atoms are kept frozen during the simulation. All snapshots are taken from this simulation.

2. Extract snapshot and define N^{solv}:
   A single frame of the MD trajectory is chosen and the coordinates and velocities of all atoms are extracted. Next, the number of solvent molecules is reduced to the desired number by only keeping the N^{solv} closest molecules to the solute. This step is then repeated for different snapshots in order to model the incoherent superposition of different solvent geometries by sampling their initial conditions.

3. Partitioning of solute and solvent basis functions:
   As before, the GWPs describing the solute are partitioned into 10 system and 20 bath modes. The 10 system modes are propagated at regular vMCG level, while the bath modes use classical trajectories. The solvent is described by a single set of GWPs using the saddleMCG approximation.

Having set up the system is this way, the same FFQD simulations can be performed, but now the MM calculations in GROMACS include explicit solvent molecules. The resulting diabatic energies, gradients (in the form of forces), and Hessians of the entire solvated system are stored in the database and the propagation continues. It is emphasised here that any solute-solvent interactions are due to the inclusion of partial charges of the solute QMD-FFs, fitted separately for each state (see Appendix B, section B.3 for specific values).

Two different solvated calculations will be presented, and as before, the population decay of the $\pi\pi^*$ state will serve as the central point of discussion. In the first set of simulations, the effect of specific solvent configurations is evaluated by considering a total of 20 H_2O molecules, described by 4 GWPs, and carrying out 5 separate calculations with 5 different initial configurations sampled from the classical trajectory. After that, the dependence of the rate of internal conversion on the number of solvent GWPs, N^{solv}_{GWP}, is tested by including only 4 water molecules, but gradually increasing the size of the basis set. In both cases, the number of GWPs used for the solute was reduced to compensate for the additional computational cost. The system modes

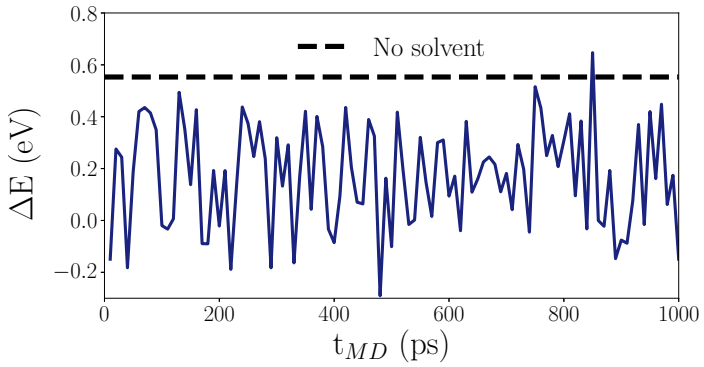


Figure 4.17: Energy difference between $\pi\pi^*$ and $n\pi^*$ states of Ura calculated for 100 snapshots (with $N^{solv} = 20$) taken at 100 fs intervals of a 1 ns MD simulation. Dotted line shows energy difference of the states without any solvent present.

are described by 21 functions, whereas the bath modes use 11. This partition is also adopted by a new reference calculation without any solvent, to which the solvated results are compared (i.e. DD-vMCG//(V10/21,C20/11)).

4.6.2 Effect of solvent configuration

The static configuration of solvent molecules is known to affect the non-adiabatic dynamics of an excited solute molecule (referred to as the static disorder effect).[228–230] To show this effect in the present case, a series of 100 snapshots is taken from the classical MD trajectory and the nearest 20 water molecules are kept in each case. Using these initial geometries, the energy of the two states of Ura are calculated. In Figure 4.17, the energy difference is plotted to reveal the effect of randomly sampled geometries on the energetics of the system.

The observed variations indicate that on average, the energy gap is reduced when Ura is solvated by water. This is indeed expected and is the result of hydrogen bonding to the carbonyl groups in the molecule, which destabilises the $n\pi^*$ state, bringing it closer to $\pi\pi^*$.[252,253,267] Notice that in some configurations, the destabilisation is so pronounced that the state ordering is flipped. This makes it abundantly clear that the addition of solvent is certainly expected to influence the population transfer between these two states.

The results of FFQD simulations using a selection of only 5 of the previously extracted snapshots (still with $N^{solv} = 20$) are shown in Figure 4.18. Note that for the benchmarking of large solvated systems, only the first 50 fs of dynamics are

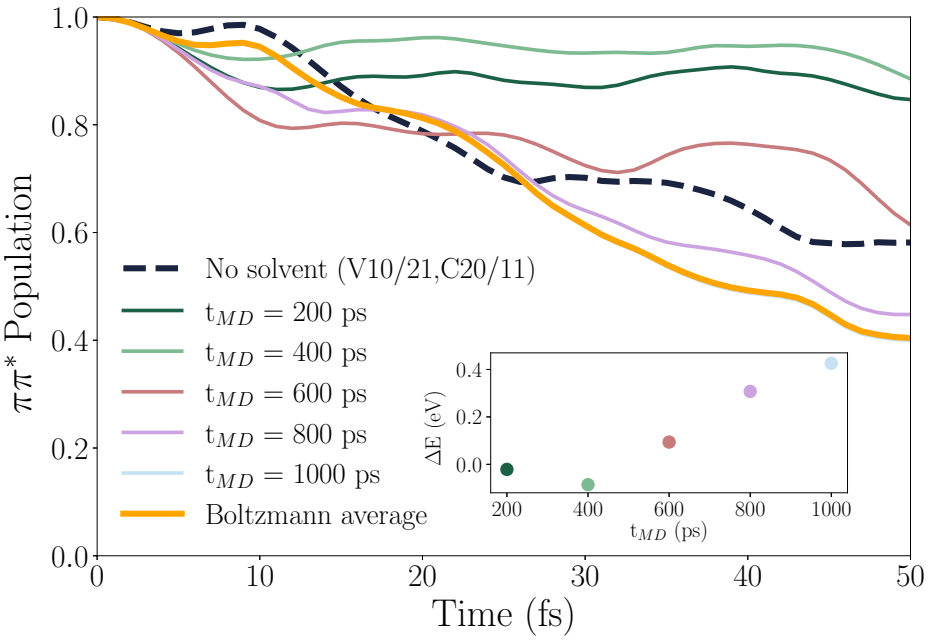


Figure 4.18: Calculated population transfer of Ura + 20 H_2O: hybrid DD-vMCG with 21 GWPs following variational trajectories and 11 GWPs following classical ones (black, dashed), and solvated calculations starting from different initial configurations sampled at 200 ps intervals (coloured lines). Boltzmann averaged density is shown in orange.

simulated rather than 100 fs. The inset in this case shows the specific energy gap for each initial configuration. The computational cost of all calculations was essentially the same, taking 410 h to run 50 fs on a single CPU. In these simulations, the static disorder effect is immediately clear given the range of different internal conversion rates observed. These variations can be explained by the obvious correlation between the energy gap of the electronic states, and the amount of population transferred between them. The smaller the gap, the slower the transfer.

In order to take into account the incoherent superposition of the solvent, the data corresponding to the 5 individual calculations needs to be consolidated by properly averaging them, yielding a single result for the Ura + 20 H_2O system. As stated previously, the most suitable procedure has not yet been worked out, but a good starting point is to simply take the normalised Boltzmann weighted average of the expectation values calculated for each separate calculation (see section 4.4.3). The Boltzmann factors are calculated using the total energy (kinetic + potential) of the system in its electronic ground state (i.e. using the ground state QMD-FF). These energies and corresponding factors are given in Table 4.2 for the 5 snapshots, denoted by the time-step

at which they were extracted (t_{MD}). The weights in the last column of the table are

Table 4.2: Total energy (kJ mol^{-1}) of solvated Ura for different initial solvent configurations, calculated using the QMD-FF of the ground state. Normalised Boltzmann factors are calculated in the last column, with $T = 298$K.

t_{MD} (ps)	E_0	w^{Bz}
200	-243.10	0.00
400	-321.48	0.01
600	-248.96	0.00
800	-308.81	0.00
1000	-340.25	0.99

used to obtain the average population density of the $\pi\pi^*$ state, corresponding to the orange line in Figure 4.18. As expected, the dominant configuration is the one with the lowest energy, but it is surprising that all other configurations have negligible weights. The maximum energy fluctuation is about 100 kJ mol^{-1} which roughly corresponds to bond energies, and thus seems unreasonably high for an equilibrated simulation at 298 K. Intuitively, one would expect each configuration to have almost equal weights. Further analysis of these energies is thus required but has not yet been carried out at the time of writing. However, one current hypothesis is that the frozen solute molecule causes artificially large fluctuations in the energy during the MD simulation.

4.6.3 Effect of solvent GWPs

Although it is unlikely that the current FFQD calculations can ever be fully converged, it is worth investigating how the number of solvent basis functions affects the overall rate of population transfer. In Figure 4.19, this dependence is plotted for the first 100 fs of FFQD simulations using $N^{solv} = 4$. Although no reliable benchmark result can be provided here, the effect of adding more basis functions to describe the solvent is clearly relatively minor (at least within this limited time-frame). The computational cost however, does increase almost linearly as expected. On the one hand, this is a promising conclusion since it suggests that describing the solvent with relatively few GWPs achieves similar results to larger basis sets. On the other hand, it is possible that to properly describe the many solvent DoFs a very large number of basis functions is required, making such calculations unfeasible. Those considerations shall be a main focus of the ongoing future development of FFQD.

Finally, it is pointed out that based on experimental evidence, the presence of water

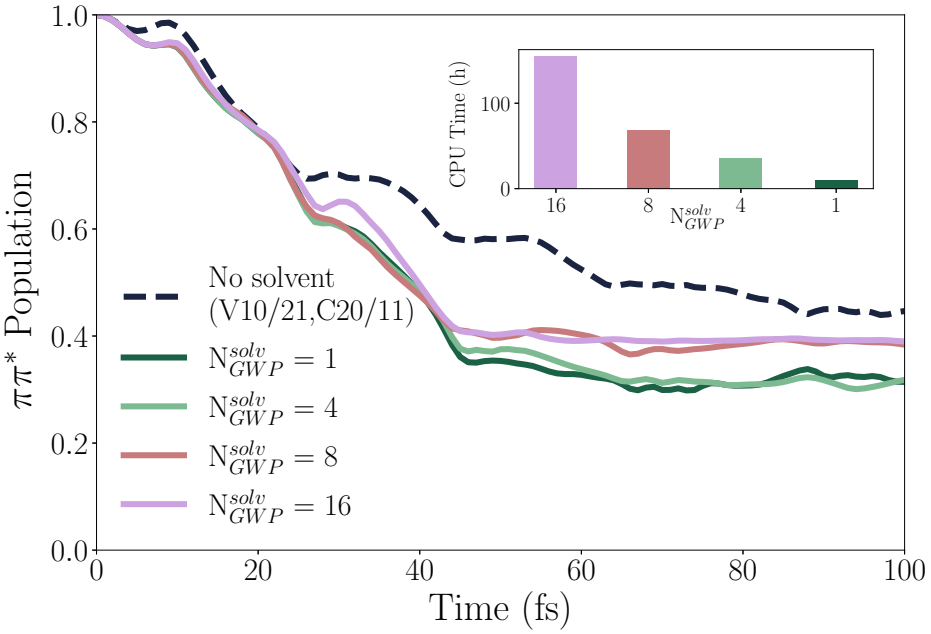


Figure 4.19: Calculated population transfer of Ura + 4 H_2O: hybrid DD-vMCG with 21 GWPs following variational trajectories and 11 GWPs following classical ones (black, dashed), and solvated calculations with increasing number of solvent GWPs (coloured lines).

is expected to slow down the populations transfer from $\pi\pi^* \rightarrow n\pi^*$ in Ura.[252,253,267] In the case of Ura + 20 H_2O, the calculated average result suggests the opposite, even though three of the individual calculations do predict a slower rate than the gas-phase result. It must be said that the preliminary nature of these results makes it difficult to judge whether or not the relevant effects of the solvent are properly described by the simulations. However, a correlation between the energy gap and the rate of internal conversion makes it clear that when the two states are close together, population transfer is slowed. It is expected that the addition of more water molecules will further decrease this gap, thus slowing down the population transfer in a larger proportion of the sampled solvent configurations. This is somewhat supported by the fact that the simulations of Ura + 4 H_2O show a faster decay in $\pi\pi^*$ density than Ura + 20 H_2O, but much more thorough comparisons will need to be made in order to determine how many solvent molecules are ultimately needed to represent bulk solvation, and how many initial configurations should be sampled.

4.7 Conclusion & Outlook

Non-adiabatic QD simulations have proved to be very useful tools to explore and understand, in great detail, the nature of photophysical and photochemical processes. Unfortunately, their poor scaling with respect to system size means that most applications of such simulations are limited to single molecules in a vacuum. This chapter addresses the need for methods that are able to go beyond the gas-phase, but still retain an accurate description of possible photoinduced processes. The central idea is to simulate an explicit environment without the increase in system size rendering the QD calculations unfeasible. The presented approach, referred to as FFQD and the AQUATICS procedure, is based on two fundamental ideas that aim to break the two biggest bottlenecks in QD simulations: scaling of nuclear dynamics propagation, and obtaining accurate potentials. The first bottleneck is addressed with the possibility of partitioning a system into different sets of DoFs within the vMCG ansatz. Provided a sensible partition is made, this can greatly reduce the overall computational cost. Furthermore, the fact that the coordinates in separate partitions can be propagated using different variations of the typical EoMs in vMCG leads to a hierarchy of classical approximations, which provides a great deal of flexibility with respect to the cost-accuracy trade-off.

The second bottleneck is avoided through the use of MM to provide the necessary properties for the wavepacket propagation, i.e. potential energies, nuclear gradients and Hessian matrices. This allows one to exploit the knowledge acquired in the last few decades in the field of MM and overcome the typical electronic structure problem, admittedly at a lower level of accuracy. By generating tailored QMD-FFs for different excited states of small solute molecules, the powerful MM engine GROMACS can be used to include any available solvent, or indeed any molecular environment for which a FF exists, fairly trivially. Although this work focuses on solvated systems, future applications could certainly also include surface chemistry,[268] or the study of excitations in molecular aggregates.[269]

It is also worth emphasising that the properties of the entire system are being calculated by a single MM calculation. This means that there is no need for additive or subtractive schemes to obtain the correct total energy, like in QM/MM or other layered models. Consequently, problems associated with interactions at the boundary

between layers are avoided, or rather, replaced by interactions between two separate FFs. Similarly, the propagation itself treats the entire nuclear wavefunction as one single quantum-mechanical object, albeit with a partitioned basis set and using some classical approximations.

Most of the work presented here serves as a proof-of-principle, and although the full potential of the proposed approach has not been explicitly demonstrated, some exciting avenues for further development have been opened up. The fact that a full 100 fs quantum dynamics simulation of a 30 mode molecule described by 2.500 basis functions could be performed on a single CPU core in 140h (DD-vMCG//(V10/61,V20/41)), and that a similar result was achieved in a quarter of the time when some approximations are used (DD-vMCG//(V10/61,S20/41)), is a promising sign that this approach can be pushed to significantly larger systems. Indeed, preliminary results were presented where AQUATICS was applied to propagate a solvated system including 20 water molecules (a total of 210 nuclear DoFs), but it is hard to judge the quality of the results at this stage. Nevertheless, it is clear that including a solvent environment around Ura has a pronounced effect on the amount of internal conversion between the $\pi\pi^*$ and $n\pi^*$ states, mostly due to the modulation of the energetic gap between the states. Propagating this system for 100 fs currently takes just over a month of CPU time on a single core, but there are some inefficiencies in the current implementation that can be relatively easily addressed to achieve better computational performance. These include sub-optimal procedures associated with reading and writing information to and from the database, as well as a lack of parallelisation when it comes to building the necessary mean-field operators which account for the interaction between different basis set partitions.

Despite the promising prospects of FFQD, some important weaknesses ought to be mentioned. The first relates to molecules that are very flexible, e.g. long saturated carbon chains, in which case it will be more difficult to generate a suitable QMD-FF since all the torsional parameters will have to be defined accurately. Alternatively, care will also have to be taken when studying molecules which have unbound exited states. Including dissociative motion in a FF is not straightforward and prior quantum chemical calculations will have to be carried out to parametrise the functional form of such a potential. On a different note, it may also get increasingly challenging to obtain appropriate coupling vectors between the states of interest as the system

size increases. Currently, NACs are obtained from a vibronic coupling model, but in principle, couplings may be used from any available source (e.g. from previous work, or using approximate methods based on, for example, the energy gap and its second derivative,[270] or curvature-based methods[271,272]). A final important weakness relates simply to the fact that the interaction between system and environment is based purely on interacting classical FFs. This means that potential covalent interactions become difficult to describe, as well as the proper inclusion of polarisation effects (now approximated as changes in partial charges of the solute).

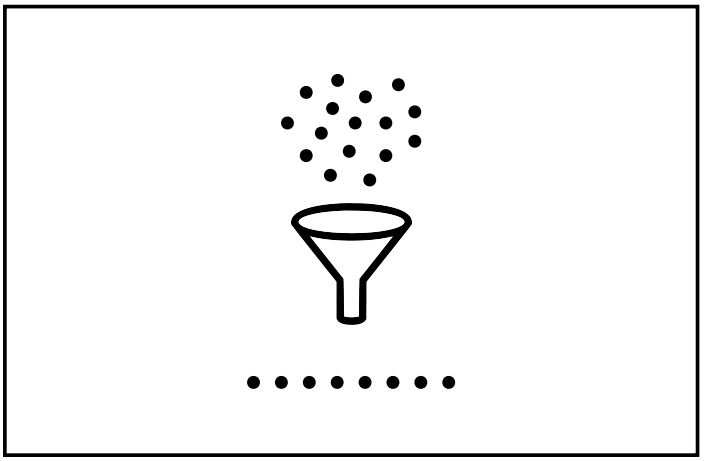
Keeping in mind that this work is explorative, further extensions to the methodology that increase its accuracy are anticipated. An example would be to remove the need for QMD-FFs by actually performing *on-the-fly* electronic structure calculations on the solute molecule. The energy of the solvated system could then be evaluated by MM calculations in GROMACS, after which a typical substractive QM/MM scheme could yield the final, corrected total energy according to Equation 4.6. Alternatively, a way in which solute-solvent interactions could be more accurately modelled within the FFQD framework involves the use of different electrostatic embedding schemes based on, for example, higher order multipole expansions. However, these would have to be fitted for each electronic state of the solute, thus introducing additional practical and computational challenges. An exciting avenue which alleviates those kinds of problems is the use of machine learning techniques to accelerate electronic structure calculations and/or the evaluation of couplings (see the recent application of such methods to non-adiabatic simulations of a molecular crystal[273]). Such approaches are becoming increasingly popular and it is expected that they will be invaluable in the quest to accurately describe the dynamics of large systems (this is briefly discussed in the next chapter).

Before exploring those directions, the current AQUATICS procedure will need to be further refined, and its limits properly tested in order to reveal the most important weaknesses, and by extension, the most promising future applications. Once the implementation is fully optimised, a full benchmark study can be carried out, similar to the one performed for Ura in the gas-phase.



CHAPTER 5

Towards accessible quantum dynamics workflows



5.1 Motivation & Context

The penultimate chapter of this thesis differs slightly from the preceding ones, both in style and scope. Rather than presenting published results, the main purpose of this chapter is to outline the procedures involved in successfully performing QD simulations, and through this analysis, shed light on some of the biggest outstanding challenges in the field. As such, the work presented here should be seen as a starting point for further consideration and development. Necessarily, the text will have a slightly more personal tone since the chosen topics only represent a fraction of the possible aspects of QD worth discussing, and were chosen based on personal experiences during the research period. However, many of the topics included here reflect a general sentiment present within the community of users and developers of computational QD methods.

Recently, two community-wide initiatives were undertaken which clearly identify major challenges in the field, and these will be a reference point for the upcoming discussions. The first was a challenge aimed at the QD community in which the goal was to predict the outcome of an upcoming experiment, hence testing the predictive power of QD simulations.[274] Many people responded to this challenge, each with a different combination of both electronic structure and dynamics methods. The outcome of this prediction challenge was ultimately not conclusive due to the difficulty of comparing different methods to each other properly, and the fact that the underlying electronic structure calculations do, as usual, have a large impact on the QD simulations. Additionally, comparison to experiment proved to be less than ideal for the purpose of evaluating the performance of the different calculations. Drawing attention to these issues is important and this challenge did exactly that, strengthening the consensus among the community that initiatives like these are necessary for the field to progress. One cannot straightforwardly trust the outcome of a given simulation, regardless of what method is used, and hence the search for ways to properly validate results, beyond comparison to experiment or to potentially very different computational methods, should be considered a priority.

The second initiative mentioned here is the development of a general roadmap towards the benchmarking of non-adiabatic QD methods. This work describes the outcome of a meeting of users and developers of QD methods, and is currently under

review for publication in the Journal of Physical Chemistry. A pre-print is also available on arXiv.[275] It shows that there is a strong desire from within the community to establish a set of standard procedures which would allow for the proper testing and comparison of existing or entirely new QD methods. This task is made exceedingly difficult by the fact that for sufficiently large molecular systems, no exact solution is available and hence many different approximations are employed, which are hard to compare directly. The final results are equally difficult to compare to one another since they are typically presented as a set of observables with a strong time and/or energy dependence, and hence are sensitive to the particular simulation conditions, which can be hard to reproduce across methods.

Apart from discussing these difficulties, some suggestions are also made about how to move forward. As a first step, 4 key phenomena out of the wide range of possible photophysical and photochemical processes were chosen to represent the most important applications of QD simulations in the molecular sciences. From there, the search for suitable benchmarking systems can be narrowed down to those exhibiting one or more of these phenomena, and an initial benchmarking set can be suggested. Also, the specific initial conditions, including the level of electronic structure theory, will need to be agreed upon, together with suitable metrics to evaluate the performance of a given simulation (this last point is especially relevant when no exact results are available). The hope is that this roadmap will allow the community to come together and work towards developing best practices and raising awareness of the challenges ahead, which would be a big step towards the maturity of the field as a whole.

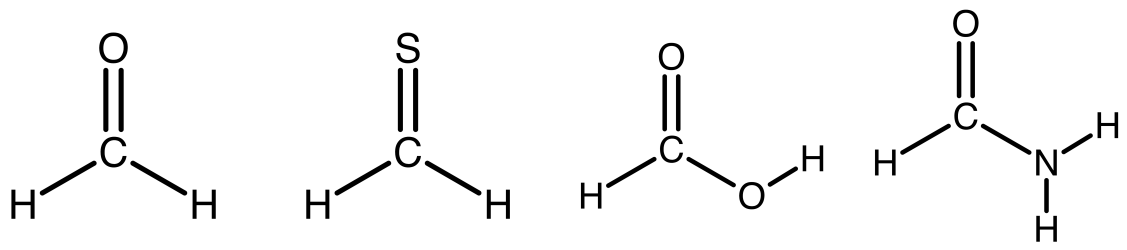
In the past, attempts have been made at providing the community with benchmark systems for which exact results are available. These include the well-known Tully models[143] which describe three idealised scenarios that allow one to evaluate the performance of a given QD method, but in a very simplified way. Other low-dimensional model systems have been used for this goal but there are doubts about how to extend the conclusions from these benchmarks to larger, more complex systems. With this in mind, the molecular Tully models were put forward as a more realistic set of benchmarks.[276] Even though these were successfully used to evaluate and compare some methods, they lack the generality required for widely adoptable benchmark systems. Hence the need for a discussion, hopefully leading to a consensus, about numerous open questions such as what reference to use when comparing

results (accurate simulations or experiment?), which observables to use for comparison, and how to ensure a consistent definition of a system across methods based on fundamentally different theoretical frameworks (e.g. wavefunctions vs trajectories)?

Within that context, this chapter will examine the various steps required to successfully run a DD-vMCG simulation. The questions raised in the discussion related to benchmarking will be considered at each stage, and discussed in the context of the method. Some considerations will be specific to DD-vMCG, while others may apply more generally to many (or all) QD methods. The aim is then to conclude whether or not a given step can be simplified, automated, or perhaps skipped entirely. Performing such an analysis should help streamline the procedures and thus make the method more accessible, both conceptually, but also with respect to practical details of the implementation (i.e. in QUANTICS, in this case). The overarching goal here is to make complicated computational methods more accessible to non-expert users by removing barriers and creating clear and sensible workflows. Such a project should also contribute to the benchmarking initiative by allowing a wider participation and, additionally, the introduction of automation in workflows can ensure consistency among different research groups by making sure a particular aspect of a QD study is performed in exactly the same way. Additionally, automation can also remove the need to perform some tedious and laborious tasks and hence speed up the entire process. Of course, it is important to thoroughly test any of these automatic procedures, but this is exactly the aim of the development of good benchmarks; to allow one to test and evaluate various aspects of one's methodology.

5.2 Proposed approach

In order to evaluate the steps behind a DD-vMCG calculation, a set of molecules was chosen which were simulated in exactly the same way, i.e. in terms of electronic structure method, initial conditions, propagation, and analysis. The molecules in question are all carbonyls, starting with formaldehyde (i.e. the simplest one, Figure 5.1a), and then adding more complexity by replacing the oxygen atom with sulphur (i.e. thioformaldehyde, Figure 5.1b), adding a hydroxyl group (i.e. formic acid, Figure 5.1c), and adding an amine group (i.e. formamide, Figure 5.1d). These four molecules will serve to exemplify and highlight various practical challenges associated with QD



(a) Formaldehyde (b) Thioformaldehyde (c) Formic Acid (d) Formamide

Figure 5.1: Set of carbonyl molecules chosen for DD-vMCG analysis and comparison.

calculations.

Admittedly, since they are very similar, the conclusions made at the end of this analysis may not hold generally. On the other hand, a clear increase in general molecular complexity is observed. Properly defining complexity in molecules is not trivial however, and remains an active area of research.[277] The first quantitative metric was proposed by Bertz based on topological arguments from graph theory, and defines molecular complexity in organic molecules as a combination of the degree of branching, and the number of heteroatoms.[278] This simple index is used here to get a sense of how quickly complexity grows for the four carbonyl molecules (Table 5.1). Notice,

Table 5.1: Bertz complexity index (C_T).

Molecule	C_T
Formaldehyde	2.0
Thioformaldehyde	2.0
Formic acid	10.3
Formamide	12.3

first of all, that the first two molecules are equivalent since no distinction is made between types of heteroatoms. This is clearly an oversimplification since the nature of a heteroatom can cause significant complications when it comes to electronic structure calculations, as well as during QD simulations. On the one hand, there is an increased computational cost associated with atoms containing relatively many electrons (as is the case for sulphur), and on the other hand, heteroatoms can cause changes in the energies of excited states, often leading to many low-lying states of various spin multiplicities which complicates the accurate modelling of non-adiabatic phenomena.

Apart from this caveat, the quantities in Table 5.1 seem to reflect the increase in

complexity that is intuitively clear by looking at the respective molecular structures in Figure 5.1. Given this observation, it should be possible to generalise some conclusions in terms of complexity, rather than the specific identity of molecules. This shall be commented on later when the results of the QD simulations of the carbonyl molecules are discussed.

In the first part of the chapter, the steps required to complete a DD-vMCG study are described, followed by an analysis of their relative practical and/or conceptual difficulty. This section is followed by some suggestions/solutions related to the previously identified challenges. Where appropriate, the discussion in both of these sections is supplemented by relevant case studies which illustrate and clarify the ideas being put forward by using one of the four carbonyl molecules as a concrete example. Lastly, a more complete analysis of the results of each molecule is presented and analysed, using the improvements suggested earlier on in the chapter.

At this point it is important to mention that although all four of these species are relevant in either industrial,[279–281] atmospheric,[282,283] and/or interstellar[284] chemistry, this is not why they were chosen. Nor do the results presented in this chapter promise to answer or explain any outstanding questions related to the photochemistry/physics of any of these molecules. The focus here is primarily on the practical aspects of the method used to study the dynamics of molecules in general, and the results related to the chemistry of any of the chosen molecules should be seen as preliminary. In fact, a choice was made to not spend a lot of time investigating the existing literature on most of these molecules so as to not bias the interpretation of results (previous work was mostly used to check that the underlying electronic structure calculations were suitable, and predicted the correct states at approximately the right energy). The effect of this will also be discussed. All that is not to say that the simulations presented here could not be further developed to obtain accurate, quantitative data on any of the above species, but such an analysis is outside the scope of this chapter

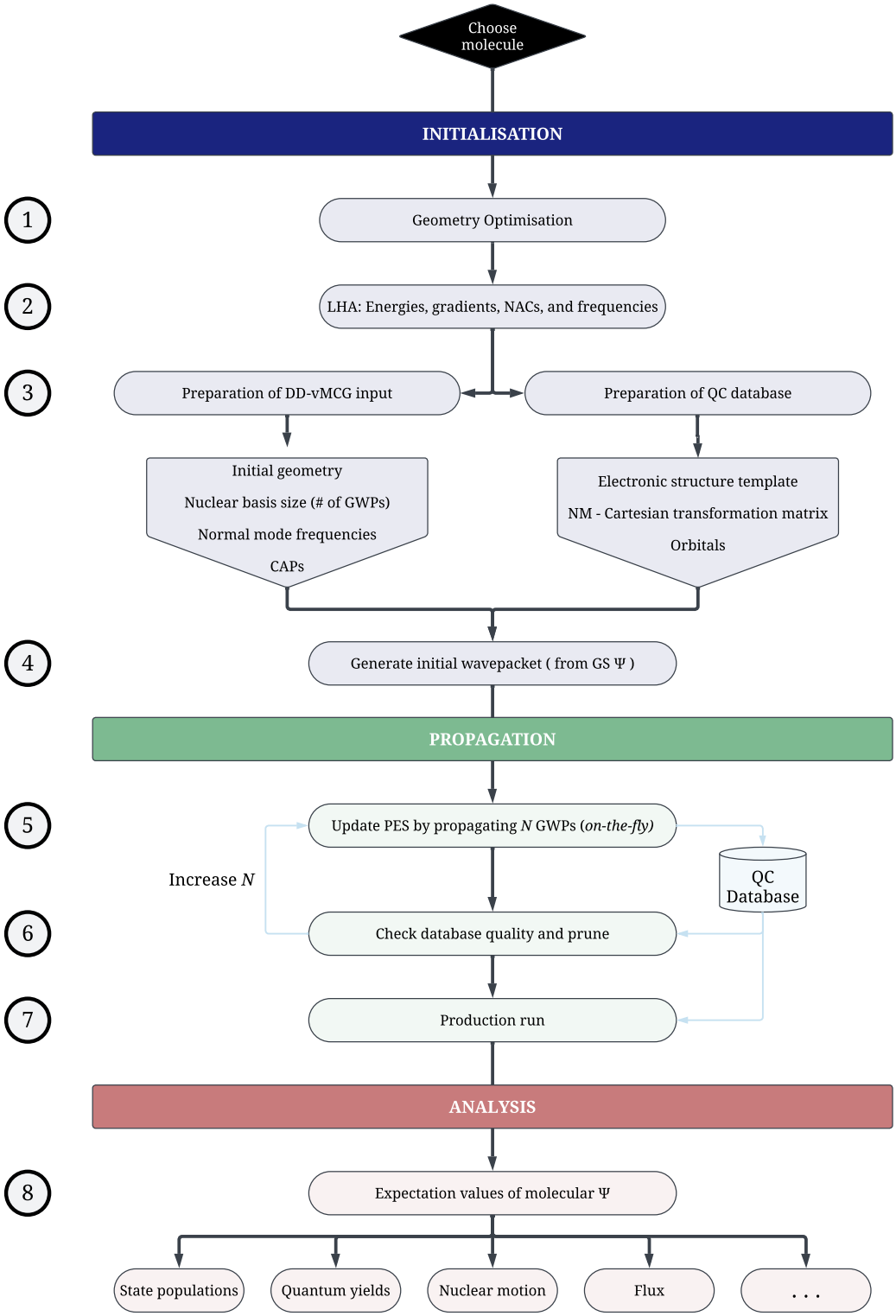
5.3 DD-vMCG workflow

In this section, the general workflow behind DD-vMCG is described. The individual steps required are outlined and labelled in Figure 5.2. This chart will serve as a guide throughout this section, which itself is split up in the 3 phases of the workflow:

initialisation, propagation, and analysis. In the first two phases, care was taken to treat the 4 molecules in the same way in an effort to limit any system specific effects. This generally applies unless stated otherwise. In the analysis phase, it is of course more appropriate to consider the results of each molecule individually, although some general considerations can still be made here as well. In addition to simply describing the specific steps and giving examples, this section also provides an evaluation of their ultimate difficulty. This evaluation sheds light on some of the major bottlenecks in the field, and serves as an introduction for the next section, which proposes solutions/improvements to a few problematic elements identified in what follows.

5.3.1 Choice of molecule

The very first step in any investigation is of course the choice of molecule to study. In theory, one could choose from the entire chemical space of molecules, estimated to contain over 10^{60} possible structures.[285] Two factors narrow down this vast space however. The first relates to the fact that, in general, researchers tend to study molecules that are relevant in our own bodies, our own atmosphere, on our own planet, or in our own galaxy. Hence, relevant research avenues ultimately dictate what kinds of molecules will be of interest. The second factor is the fact that in computational quantum chemistry, one is often severely limited by molecular size. Although this does massively reduce the number of systems that one could possibly study, it is a big limitation as well. A common way of dealing with this issue is to use model systems that recreate the essential features of a target molecule. The term "model" can have two different meanings in this context. The first refers to a system which does not resemble a molecule at all, in the sense that no atoms are defined, but simply consists of a number of pre-defined parameters which govern the motion of one or more basis functions which constitute the "particles". Such models are often called toy-models and are typically used to test a specific aspect of a computational method. A good example are the Tully models,[143] consisting of 2 potential energy surfaces defined by one-dimensional analytical functions, with another function describing the strength of the coupling between the two surfaces. Such a model, despite being somewhat arbitrary, can capture the necessary physics to test how a given QD method handles specific situations. However, the conclusions cannot necessarily be extrapolated to more realistic systems.



NAC = Non-adiabatic Coupling, LHA = Local Harmonic Approximation, GWP = Gaussian Wave Packet, CAP = Complex Absorbing Potential, QC = Quantum Chemistry, NM = Normal Mode, GS = Ground State, GGP = Gross Gaussian Population, PES = Potential Energy Surface

Figure 5.2: Workflow of a DD-vMCG simulation.

The second kind of models are those which do consist of the description of an actual molecule, but include some restrictions or approximations. For example, bulky groups can be replaced by smaller ones provided they do not actively participate in the reactivity of a molecule. A common practice in computational photochemistry is the study of an isolated chromophore as a means to understand the mechanism of a much larger molecular system, containing that chromophore (examples include the study of retinal by means of small protonated Schiff bases,[286] or modelling the large GFP protein by considering only the active HBDI chromophore, and a few key surrounding residues[287]). Alternatively, it may also be desirable to freeze specific DoFs in a large molecule to reduce the dimensionality of the calculation. Notice that these types of approximate models all rely on an intuition and/or knowledge about the chemistry of the system in question.[169] Obviously, such knowledge can be wrong or simply absent. In those cases, *on-the-fly* dynamics methods provide a potential exploratory tool, allowing a sampling of the vast phase space. This idea is further explored in section 5.4.3. The molecules considered in this chapter however, are small enough to be modelled in their full dimensionality.

5.3.2 Initialisation

The first important part of any QD study, or indeed any computational study, is to properly define the system of interest. In this case, this refers of course to the chosen molecule. In step **1** (see Figure 5.2), the starting molecular structure is generated by a simple optimisation which finds, hopefully, the ground state global minimum energy geometry, taken to be the Franck-Condon (FC) point. The second step consists of calculating excited state energies, and their properties (frequencies/Hessians, gradients, and all inter-state coupling vectors) at this point. Having obtained the Hessian, note that it is now also possible to go from the molecular Cartesian coordinates, to normal mode coordinates. All this information is required to parametrise the initial PES using the LHA.

Although these first two steps seem trivial, only requiring some knowledge of electronic structure methods and codes, it is here that one has to decide what calculations will be run during the *on-the-fly* evaluation of the potential. This is arguably the most important stage of the entire procedure since the quality of the potential can determine the ultimate quality of the resulting dynamics.[288,289] So, although automation of the

physical steps required to set up and run the calculations in steps **1** and **2** is trivial, the method of choice is an extremely important decision for the user. How to make this choice is not the topic of the current discussion, but it is worth mentioning some recent advances in the direction of facilitating this task. Firstly, the automatic generation of molecular active spaces for use in methods like CASSCF and its variants have been proposed, and have the potential of reducing the practical challenges associated with such methods.[290,291] Another exciting development that was recently put forward is the use of an automatic benchmarking program which identifies optimal methods for excited state calculations, taking into account cost vs accuracy.[292]

For the present work, rather than spending a lot of effort benchmarking specific electronic structure calculations for the 4 molecules of interest, they were all described at the same level of theory: State-Averaged Extended Dynamically Weighted CASPT2 (XDW-CASPT2).[206] The only exception is thioformaldehyde, for which multi-state CASPT2 was used, following previous work.[289] The active space of each molecule was also kept as similar as possible, making sure to include the π and π^* orbitals of the carbonyl group (which acts as the active chromophore in all systems), as well as carbonyl oxygen/sulphur lone pairs and the main σ and σ^* orbitals to allow the description of dissociating bonds. Having identified the lowest lying excited states for each molecule, the results were compared with some previous studies in the literature to make sure that the important states were at least present and roughly in the correct place. The results of the CASPT2 calculations are collected in Figure 5.3, showing the low-lying singlet states for all molecules, as well as the two lowest triplet states of formaldehyde and thioformaldehyde, which may or may not be of interest (discussed later).

To recap, the initialisation of a DD-vMCG simulation consists first of the calculation of key properties at the FC point in order to set up the LHA, represented symbolically as follows: $\mathbf{x} \rightarrow \{E_i, F_i, H_i, \Lambda_i\}_{i=1}^{S}$ where $\mathbf{x} = \{x_i, y_i, z_i\}_{i=1}^{N}$ are the Cartesian coordinates of N atoms, and $\{E_i, F_i, H_i, \Lambda_i\}_{i=1}^{S}$ is the set of energies E, gradients F, Hessians H, and couplings Λ, for a total of S electronic states. This information is used in step **4** to generate the initial wavepacket, $\Psi_0(\mathbf{x}, t_0)$. A big advantage of the DD-vMCG method, or any other method which propagates a fully coupled, "quantum" wavefunction, is that this step can be made trivial by employing the vertical (sudden) approximation, described in chapter 2, section 2.3.4. An exact wavepacket is generated

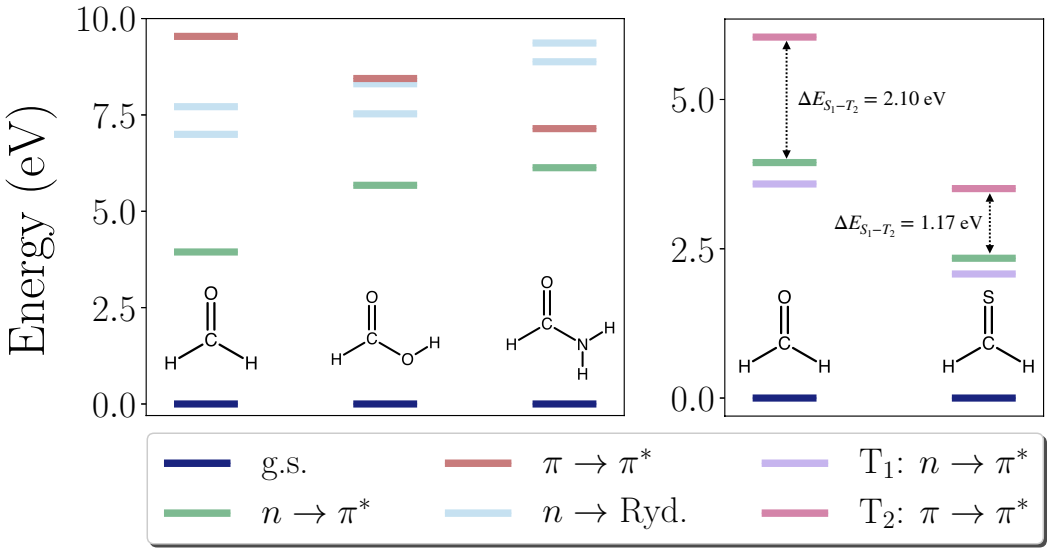


Figure 5.3: *Left:* Singlet electronic excited state energies and characters of formaldehyde, formic acid, and formamide. *Right:* Singlet and triplet states of formaldehyde and thioformaldehyde. Energies computed with XDW-CASPT2//aug-cc-pVDZ (for thioformaldehyde, CASPT2 was used), implemented in OpenMolcas.[192] "g.s." = ground state, "Ryd." = Rydberg orbital.

by projecting the vibrational ground state wavefunction on any desired excited state, by setting the transition dipole to unity. The specific parameters defining the excited state wavepacket can be adjusted, but this is the default approach in DD-vMCG. Next, the details of the calculation itself need to be defined using suitable parameters (timescale, initial state, integrator etc.), and an initial database, DB_0, created. Having done that, QUANTICS will automatically generate an initial wavepacket, which is equivalent to carrying out the following transformation: $\mathbf{x}, \{E, F, H, \Lambda\} \rightarrow \Psi_0$.

Automating or simplifying these two steps is, at least in principle, straightforward. For example, the use of a Graphical User Interface (GUI) can reduce the barrier of learning how to use an unfamiliar computer package. If implemented well, the only decisions that need to be made to initialise a DD-vMCG calculation are the choice of molecule, and the level of electronic structure theory. The addition of a GUI to the QUANTICS package is currently a work in progress.

5.3.3 Propagation

Having chosen a suitable electronic structure method, generated the initial wavepacket, and chosen simulation parameters, a propagation can be launched. Remembering that the PES is initially given by a harmonic approximation of all the excited state surfaces,

the direct dynamics algorithm will propagate GWPs until the calculation of a point on the potential is requested and carried out. In this way, the PES is gradually built up. It is generally advisable to start the initial propagation with few basis functions and for a relatively short time. This allows one to more easily inspect the results and determine whether or not everything is working properly. One should inspect the calculated points and evaluate the potential that was produced, as well as confirming the energy and norm conservation during the simulation.[†] Only after being satisfied with these tests should one continue to run more simulations with a larger basis size and a longer total simulation time. This loop of building up the potential over multiple separate simulations is represented in Figure 5.2 between steps **5** and **6**. For an example of the application of this process, see Case I (section 5.5.1) which illustrates the build-up of the formic acid database.

Aside from making it easier to spot potential errors, the gradual build-up of a database using sequential calculations benefits from improved initial conditions given by the previously generated surfaces. As such, each additional calculation will more optimally sample relevant geometries because the propagation is started on increasingly accurate surfaces (in contrast, the propagation of a direct dynamics calculation on a database containing only the first point starts on harmonic surfaces, which may inadequately capture the initial motion of some wavepackets). Such a procedure is reminiscent of the Grow method of Collins and coworkers, in which multi-dimensional PESs are "grown" by performing multiple classical[293,294] or quantum[295] dynamics simulations in order to automatically sample relevant geometries.

It is also worth pointing out that one can perform simulations that start on different electronic states, and then merge the resulting databases together at the end. This is useful if excitations to multiple states are being considered (e.g. when studying the wavelength dependence of product ratios). Since each simulation can be run simultaneously the relevant space can be sampled in an effectively parallel way.

A key question now arises: can the evaluation of the quality of database records be assessed in a simple and/or automatic way? Firstly, it is stressed that the issue being discussed here is not related to the *procedure* used to generate new points in the database. It is assumed that the geometries being sampled are of good quality. Rather,

[†]A notable exception occurs when using CAPs, in which case neither the norm nor the energy is conserved.

the question is about determining whether or not the QC points in a given database represent an accurate potential, not only in terms of the energy but considering the gradients and couplings as well. Assuming that the chosen electronic structure method produces reasonable data for most points in the configuration space, this exercise boils down to being able to "prune" a database, meaning to selectively remove failed or unphysical points which do not contribute to smooth PESs. Such failures can be due to a number of reasons and can occur, for example, at geometries corresponding to a CoIn where electronic structure methods such as TD-DFT and even CASPT2 are known to struggle.[40,86,288,296–298]

Considering first the evaluation of potential energies, one could try and define a reasonable energy range for a particular system. By setting an upper bound, QC calculations that fail massively and produce unphysically high energies can be prevented from contributing to the potential. Similarly, negative energies (relative to the ground state equilibrium) should be removed as well. This same idea can be applied to the other quantities, provided one can decide on suitable threshold values, which is not trivial for gradients and couplings (see section 5.4.1 for suggestions). In any case, such a procedure would only guarantee to pick up on clear outliers in the dataset, while any subtle discrepancies will be missed.

Making matters worse, the database points need to be diabatised to obtain the required diabatic potentials from the adiabatic QC data. Currently, the propagation diabatisation technique is used (chapter 2, section 2.3.7), which may not always produce exactly the desired outcome. However, there is a trade-off here between accuracy and ease-of-use. The fact the this procedure automatically diabatises an arbitrary number of states, and performs well enough that often only minor adjustments can lead to more than sensible surfaces, means it is a suitable solution to a problem that would otherwise be a major bottleneck of the method.

In summary then, even though some standard checks exist, it has to be concluded that determining the suitability of a given database for use in a direct dynamics calculation should be carried out by a knowledgable user who can judge the quality of a surface based on intuition and preferably some benchmark calculations (e.g. bond scans) or literature references. Such a conclusion is not very satisfying and points again to the importance of choosing a good electronic structure method to begin with. It is worth pointing out that virtually all other implementations of *on-the-fly* dynam-

ics methods do not store any of the calculated points. This means that it is not straightforward to inspect and evaluate the PESs on which the dynamics proceed, and is therefore rarely done. Given that it is well-known that electronic structure methods may fail at certain points, leading to discontinuous surfaces, this could be seen as willful, and certainly blissful, ignorance.

Moving on now, once a suitable database is obtained that passes all the checks, a final "production run" is carried out (step **7**) in which a database is read, but no new points are added to it. The input parameters are exactly the same as during the preceding calculations and the number of GWPs should now be at its maximum. One may ask: what is the optimal number of basis functions to use for the production run? Unfortunately, finding a suitable number is not dictated by fixed rules since the level of convergence of a DD-vMCG calculation cannot straightforwardly be determined. Rather, it may be very system-dependent since the complexity of the dynamical processes occurring in a given molecule will determine how flexible the wavefunction needs to be, and hence how large the basis set should be. There are some rules of thumb however, which can provide some guidance. Consider for example the following: $N_{GWP} = n \cdot 2d + 1$, where d is the dimensionality in terms of normal modes $(3N - 6)$, and n is any integer. The rationale here is that by having multiples of twice the number of dimensions, there will in theory be at least two wavepackets for every normal mode coordinate, one in each direction. Adding one more basis function ensures that there can remain at least one wavepacket in the centre. This approach has also been employed in quantum-Ehrenfest studies using the same Gaussian basis functions.[299]

Even though this is a good starting point, in no way does it ensure a large enough basis. If one really wants to push a simulation to convergence, a reference result obtained from a method with well-defined convergence criteria is desirable (e.g. MCTDH). When such a reference is not available, as is usually the case, one could look at the difference in the change of a certain quantity with respect to the number of basis functions. An example of using the diabatic state populations as a measure of convergence is illustrated in the bottom part of Figure 5.8 of Case I. Other than convergence, some other checks can be performed to evaluate the quality of a simulation. A first example of a simple test is the conservation of energy and norm. These should both be conserved in a closed system and hence can be an indication of errors

if this is not the case. However, as stated earlier, introducing CAPs in a simulation has the effect of making the Hamiltonian matrix non-Hermitian and thus the energy and norm are no longer guaranteed to be conserved.

As a final note, it is clarified that determining the ultimate accuracy of a computational result with respect to physical reality is an enormous challenge and is not the topic of this chapter. Instead, the notion of "quality" refers here simply to the extent to which a computational procedure produces physically plausible results based on the input data.

5.3.4 Analysis

Once a production run has successfully finished (meaning without producing errors during the propagation) and passed the basic tests mentioned above, the output needs to be analysed to extract useful information about the system. On the one hand, appropriate analyses will be system-specific and based on the particular phenomena being studied. However, there are some common data points that are important and useful in many applications of QD simulations, especially in the context of photochemistry/photophysics. These include electronic state populations, quantum yields or reaction rates for example. For DD-vMCG and any other methods which produce the full molecular wavefunction, extracting this information is not particularly difficult in theory. All that is required is applying a suitable operator and computing the expectation value. The QUANTICS program in particular contains a library of analysis programs that perform the evaluation of a wide range of expectation values (see chapter 3, section 3.2.3 for the theoretical basis of some common analysis techniques). Their interpretation however, is more cumbersome and requires some level of experience and intuition in order to draw sensible conclusions.

Again, the use of a GUI could be useful here. If designed well, data can be extracted automatically and visualised in ways which makes interpretation much easier. Alternatively, the use of workflow managers can be useful when carrying out more advanced analysis. As an example, Barbatti and co-workers have recently published a package to aid with the data analysis of simulations obtained using Newton-X[300] (a non-adiabatic dynamics code used mainly for TSH simulations).[301] It consists of a collection of tools used to extract useful information from trajectories, such as critical geometries, key active coordinates and groups of similar trajectories. Some of these

tools resemble the ones that have been applied here (described in section 5.4), which confirms that such developments are generally valuable within the field.

Before describing some proposed improvements related to the final analysis of DD-vMCG calculations, a few general considerations are given with respect to the interpretation of commonly used quantities: diabatic state populations (and associated nuclear motion), product ratios, and reactive flux.

State population

Having access to the full wavefunction allows for a straightforward calculation of the individual diabatic state populations during a DD-vMCG simulation (see chapter 3, section 3.2.3). Analysing the change in population of various states during a simulation can reveal detailed mechanistic information about the role played by those states, and can additionally help to determine the level of convergence of a calculation, as described earlier. Using the state populations for the latter application is simply a matter of comparing how the population of a particular state changes in different simulations and assessing when the difference is small enough to consider the calculation converged. Extracting mechanistic information however, is much more difficult. With experience, one learns to spot some common patterns and connect those to a physical interpretation, but this typically requires system-specific knowledge, a good conceptual understanding of excited state population transfer processes, and an idea of the shape of the excited state PESs. As a result, it is not obvious how one would generalise this analysis.

A first useful step might be to break the problem down into different classes of phenomena. For example, one is often interested in knowing where the density ultimately accumulates, after the initial ultrafast relaxation processes have taken place. Identifying this state is of course not difficult since it simply requires one to identify the state with the highest population at the final simulation time. However, it would be a shame to stop there since much more information is available about the dynamics behind the final result. For instance, it should be possible to identify what nuclear motion facilitates the population transfer to a particular state. Such knowledge is clearly valuable in the context of rational design for functional chromophores and other optoelectronic applications where specific transfer dynamics are required.

Alternatively, it may instead be interesting to know which state(s) cause a chemical

change to occur, and on what timescale. Think of unbound states for example, leading to dissociation. By tracking both the total norm of the wavefunction and the individual state populations, this information becomes available, but not necessarily in a manner that is easy to interpret (see for example chapter 3, section 3.3.2 where the complex photodissociation mechanisms in methanol are analysed).

To provide an example of the first kind of system, in which the final state populations are of interest, the excited state dynamics of thioformaldehyde are analysed in Case II (section 5.5.2). From this example, it is clear that simply looking at the change of diabatic state population cannot reveal the underlying factors causing the observed dynamics. The shape of the PES and the motion of important nuclear DoFs need to be considered to arrive at valid conclusions. With respect to this point, it is important to recognise that the challenge lies not in obtaining the relevant information, but rather knowing what to look for. It is reasonable to suggest that this problem, therefore, is highly system-specific since understanding electronic transfer dynamics depends on the complex coupling between electronic and nuclear motion, which can vary greatly from one molecule to another. Nevertheless, some general trends are sometimes observed when comparing different molecules, which will be demonstrated in section 5.6.1 when the state populations of the other three carbonyl molecules are presented, though in less detail.

Product ratios

Determining the nature of different possible photoproducts, and their relative yields, is one of the most desirable results when investigating photochemical reactions leading to large-amplitude nuclear motion, such as dissociation, isomerisation or excited state proton transfer (all key phenomena identified in the benchmarking roadmap[275]). Naturally, the first step towards this goal requires an analysis of the changes in molecular geometry during the dynamics. This is most easily done by extracting the nuclear trajectories of individual basis functions and visualising the change in geometry over time, assigning every individual trajectory to a certain product. Then, to extract a quantitative ratio, each trajectory corresponding to a certain product is weighted by its GGP, all weights are summed up, and finally divided by the total. This approach is very straightforward and not particularly difficult if one has some experience looking at molecular models. However, as computational resources and algorithmic efficiency

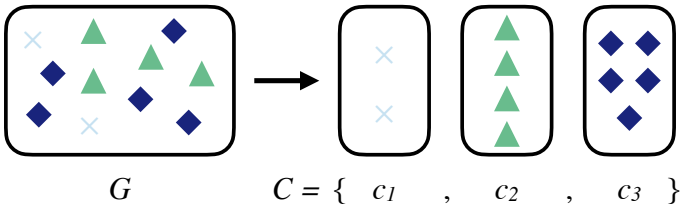


Figure 5.4: Graphical representation of the separation of 3 clusters of similarity in a dataset.

grows, QD calculations are able to tackle larger molecules and employ larger basis sets. At a certain point, both these factors will make the manual inspection of trajectories a very time consuming process, and not a particularly interesting or stimulating one.

Thinking about this problem, one soon realises that the analysis of molecular trajectories is not unique to the field of QD, far from it. Classical molecular dynamics, which has been around for much longer, is also faced with the challenge of interpreting changes in configurations. In fact, it is arguably more challenging in that field since both the size and the number of configurations are typically orders of magnitude larger than molecular trajectories resulting from QD simulations. Large biomolecules for example, don't lend themselves particularly well to visual inspection, and therefore several automated workflows have been developed that employ statistical techniques to aid the identification of, for example, metastable states in proteins[302] or the conformational dynamics of multi-body systems.[303] The common feature here is the identification of *similar* structures in a dataset, and understanding the key factors driving the conformational changes. This problem can be represented as a mapping of $G \rightarrow C$, where G is the set of all trajectories/configurations, and C represents the set of clusters of similarity extracted from G (graphically shown in Figure 5.4). In the particular case of determining branching ratios, one aims to achieve $G \rightarrow \Phi$, where Φ contains the quantitative yields of all products. In section 5.4.2, a procedure is outlined which performs this mapping, partially inspired by work done in classical molecular dynamics communities.

Flux

The last quantity that will be discussed here is the reactive flux, as described in chapter 3, section 3.2.3. It can be used as a way to extract information about the electronic populations and product ratios, simultaneously, by evaluating the magnitude of the density of a specific state going into a product chanel. In fact, analysis of the flux would

be the traditional way of determining yields of reactive scattering simulations, which lie at the origin of the field of QD. In those low-dimensional systems (often only 3 particles, e.g. H + H_2[304–306]), this quantity is definitely the best diagnostic tool to understand the finer details of the reaction. However, one could argue that as the dimensionality of a system grows, calculating and interpreting the flux becomes more challenging. The problem is a practical one in the sense that the location of a flux operator needs to be defined in order to obtain its expectation value at that point. Choosing a suitable position, i.e. one that will not capture any oscillatory vibrational motion, can be tedious and is only realistically achievable with some *a priori* knowledge of possible reaction channels, or by analysing the results of a prior simulation. Additionally, the flux can only be calculated accurately for grid-based methods. For these reasons, it was decided not to focus on it in any automated procedure. However, in the last part of this chapter (section 5.6), where the simulations of the carbonyl molecules are discussed and a suggested approach for automatically determining product ratios is applied, it will be shown that many of the the same insights are obtained from analysing the flux.

5.4 Solutions

5.4.1 Pruning a QC database

Performing sequential direct dynamics calculations in order to gradually build up a database of QC points has the advantage that one can take the opportunity to check the quality of the records being stored. This is important since it is reasonable to assume that electronic structure calculations will produce spurious results, at some point. However, manually checking if the QC data is sensible or not is only possible for small databases, and becomes increasingly unfeasible as the number of records rises. When looking for failures in larger databases, the PESs that are produced can help determine whether or not there are any major issues. However, if unphysical features do appear (e.g. kinks, oscillations, negative energies etc.), it is usually not obvious what actions to take to remedy this.

One solution is of course to start again using a different electronic structure method altogether, but it is assumed here that the QC calculations are valid in the majority of the configuration space explored, and that any chosen method is likely to break down

in some regions. Hence it would not be productive to try a different method since there is always a possibility of a few problematic points affecting the quality of the PES. It is also important to remember that it is not only the potential energy that has to be calculated properly, but the gradients and couplings as well, which are typically more prone to errors. A more efficient alternative would therefore be to simply remove problematic records from the database, and let the Shepard interpolation fill in the gaps. In the context of populating a database by running sequential QD calculations, this proposed pruning would be done at the end of each individual simulation, until the production run is performed using the final database.

Given that assessing the validity of a database based solely on chemical intuition, or a similar manual inspection, is not feasible when typical databases contain of the order of 10^3 - 10^4 number of records, it is clear that more sophisticated methods are needed. The idea explored here is based on being able to detect outliers and provide reasonable threshold values for each property, based on the available data in a given database. Statistical techniques should be ideally suited for such a task. If a probability distribution function can be fitted to all energies, gradients, and couplings in a given database, a cut-off value can be defined based on statistical measures (e.g. variance), and any records not falling within this cut-off can then be deleted. Note that the Hessian matrices are not considered here since these are usually obtained by a Hessian updating scheme, rather than being computed explicitly.

To determine what type of distribution is most appropriate, the formaldehyde database is used as a test case. In Figure 5.5, the distribution of the energies, gradients, and couplings are plotted, along with three different probability distribution functions; normal distribution, Lorentz (or Cauchy) curve, and Maxwell-Boltzmann distribution, defined as follows

$$\begin{aligned}
\text{Normal}: \quad & f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad , \{x \in \mathbb{R}\} \\
\text{Lorentz}: \quad & f(x) = \frac{1}{\pi(1+x^2)} \quad , \{x \in \mathbb{R}\} \\
\text{Maxwell}: \quad & f(x) = \sqrt{2\pi} x^2 e^{-x^2/2} \quad , \{x \in \mathbb{R} : x \geq 0\} \quad .
\end{aligned}$$

Before commenting on which distribution is most appropriate for which quantity, it is remarked that what matters here is that the centre (mean) and the spread (variance) of the data are well described, not necessarily the shape of the distribution. In that

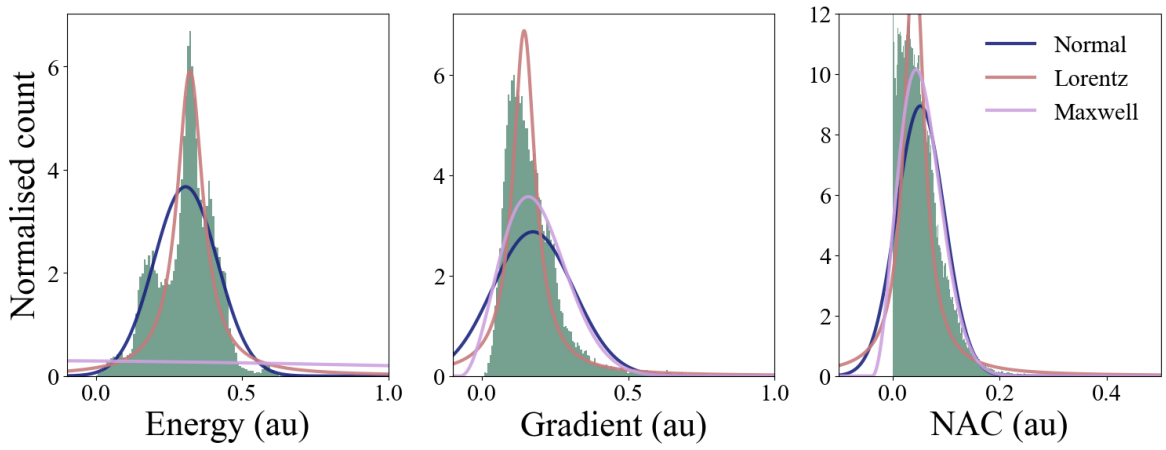


Figure 5.5: Normal, Lorenz, and Maxwell probability distributions fitted to (from left to right) energies, gradients and couplings (all in atomic units). All data produced by DD-vMCG calculations of formaldehyde, with $N_{rec} = 4420$.

respect, the normal distribution is the best choice to describe the energies, despite the fact that the actual distribution appears to be bimodal. However, this is not the case for gradients and couplings. Note first that what is plotted are the magnitudes of the vectors, which is why all entries are positive (the energy values are also expected to be strictly positive, but failure of electronic structure calculations may lead to negative energies). This allows for the use of a Maxwell-Boltzmann distribution, the domain of which is restricted to all positive real numbers, including 0. Note that for this reason, it cannot be applied to the energies, which may be negative, leading to the flat distribution in the first panel of Figure 5.5. Such a function seems to best describe the asymmetric distribution of both the gradients and couplings. It is of course possible that these conclusions only hold in the case of formaldehyde, but the same analysis was repeated on the databases of the other three molecules and similar trends are observed (see Appendix D, section D.2).

Moving forward with a normal distribution for energies and a Maxwell-Boltzmann distribution for gradients and couplings, a given database can be pruned by removing points that fall outside a certain range. Specifically, the range will be given by $[0\ ,\ \mu + n \cdot \mathrm{Var}(x)]$, where μ is the mean, and $\mathrm{Var}(x)$ is the variance of the distribution. Note again that this range is strictly positive since all energies should be positive (relative to the FC point) and the magnitudes of the gradient and coupling vectors are positive by definition. The integer n is used to adjust the strictness of the pruning. Since the number of outliers is expected to be very small compared to the total size of the

database, it is advisable to use a relatively large value of n. Ultimately though, the most suitable value will vary across databases and depends entirely on the how many electronic structure calculations ultimately fail. By way of example, appropriate values for n are determined for formaldehyde in Case III (section 5.5.2).

At this point, the main drawback of this approach is that the value of n needs to be determined based on inspection of the data, rather than being defined automatically. However, it is possible that with enough testing some good default values are determined and this pruning procedure can then be applied after every direct dynamics simulation, before starting the next one. Going a step further, one might even imagine evaluating the suitability of a calculated point before it is ever stored in the database. The ability to perform such a check during a propagation would be a valuable feature of DD-vMCG, and avenues in that general direction are currently being explored.

5.4.2 Calculation of product ratios

The task of identifying which trajectories lead to which products amounts essentially to pattern recognition, something humans are naturally good at, but which has historically been much more challenging for computers to perform, especially when dealing with high dimensional data. Fortunately, as data science has become increasingly important, many methods have been developed (and are readily available) that take a large multi-dimensional dataset and reduce its dimensionality by identifying key features responsible for different patterns. Furthermore, a different class of algorithms can be used to group together patterns based on some underlying similarity. These are termed clustering algorithms.

It is a rather obvious step to try and apply both of these methods to design an automated workflow which identifies photoproducts, and quantifies their relative yields. The multi-dimensional input data then consists of different molecular geometries. These geometries can be obtained from a QD trajectory, and the characteristic coordinate(s), or feature(s), of a particular trajectory could then be identified by dimensionality reduction. However, one can go one step further and apply a dimensionality reduction technique to the set of all trajectories, G. The aim is then to construct a low-dimensional map of points, each representing an individual trajectory. The distribution of these points reveals information about, for example, the similarity between two trajectories. By separating them into clusters of similarity, it should then

be possible to identify the different products formed during a dynamics simulation, and which specific trajectories lead to those products. If such a procedure can be performed automatically at the end of a QD calculation, the branching ratios of the different products could easily be obtained.

Variations of this idea have been previously applied in the analysis of QD simulations.[307–312] Presented here is the specific implementation used, as well as an evaluation of, and justification for, the choices that were made with respect to the methods used. Different dimensionality reduction techniques are discussed first, followed by various clustering algorithms, and finally the choice of similarity metric.

Dimensionality reduction

Methods that identify key features in high-dimensional datasets are typically based on unsupervised machine learning algorithms, which basically solve an optimisation problem. Among these approaches, the most popular include Principal Component Analysis (PCA),[313] Diffusion Mapping (DMAP),[314,315] Isometric Feature Mapping (ISOMAP)[316,317] and Multi-Dimensional Scaling (MDS).[318] Although varying in their respective implementations, all these methods are largely based on the same assumption; that a low-dimensional manifold exists, embedded in the high-dimensional data. The methods then differ in the way this manifold is found. Generally speaking, the idea is to find, for a n-dimensional matrix, the m highest (most important) eigenvalues and discard the dimensions that contribute to the remaining, less important eigenvalues. Each point in the original data is then represented in this m-dimensional space.

In relation to their application to molecular dynamics, it is worth looking at some properties of the four methods mentioned here. These are summarised in Table 5.2. For a more detailed discussion of the similarities and differences between dimensionality reduction methods, see for example the review by van der Maaten *et al.*.[319] The first

Table 5.2: Comparison of dimensionality reduction methods.

Method	**Linearity**	**Distance preserved**
PCA	Linear	Not preserved
MDS	Both	Euclidean
DMAP	Non-linear	Diffusion
ISOMAP	Non-linear	Geodesic

important feature to consider is whether the method is linear or non-linear. When considering chemical dynamics, in which large-scale motion involving multiple coordinates is likely, the use of a linear method may not be suitable. Imagine a torsional motion around a bond for instance, which could either be viewed as a linear displacement of a normal mode which varies a dihedral angle, or equivalently, as a non-linear change in Cartesian coordinates. This coordinate dependence thus limits the use of linear methods such as PCA for the current application. Additionally, PCA is not guaranteed to preserve distance between data points, meaning that similar geometries may not be close together in their reduced dimensions. This is important for the current application since the aim is to identify structural similarities between trajectories.

With this in mind, the remaining three methods all seem suitable for the analysis of molecular geometries, and they have indeed been used for that purpose in the past. Nevertheless, each one does have its particular advantages and drawbacks. The fact that MDS is based on the Euclidean distance (between the elements of the lower-dimensional manifold) means that it may not perform well when the underlying manifold is curved. Additionally, when using the non-linear version of the algorithm, an intuitive interpretation of the reduced dimensions in terms of molecular motion is difficult. A similar issue might arise when using DMAP since the diffusion distance is also related to the Euclidean distance of the points in the embedded space. Furthermore, some extra work may need to be done to extract a meaningful physical coordinate from the diffusion distance. Another drawback of DMAP is that one needs to define several variables before running the algorithm, such as the specific diffusion kernel being used and the type of diffusion being modelled. These tend to be determined empirically. In that respect, ISOMAP offers a much more user-friendly "black-box" alternative, especially when remembering that the aim is to embed this technique in an automated workflow. Additionally, the fact that ISOMAP retains geodesic distance, which takes into account the curvature of the embedded space, results in a less coordinate dependent, and physically more intuitive interpretation of the reduced dimensionality coordinates. As a final comparative note, although DMAP is the most robust algorithm with respect to noise (or outliers), this will not be an issue since noisy data can be identified and removed by a clustering algorithm, discussed later on. Having made these considerations, the ISOMAP method was ultimately chosen to be the most suitable.

Clustering

Moving on now to the second step in the automatic detection of reaction products. Whereas dimensionality reduction serves to identify key (structural) features in a set of trajectories, a separate technique is needed to separate different clusters of similar trajectories leading to the same product. Again, one is faced with a number of different clustering algorithms, the performance of which depends on the type of dataset it is applied to. A table is made with some common clustering methods and some of their properties (Table 5.3), namely K-Means,[320,321] Spectral clustering,[322] Agglomerative clustering,[323,324] and (Hierarchical) Density-Based Spatial Clustering of Applications with Noise ((H)DBSCAN).[325,326]

Table 5.3: Comparison of clustering methods.

Method	**Type**	**Linearity**	**Robust to noise?**
K-Means	Centroid-based	Linear	✗
Spectral	Graph-based	Non-linear	✗
Agglomerative	Hierarchical	Non-linear	✗
(H)DBSCAN	Density-based	Non-linear	✓

Without going into too much detail, some key considerations are made for each method, starting with K-Means. This very simple and efficient algorithm is widely used exactly because of its simplicity and its wide availability. It assumes a dataset with globular clusters and partitions the data points based on the distance away from the calculated centre of the cluster. Although these features already present serious limitations, the biggest drawback is that clusters are assumed to have the same shape. Additionally, one has to define the number of expected clusters beforehand. Such a method is clearly not suitable for the current application.

Next, the Spectral Clustering method improves on some of the weaknesses of K-Means by constructing a graph to map the original space to a lower dimensional manifold, on which K-Means clustering is then performed. Thus, some of the weaknesses of K-Means still remain, such as having to guess the total number of clusters.

A different approach altogether is Agglomerative Clustering. As the name suggests, this method repeatedly merges small clusters together, based on certain criteria, until a single cluster remains. The result is a hierarchical structure where each level is composed of an increasing number of clusters, and one then has to decide on the desired

level of clustering. Although much more flexible than the two previous methods, agglomerative clustering essentially still requires one to guess the number of clusters in a dataset, and is hence not deemed to be suitable.

Finally, DBSCAN was chosen. This method assigns clusters based on the density of the data in a given region. This being a relative measure, a density parameter is introduced to provide a cut-off for defining new clusters (so-called EPS parameter). Regions of the data that are too sparse to fall within this clustering threshold are considered noise, and are labelled as such. This is a very desirable feature in the case of clustering molecular trajectories since it is possible that a given trajectory may fail, in the sense that it produces unphysical results due to a runaway wavepacket for example. In that case it is essential that it is identified as noise, so as not to influence the clustering of the remaining trajectories. Neither one of the previously discussed methods has this robustness to noisy data.

The main drawback of the DBSCAN algorithm is that all clusters are assumed to be of similar density, i.e. the average distance between the points in each cluster is comparable. The user-defined EPS parameter defines a suitable density for a given dataset and thus introduces an element of decision making which is typically undesirable in an automated workflow. In this respect, the hierarchical implementation of DBSCAN, HDBSCAN, potentially offers a solution by allowing clusters of varying density. In effect, it removes the need to choose a suitable distance parameter by determining the optimal number of clusters.

Ultimately though, it was decided to move forward with the original DBSCAN implementation since it offers a level of flexibility that can be useful when analysing the clusters produced by the algorithm. For example, one might find that multiple clusters are identified which actually correspond to the same product. In that case, it may be sensible to reduce the "strictness" of the clustering by increasing the distance parameter, which relaxes the criteria related to the similarity of different points/clusters, and might therefore merge clusters corresponding to a single product. A simple sensitivity analysis can help to determine the optimal distance parameter (see Appendix D, section D.3 for an example). It is added that DBSCAN has been used previously to cluster molecular trajectories, and found to be suitable for that application.[311,327–329]

As a final note on this section about clustering, a brief comment is made about the choice to use widely available algorithms, as opposed to developing a new ap-

proach that is ideally suited to the present use-case. The advantage of the latter is of course that some of the weaknesses associated with more traditional, general algorithms could be eliminated by focusing only on one specific application. On the other hand however, this may limit the availability of the workflow to the wider community of interested users. As an example of a newly developed clustering scheme which is particularly well-suited for the analysis of photochemical dynamics trajectories, see the recently developed algorithm by Kára *et al.*.[330] This methodology was designed specifically to group together trajectories from TSH simulations, and addresses some of the shortcomings of the methods listed in Table 5.3. In particular, its main advantage is arguably that it removes the need for tunable parameters (like EPS in DBSCAN). As such, it is ideal for incorporation into a "black-box" analysis workflow, perhaps more so than DBSCAN. However, its novelty does come at a cost, specifically the lack of rigorous testing on various datasets generated by QD simulations. Additionally, it is much more computationally demanding than, for instance, DBSCAN. In the end, the choice was made to not use this new algorithm in this work, but it may be given some consideration in the future.

Measuring similarity

Having decided on both a dimensionality reduction and a clustering method, a final consideration needs to be made about the type of input data. Of course the trajectories generated from a QD calculation are required, but since the key idea here is the concept of *similarity*, it is useful to look for a suitable metric to objectively quantify the resemblance of two trajectories. Looking at this from a purely mathematical standpoint, a procedure is needed which measures the similarity between two curves in arbitrary dimensions. Again, several such methods exist but of particular importance is a metric called the Fréchet distance. More specifically, the discrete variant of the exact formulation of this metric can be used to compute a distance between two arbitrary curves, approximated as polygons (i.e. composed of discrete points rather than a continuous curve).[331–333] Coming back to the specific use-case, this method is ideally suited to compute the distance between two trajectories, where the discrete points are given by the individual geometries, changing over time. An added benefit, although not relevant here, is that two trajectories need not be of the same length. In coupled wavepacket propagations this is never an issue, but when using independent

trajectory methods it is possible for different trajectories to finish at different times. Thus, for the sake of generality, it is considered an important and necessary feature.

The mathematical derivation of the approach is detailed in the literature and, reassuringly, is formally proven (for the exact case). It suffices to say here that the Fréchet distance is based on the similarity of *all* points along a given trajectory, compared to all points of another trajectory. This is in contrast with other methods which may only compare equivalent points (i.e. at one given time-step). This fact gives the method its characteristic flexibility with respect to the shapes of the curves being compared. It does mean however, that the computation of the Fréchet distance is inherently recursive and has a scaling of $\mathcal{O}(3^n)$ if one takes a naïve, brute-force approach (n = length of trajectory). Luckily however, a dynamic programming algorithm has been developed which brings the scaling down to $\mathcal{O}(n^2)$.[331]

The algorithm is shown in pseudo code in Figure 5.6, along with the subsequent steps using ISOMAP and DBSCAN, and ultimately the calculation of product branching ratios using the GGP. Notice that the computation of the Fréchet distance relies on the root-mean-square deviation (RMSD) between individual Cartesian geometries, collected in the distance matrix $\mathbf{D}$. One is free to choose any measure of (dis)similarity between geometries but RMSD seems the obvious choice, provided one removes rotational and translational contributions. Having calculated all the pairwise Fréchet distances between all trajectories, these are collected in a $N_{GWP} \times N_{GWP}$ symmetric matrix, $\mathbf{D}_{Fréchet}$, of the following form:

$$\mathbf{D}_{Fréchet} = \begin{bmatrix} 0 & F_{12} & \cdots & F_{1n} \\ F_{21} & 0 & \cdots & F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \cdots & 0 \end{bmatrix} \quad . \tag{5.1}$$

This distance matrix forms the input for ISOMAP, which assigns a 2-dimensional coordinate to each trajectory (note: a different number of dimensions can be chosen using the `n_components` variable). These reduced coordinates provide a way to visualise the distribution of trajectories in a 2-dimensional plot. Applying DBSCAN clustering to $\mathbf{D}_{Fréchet}$ finally reveals the different product channels and their corresponding trajectories. The `EPS` variable refers to the density parameter mentioned previously. Tuning this parameter can give rise to more or fewer clusters and/or noisy data points, and one should therefore always perform a sensitivity analysis by varying this parameter from

```
# Construct Fréchet distance matrix, DF, between ALL trajectories:

  1) Build pairwise distance matrix, D, between trajectories A and B:

  for i in len(A):
    for j in len(B):

      D_AB[i,j] = RMSD(A_i , B_j)

  2) Calculate Fréchet distance, F, between A and B, from D,
    via dynamic programming:

  F_AB = np.ones(len(A) , len(B)) * (-1)
  F_AB[0,0] = D_AB[0,0]

  for i in len(A):
    F_AB[i,0] = MAX(F_AB[i-1,0] , D_AB[i,0])

  for j in len(B):
    F_AB[0,j] = MAX(F_AB[0,j-1] , D_AB[0,j])

  for i in len(A):
    for j in len(B):

      F_AB[i,j] = MAX( MIN(F_AB[i-1,j] , F_AB[i,j-1] , F_AB[i-1,j-1]) , D_AB[i,j] )

  F_AB = F_AB[len(A) , len(B)]

for i in N_GWP:
  for j in N_GWP:

    D_Fréchet[i,j]= F_ij

# Map trajectories to Reduced Coordinates, RC, using ISOMAP:

RC = ISOMAP(n_components = 2).fit(D_Fréchet)
PLOT(RC[:,0],RC[:,1])

# Clustering using DBSCAN:

C = DBSCAN(EPS, min_samples = 2).fit(D_Fréchet)

# Compute ratio for each cluster, i, in C:

  for i in len(C):
    for each j in C[i]:

      cum_ggp[i] = cum_ggp[i] + GGP[j]

    Ratio[i] = cum_ggp[i] / GGP_total
```

Figure 5.6: Pseudo code for automated calculation of product branching ratios. `len(X)` returns the length of `X`, i.e. number of geometries if `X` is a trajectory, and number of clusters if `X` = `C`.

$0 \rightarrow 1$ in fixed intervals. To finally calculate quantitative ratios after having obtained suitable clusters, one simply sums up the GGPs of every trajectory in a given cluster to get the cumulative GGP (`cum_ggp` in Figure 5.6), and then dividing by the total GGP of all trajectories, excluding those labelled as noise. An illustrative example of this workflow applied to formic acid simulations is presented in Case IV (section 5.5.4).

To summarise without considering any of the details described previously, the proposed procedure boils down to a simple program that has as its only inputs $G = \{\mathbf{g}_i\}_{i=1}^{N_{GWP}}$, i.e. the set of all trajectories $\mathbf{g}$, and their corresponding weights $W = \{GGP_i\}_{i=1}^{N_{GWP}}$, given by the GGP of each trajectory. The program then carries out the following mapping: $G \rightarrow \Phi$, where $\Phi = \{\phi_i\}_{i=1}^{N_p}$, i.e. the set of all branching ratios ϕ, with N_p being the number of distinct products. Such a transformation is achieved by determining the set of clusters, $C = \{c_i\}_{i=1}^{N_C}$, with $N_C = N_p$, and, for each cluster c, computing the ratio

$$\phi^{(c)} = \frac{\sum_{i=1}^{N_{traj}^{(c)}} GGP_i}{\sum_{c=1}^{N_C} \sum_{i=1}^{N_{traj}^{(c)}} GGP_i} \quad , \tag{5.2}$$

with $N_{traj}^{(c)}$ being the number of trajectories in a particular cluster. Such a program can in principle be built in a way that does not require any additional user input. Of course, in practice, there are several checks that one should always perform, which cannot straightforwardly be automated. These include checking the mean geometries of the clusters to confirm that they do indeed correspond to distinct products. Due to symmetry, it is possible that two separate clusters actually correspond to the same molecular products. This could be fixed by employing a more sophisticated measure of similarity between structures/trajectories, but since this doesn't constitute a major weakness of the proposed approach, no steps have taken in that direction (yet). Next, one should always check the noisy trajectories as some of these may represent a reaction channel with very low probability, and hence only explored by one single wavepacket. Such a trajectory will not be assigned to a cluster since a minimum of 2 elements are needed.

Another factor that makes this methodology not entirely black-box is the `EPS` parameter in DBSCAN. This parametrisation can be avoided by employing the HDBSCAN algorithm, however, it was found that a lack of control over the clustering led

to cases where the set of trajectories was "over-clustered". An example of this is given in Appendix D, section D.3, where different clusters of thioformaldehyde are computed with HDBSCAN, corresponding to exactly the same product. Instead, it was way more effective to perform a simple sensitivity analysis with respect to the density parameter. Ambiguous cases are bound to come up, where a trajectory may belong in more than one cluster, and in that case it is desirable to have some control over how this decision is made. Also in Appendix D (section D.4) is a comparison of all dimensionality reduction and clustering algorithms discussed in this section, applied to the trajectories generated from the $\pi\pi^*$ excitation of formaldehyde.

Finally, the analysis outlined here was applied to trajectories obtained from methanol DD-vMCG simulations, presented in chapter 3. This was done to confirm that the same product ratios are obtained compared to the manual analysis of trajectories, which was indeed found to be the case (see Appendix D, section D.5).

5.4.3 Direct dynamics as an exploratory tool

This section addresses the potential lack of accuracy of DD-vMCG calculations caused by not reaching convergence. There is no point in trying to define absolute accuracy since it will heavily depend on what one is trying to achieve, which type of system is being studied, and what kind of method is being used. However, the possibility of getting fully converged, numerically exact results from a MCTDH calculation for example, is an impressive fact and it is worth thinking about how obtaining such results could be facilitated. The multi-layer implementation of the original algorithm addresses this issue by extending the reach of MCTDH to potentially very large systems (this is briefly mentioned in chapter 2), but the requirement of globally defined potentials in the correct sum-of-product form still remains a practical bottleneck. In that respect, direct dynamics calculations may turn out to be helpful.

The main problems with globally defined, analytical PESs can be broken down in two parts; the fact that much of the configuration space is not relevant during a given dynamics simulation, and the need to calculate and fit points in a multi-dimensional space. The first problem can essentially be solved by sampling the relevant space in advance of parametrising global functions. By realising that this sampling cannot be based solely on information calculated at the FC point, the use of DD-vMCG simulations could provide an ideal way to explore important regions of the

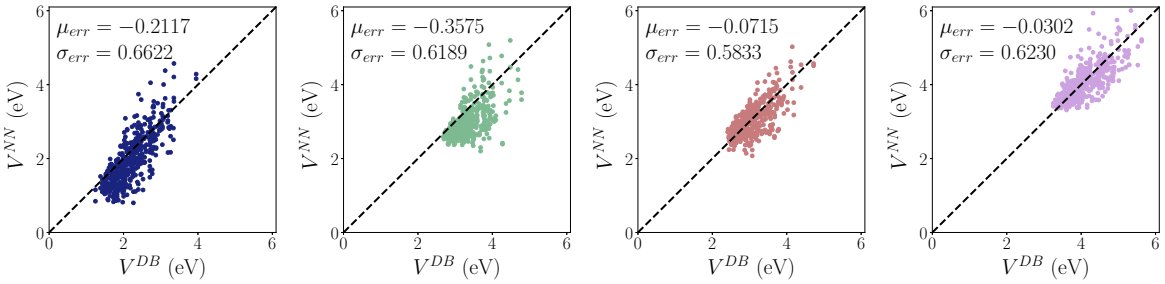
configuration space. Secondly, the QC data defining the PES in those regions is stored in a database, allowing it to be used as the starting point of a fitting procedure to obtain surfaces in a form suitable for MCTDH calculations. Like killing two birds with one stone, the use of direct dynamics calculations seems to be able to address the two issues related to one of the biggest drawbacks of MCTDH, or indeed any grid-based QD method with the same requirement for analytical surfaces.

Coming back to this second point, a procedure for fitting electronic structure data into analytical potentials is currently being implemented into QUANTICS by S. Patra at UCL. This approach is based on previous work which demonstrates how neural networks can be used to obtain PESs in the required sum-of-product form for use in MCTDH simulations (see Equation 2.57 in chapter 2).[334] To give a general outline, the procedure relies on a set of parameters being propagated forward through a network of layers, until the final layer produces the desired output from a set of inputs.[335] In this case, a 3-layer model is used with an input layer corresponding to a set of molecular geometries $\{\mathbf{x}\}$ taken from a QC database, one hidden layer, and one output layer which corresponds to the respective adiabatic potential energies $\{V^{NN}\}$ (the superscript "NN" refers to the fact that these energies are estimated through a neural net).

To achieve this mapping from geometry to energy, a linear combination between the elements of the network (referred to as nodes or neurons) is defined based on a set of weights $\{w\}$ between nodes of adjacent layers. Additionally, the nodes of the hidden and output layers have associated biases, $b^{(hid)}$ and $b^{(out)}$ respectively. These parameters are used to define a non-linear function:

$$y(\mathbf{x}) = b^{(out)} + \sum_{n=1}^{N} w_n^{(hid-out)} \sigma(z) \quad , \tag{5.3}$$

where N is the number of nodes in the hidden layer and $w^{(hid-out)}$ are the weights connecting the hidden nodes to the output layer. An activation function, $\sigma(z)$, is responsible for connecting the inputs with the nodes of the hidden layer. As such, it is a function of $\mathbf{x}$, as well as of the weights between input and hidden layers ($w^{(inp-hid)}$), and the hidden layer biases, $b^{(hid)}$. The choice of this activation function is crucial because it determines whether or not $y(\mathbf{x})$ has the required sum-of-product form. It has been shown that using an exponential expression, $\sigma(z) = e^z$, yields a suitable



(a) S_0 (b) S_1 (c) T_1 (d) T_2

Figure 5.7: Predicted vs actual adiabatic energies (in eV) of database records not included in the neural net training phase. The mean (μ) and standard deviation (σ) of the error is given in the top left corner.

function which can be written as follows to obtain the energy of an electronic state m:

$$\begin{aligned} y_m(\mathbf{x}) &= b^{(out)} + \sum_{n=1}^{N} \tilde{C}_n \prod_{i=1}^{D} \exp\left(w_{ni}^{(inp-hid)} x_i\right) \\ &= V_m^{NN}(\mathbf{x}) \quad , \end{aligned} \tag{5.4}$$

where D is the dimensionality of the nuclear coordinates. Note that $w^{(hid-out)}$ and $b^{(hid)}$ are both collected in the combined coefficient $\tilde{C}$.[336]

Having obtained such an expression, the parameters of the network (i.e. weights and biases) need to be trained on a given dataset to "learn" how the inputs map to the outputs. Once optimised, $y_m(\mathbf{x})$ can be used to predict the energy of any geometry that falls within the configuration space explored by the simulation that produced the database. In order to evaluate the performance of the model, the training is only done on a portion of the database, after which it is used to predict the energies of the remaining geometries not included in the training phase. In this way, a correlation plot can be made where the predicted energies, V^{NN}, are compared to the actual energies taken from the database, V^{DB}.

As an example, 75% of a database created during the simulation of thioformaldehyde is used to train a model like the one outlined above. The adiabatic energies of four electronic states were then predicted for the remaining 25% of the database records. This comparison of predicted vs actual energies is shown in Figure 5.7 for all four states individually. Without commenting too much on the actual quantitative accuracy of these preliminary results, the mean and standard deviation of the errors associated with the predicted energies are briefly analysed (given in the Figure). Firstly, the mean error is negative in all cases, indicating a systematic underestimation

of energies. The standard deviation is relatively large which could potentially lead to qualitatively wrong PESs. It would thus be useful to compare the actual vs predicted surfaces along different normal modes, but this has not yet been implemented. Another interesting observation is that the triplet states appear to be better described, having a mean error closer to zero and smaller standard deviations. Although the idea is nicely illustrated here, the ultimate performance of this approach should be judged based on MCTDH calculations performed on potentials generated in this way.

5.5 Case studies

5.5.1 Case I: Building up a QC database

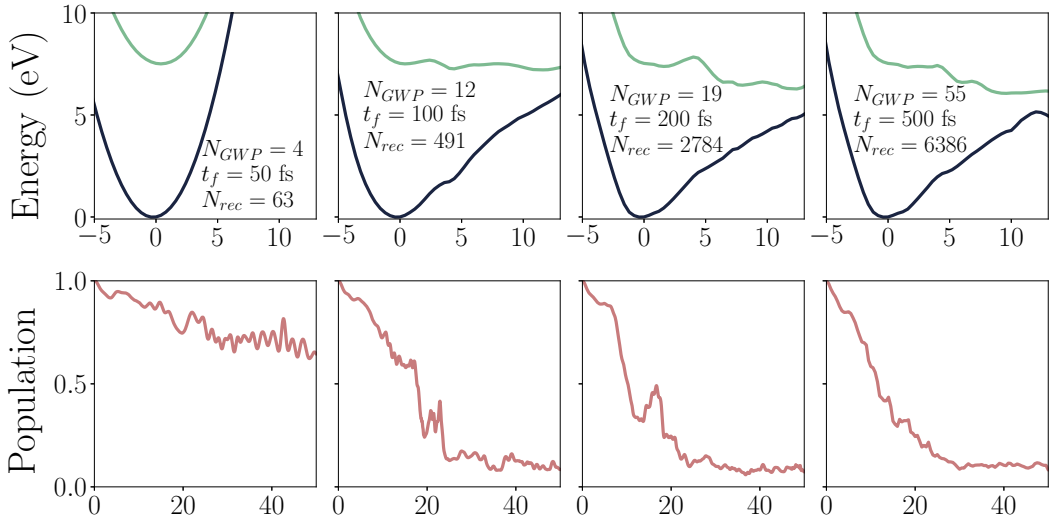


Figure 5.8: *Top:* Ground (blue) and second excited (green) state of formic acid along the O–H stretch normal mode (units are mass-frequency weighted) after 4 consecutive DD-vMCG simulations. The number of GWPs, N_{GWP}, total simulation time, t_f, and the resulting number of database records, N_{rec}, are given in each panel. *Bottom:* Decay of the population density of the initially excited state (green in top panel), during the first 50 fs.

The iterative process of populating a QC database by running increasingly large DD-vMCG calculations is shown in Figure 5.8. The plots demonstrate how a cut along the PES of formic acid changes during four consecutive simulations, each run with a different number of basis functions and for a different amount of time. Also plotted is the decay in population density of the upper state. Notice how during the first simulation, the surfaces along this particular mode do not deviate from the LHA. Only after more wavepackets are added, and the calculation is run for a longer

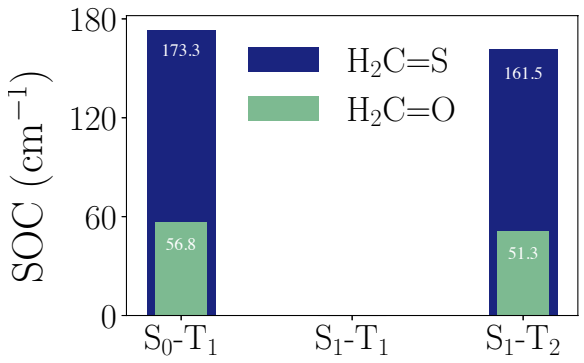


Figure 5.9: SOC strengths between selected Singlet-Triplet state pairs of formaldehyde and thioformaldehyde, at their respective FC geometries (see Figure 5.3 for state characters).

time, do they get updated (with this updating, the potential need for CAPs along dissociative states should also become apparent). Additionally, the simulations tend towards convergence as the database gets larger and the size of the basis increases. This is inferred by comparing the changes in state population, and the fact that the plots get increasingly smooth. In principle, one should continue to populate the database in this way until no more points are added, indicating that all relevant regions of the configuration space have been fully explored by the wavepackets. The number of wavepackets should similarly be increased until the change in population transfer doesn't vary any longer, indicating fully converged results. Reaching this point in practice is rare however.

5.5.2 Case II: Population transfer in thioformaldehyde

To illustrate a situation in which the population transfer processes constitute the main properties of interest, an analysis of the dynamics of thioformaldehyde is presented. First of all, the fact that the sulphur atom is heavier than oxygen in regular formaldehyde leads to an increase in SOC.[337–341] This is an example of the heavy atom effect, which can increase the magnitude of SOC by 3-5 times, confirmed by Figure 5.9.[342,343] Notice also that El-Sayed's rule is nicely obeyed, stating that only states of different orbital symmetry give rise to non-vanishing SOC terms between them.[344] In those cases, spin-forbidden population transfer becomes weakly allowed, leading to possible ISC phenomena. Importantly though, it is both the strength of the coupling *and* the energy gap between the states that dictates how efficient ISC ultimately is. Unsurprisingly then, the possibility of ISC in thioformaldehyde, along with it's small size, has

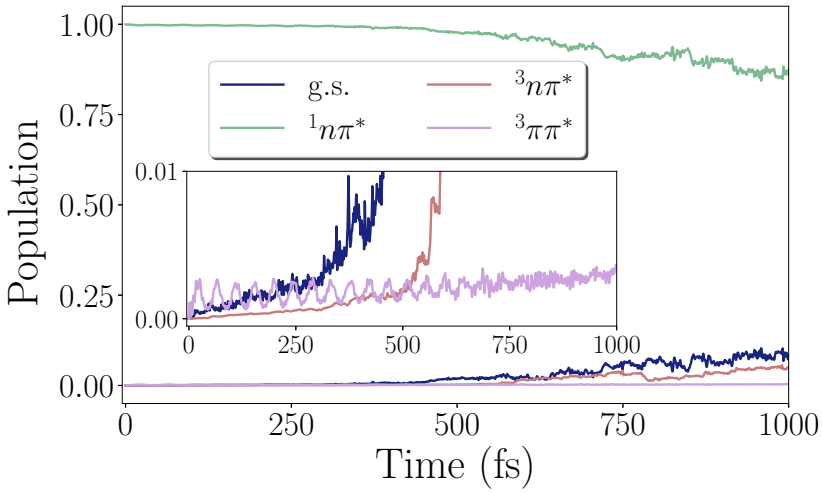


Figure 5.10: Diabatic state populations CASPT2(12,10) level of theory. Inset showing close-up of populations.

resulted in it becoming a popular candidate for high-level theoretical investigations of singlet-triplet interactions, and as such, has been the subject of several computational studies.[289,345–350]

Since non-adiabatic QD simulations are particularly well suited to the modelling of such interactions, they are the obvious tool for the quantitative determination of ISC yields and timescales. On a practical note however, the calculation of relativistic effects, including but not limited to SOC, can add a significant computational overhead to electronic structure calculations if one wishes to calculate them properly. Therefore, a route that is often taken when dealing with light elements is to ignore effects like, for example, electrons travelling near the speed of light, and focus solely on the spin-orbit relativistic effect. This partition allows for SOC to be added to the non-relativistic Hamiltonian as a perturbation, and reduces the overall cost massively. For the specific implementation of this approach used in the OpenMolcas package, see ref. [351]. In QUANTICS, the effect of SOC is included in the wavepacket propagation by adding off-diagonal coupling terms to the system Hamiltonian, corresponding to the spin-diabatic representation (see also chapter 2, section 2.3.1).

Coming back to thioformaldehyde, the electronic structure used for the DD-vMCG dynamics was multi-state CASPT2. Only 4 states are required in this model, the ground and first singlet state, and the two lowest lying triplet states (see Figure 5.3). The active space is made up of 12 electrons in 10 orbitals (given in Appendix D, section

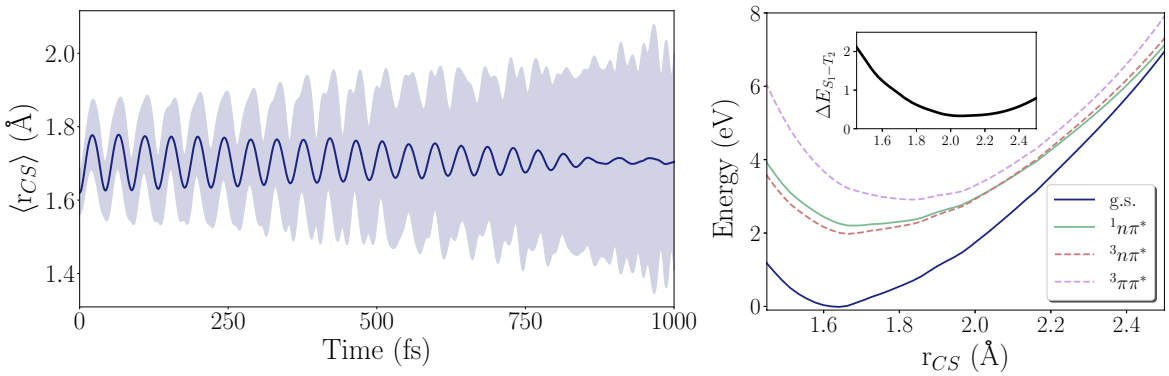


Figure 5.11: *Left:* Expectation values of C–S stretching mode in thioformaldehyde, centre of wavepacket shown as blue line with its width represented by the shaded area. *Right:* PES along C–S stretch, with inset showing variation of S_1-T_2 energy gap, calculated at CASPT2(12,10) level of theory.

D.1), which was found to be suitable in previous work.[289,348] To determine whether or not any ISC occurs, the diabatic state populations during the first 1000 fs are extracted (Figure 5.10), and the mechanism behind the different population transfers will be considered for each state, starting with T_2 ($^3\pi\pi^*$).

Even though there is non-negligible SOC between S_1-T_2, no significant ISC is observed between these states. This is in line with experimental evidence, suggesting that no ISC is expected on sub-picosecond timescales.[347,352,353] There are however small oscillations in the $^3\pi\pi^*$ population, which are investigated. First of all, it is noted that it would be unlikely that any population is transferred directly from $^1n\pi^*$ within the FC region, because the energy gap between the two states is too high at that point. Assuming then that certain nuclear motions modulate the energy gap between the pair of states, the expectation values of the normal modes were analysed, revealing that the C–S stretching mode is most strongly excited. This motion is plotted in Figure 5.11, along with the corresponding PES cut. The inset in the figure confirms that motion along this coordinate does indeed affect the singlet-triplet gap. Since the level of ISC is affected by the energetic proximity of two states, population transfer may become possible at elongated C–S bond lengths, when the gap between S_1 and T_2 is at its smallest. Even though this bond length does increase, as evidenced by the the coherent oscillation of all the wavepackets, lasting for about 800 fs, no significant build-up of density on T_2 is observed. However, the weak, periodic population and de-population of this state, seen in Figure 5.10, matches the oscillating C–S bond, which shows the same period of about 45 fs (experimentally found to be 41 fs[352]). It

seems as though a fraction of the density leaks to the T_2 state every time the C–S bond is at its maximum, but then transfers back to the S_1 state as the bond shortens again.

Next, internal conversion back to the ground state is discussed. Such a process is often facilitated by an energetically accessible CoIn between the two states. Searching for the minimum energy intersection[†], a point of degeneracy was found when the two hydrogen atoms are displaced sufficiently out of the plane of the molecule (a motion corresponding to the lowest frequency normal mode). At such a geometry, along with an elongated C–S bond, internal conversion is thought to take place efficiently.[347,355,356]

Finally, and surprisingly, the lowest triplet state shows a rising population despite being El Sayed forbidden. The most likely rationalisation of this phenomenon is a two-step process where population is first transferred to S_0 via internal conversion, creating a hot ground state. From there, vibrational relaxations brings the molecule to a crossing between S_0 - T_1, where ISC is allowed due to non-zero SOC (see Figure 5.9). Although somewhat exotic, this mechanism seems plausible given that the CoIn between S_0 - T_1 is found to be at a lower energy than the one which facilitates internal conversion, and along roughly the same coordinate, i.e. an out-of-plane geometry with slightly elongated C–S bond. All these processes are discussed in more detail in an article currently under preparation.[357] Since the conclusion of that work is still under consideration, no further elaboration will be given here.

5.5.3 Case III: Pruning the QC database

The idea of pruning a database based on the distribution of different calculated quantities is applied to data generated from formaldehyde simulations. Figure 5.12 shows the use of a suitable cut-off (i.e. value of n), defined using the distribution of a given property. Additionally, this data emphasises the need for a pruning procedure since some very obvious outliers are observed, which are sure to negatively impact the quality of the dynamics.

[†]A simple optimisation is used to find the minimum energy conical intersection by locating the intersection seam between two states, and minimising its energy.[354]

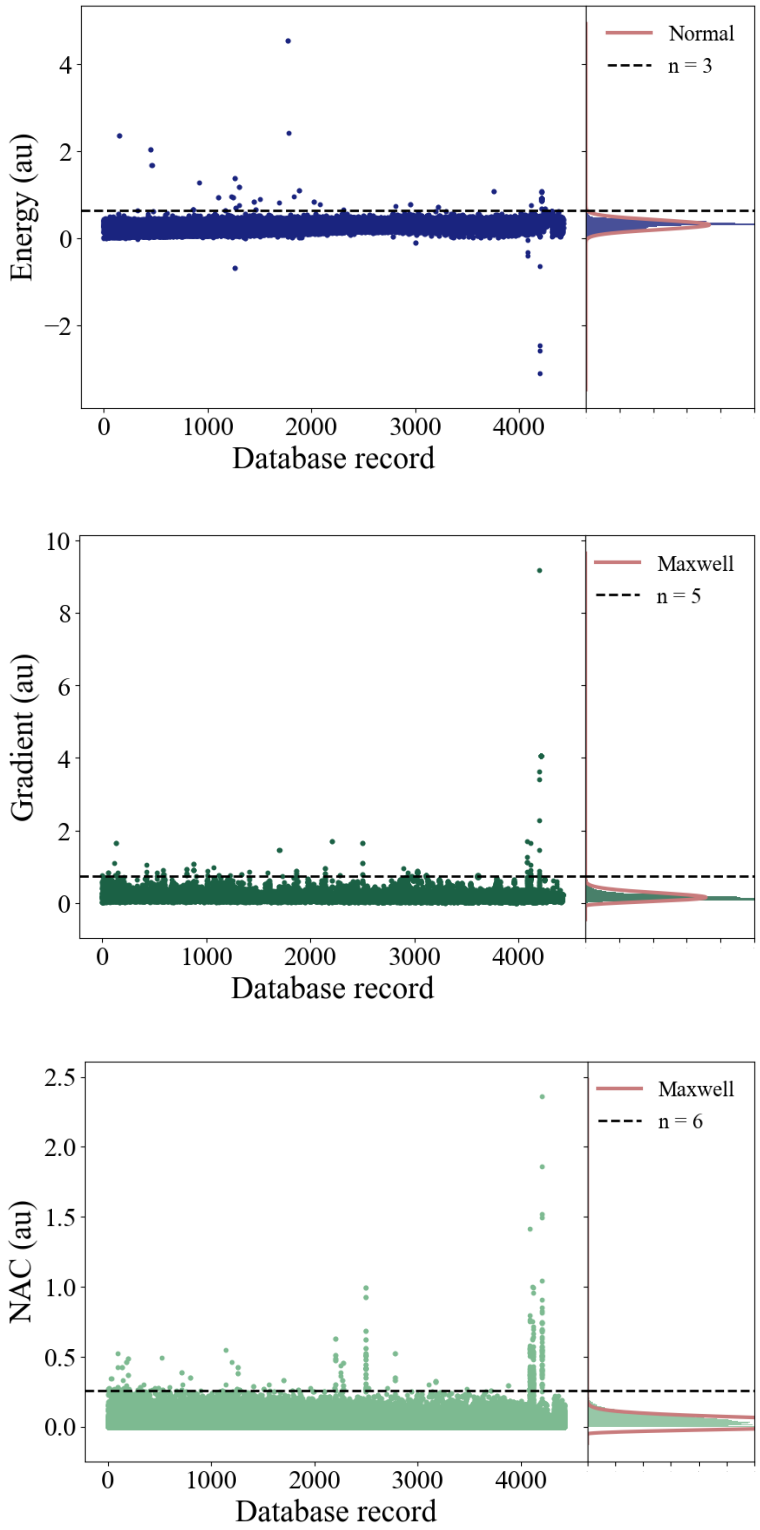


Figure 5.12: Pruning of the (from top to bottom) energies, gradients, and couplings in a QC database. *Left:* individual database entries (all in atomic units). *Right:* probability distributions of Figure 5.5, and cut-off defined by $n \cdot \mathrm{Var}(x)$. Data shown is from formaldehyde calculations.

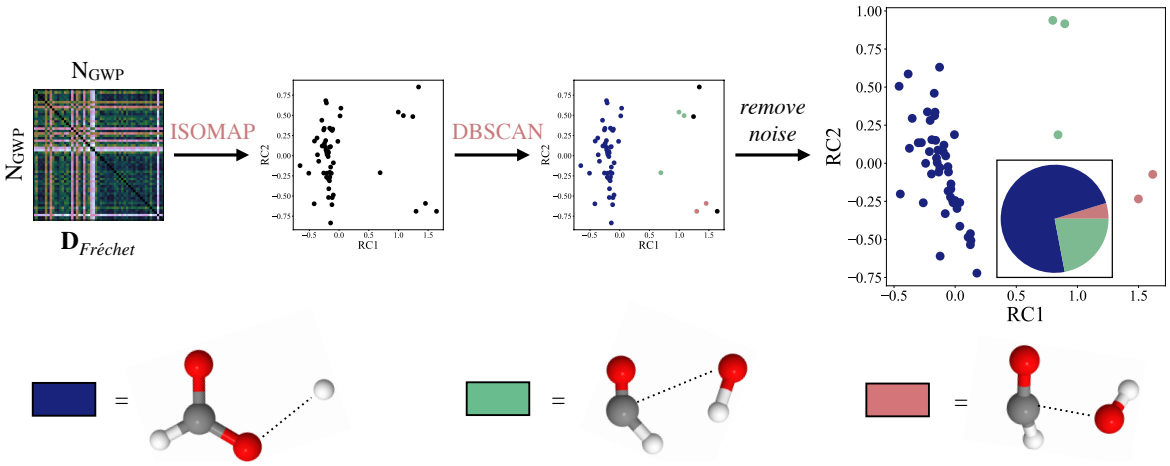


Figure 5.13: Automatic computation of branching ratios, illustrated by trajectories obtained after $\pi\pi^*$ excitation of formic acid. Average final geometry of each cluster is shown, with dotted line indicating the dissociating bond.

5.5.4 Case IV: Automated calculation of product ratios

The product ratios of formic acid are determined using the automated procedure based on dimensionality reduction and clustering. The molecule was initially excited to its $\pi\pi^*$ state, and within about 200 fs it fully dissociated into different fragments. Having calculated the expectation values of the centres of all 55 GWPs over time, these trajectories are analysed using the algorithm outlined previously. The different key steps are graphically represented by the scheme in Figure 5.13, showing $\mathbf{D}_{Fréchet}$, the dimensionality reduction to 2 reduced coordinates, the identification of 3 separate clusters, and finally the computation of the relative ratios based on the GGP of the trajectories in each cluster, shown in the pie chart. Additionally, the average trajectory corresponding to each cluster can be determined. This is done by simply computing the mean Cartesian coordinates of all member trajectories at each time-step, weighted by their respective GGPs. The final frame of these average, representative trajectories is shown in the bottom of Figure 5.13. For a more detailed analysis of these formic acid results, see section 5.6.2.

5.6 Application

Having established a general workflow of the DD-vMCG method, it is now put into practice by applying it to the photoinduced dynamics of three of the four carbonyl molecules introduced at the beginning. Thioformaldehyde is left out as it has already

been discussed in some detail in the case study about ISC (section 5.5.2). Although the discussion here won't be nearly as thorough and detailed as the one presented in chapter 3 for methanol, it should provide a useful addition to the sometimes abstract discussions in the previous sections.

Having chosen the molecules of interest, and having identified a suitable electronic structure method, the procedure outlined in Figure 5.2 can be followed. Looking back at the discussion of the initialisation stage, one would like to use as an input the Cartesian coordinates of a molecule, and obtain the initial excited state wavepacket necessary to start the dynamics simulations, as the output ($\mathbf{x} \rightarrow \Psi_0$). Such a transformation is trivial when approximating this wavepacket as the projection of the lowest vibrational eigenstate of the ground state, hence only requiring the ground state frequencies. These can be found in Appendix D, section D.6 for all four molecules, along with their optimised geometries. This step is followed by using XDW-CASPT2 calculations to obtain the set of all necessary molecular properties for each electronic state, $\{E, F, H, \Lambda\}$. A LHA is then used to provide the initial (harmonic) potential surfaces, along with nuclear gradients which will act on the wavepackets once the propagation is initiated.

As for the propagation itself, the following conventions are used for each molecule: all $3N$-6 normal modes are included, wavepackets are not partitioned (each GWP is $3N$-6 dimensional), all simulations are started at the FC point ($q(t_0) = 0$ for all normal modes), and the initially excited state is the bright $\pi\pi^*$ state. The total simulation time varies between 300 - 500 fs and the total nuclear basis size was determined based on the previously mentioned $n \cdot 2d + 1$ rule, where $n = 3$ was chosen for all molecules, except formamide where $n = 1$ (the reason being that previous work indicates that 25 GWPs is sufficient[157]). Initially, for all molecules a small calculation with only 4 GWPs was run for 50 fs, after which both the basis size and the final time were gradually increased. The energy and norm conservation was checked each time, and the databases were pruned to remove failed points where necessary. Based on the most active normal modes and the shape of the PES along these modes, suitable CAPs were introduced to capture any dissociative motion. A final production run was then carried out, the results of which are summarised next.

The analysis is based on determining both relevant electronic and nuclear motion, and using this information to make general conclusions about the behaviour of the

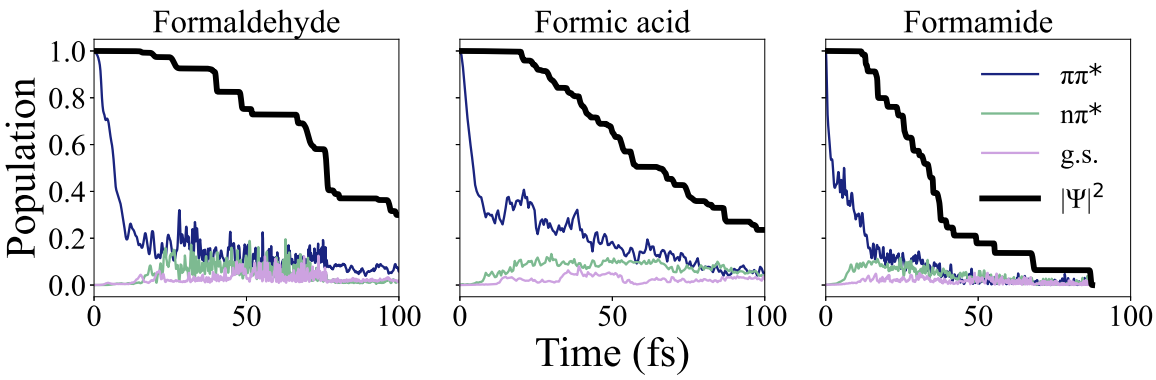


Figure 5.14: Diabatic state populations of $\pi\pi^*$, $n\pi^*$, and ground state (g.s.) for three carbonyl molecules during the first 100 fs of DD-vMCG dynamics. The total density is also plotted (thick black line).

molecules during the dynamics. Note however that the quality of the simulations is not always high enough to provide trustworthy results. This will be commented on as well. In addition, referring back to the introduction, no thorough literature search was carried out, only a superficial scan. This was done to minimise bias during the analysis, but it does mean that it becomes very difficult to judge the accuracy of a given simulation. It is stressed once again that this ultimately is not the goal of these calculations. They merely serve to illustrate the ideas put forward in this chapter about QD workflows.

5.6.1 Electronic motion: state populations

Information about the "motion" of electrons is obtained by evaluating the diabatic state populations. In Figure 5.14, the population of the initially excited $\pi\pi^*$ state is plotted, along with that of the $n\pi^*$ and the electronic ground state, as well as the total density given by the square of the norm of the wavefunction. The other states have mostly Rydberg character and tend to be of less interest in a photochemical context. Comparing the "smoothness" of the plots to estimate the level of convergence, the formaldehyde calculations can be said to be furthest from convergence, while the other two are only slightly better. In terms of general features, notice the striking similarities between the molecules. In all three cases, population is transferred very rapidly from the $\pi\pi^*$ to other states during the first 10-20 fs, after which CAPs start absorbing incoming wavepackets and the total norm eventually drops to zero.

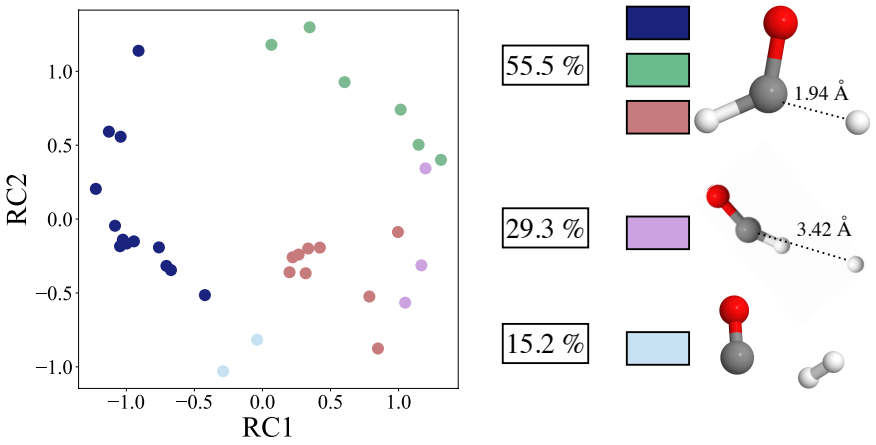
The rate of dissociation is slightly different and increases from formaldehyde being

the slowest, to formamide dissociating fastest. This means that some unbound, dissociative surfaces must exist. Identifying the nature of those pathways cannot be done using solely the state populations, but requires an analysis of the nuclear DoFs, which is carried out in the next section.

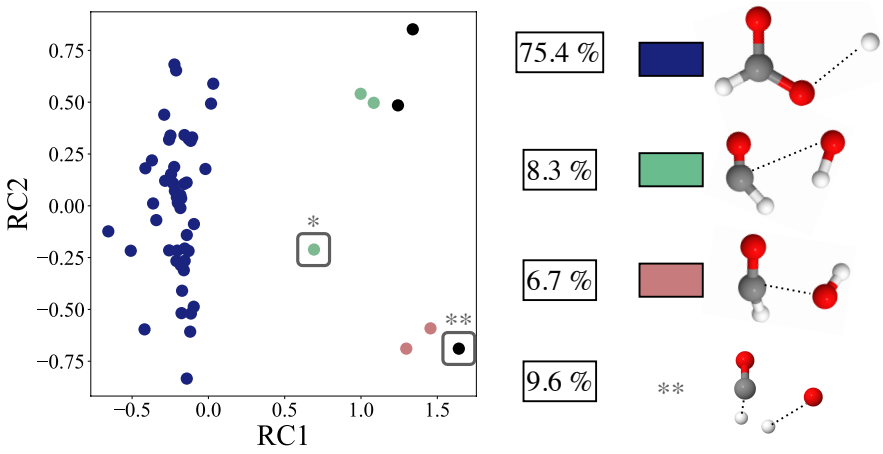
5.6.2 Nuclear motion: clustering & product ratios

Identifying key nuclear coordinates is crucial to unravel what drives the photoinduced dynamics of a specific molecule. To this end, a procedure for the automatic determination of different photoproducts, and their respective yields, is given in section 5.4.2. It is based on a similarity measure between trajectories which is then used for dimensionality reduction and clustering. The application of this procedure to the three carbonyl molecules is presented in Figure 5.15, showing the outcome of ISOMAP and DBSCAN, i.e. 2-D plot of trajectories, coloured according to their corresponding clusters. Shown to the right of these plots are the weights of each cluster corresponding to the relative yields of different photoproducts, and the average molecular trajectory characterising them. In combination, this information allows one to determine 1) the most important/active DoFs, and 2) the nature and branching ratios of any photoproducts formed during the dynamics.

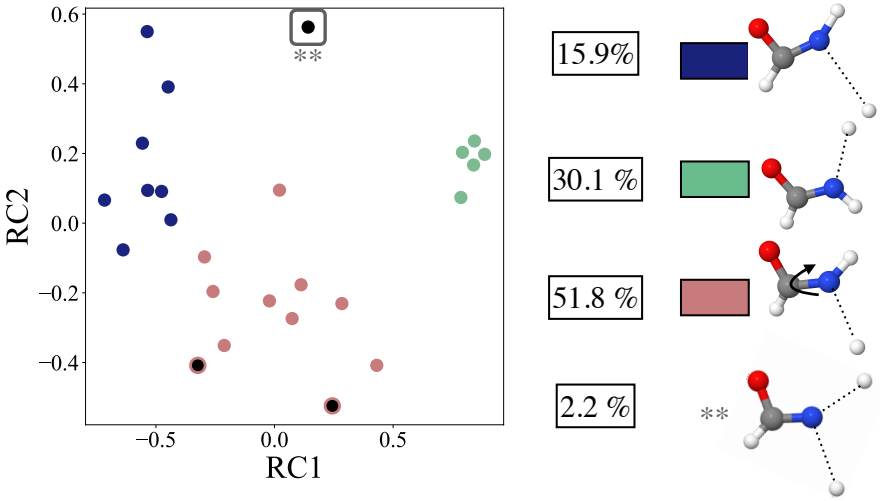
A few general comments about the level of automation of this analysis are in order, before continuing with a molecule-specific discussion of the results. It is clear that clustering trajectories still requires some manual input from the user. Firstly, the EPS value needs to be chosen, which is done here based on a sensitivity analysis. As commented previously however, having this flexibility of defining the "strictness" of the clustering may be a useful feature. A more important point to highlight is that for each molecule, some adjustments have been made even after a suitable EPS parameter was chosen. In Figure 5.15a for example, 3 clusters are merged together since they represent essentially the same product. It is possibly a matter of opinion, or perhaps of scope, whether or not products are indeed the same, and therefore intuition might play a role here, taking away somewhat from the desired black-box nature of the approach. Next, in Figure 5.15b, the trajectory marked with "**" is labelled as noise by DBSCAN, but upon inspection it was found to correspond to a potentially viable product channel. Similarly in Figure 5.15c, some of the noisy trajectories are manually added to existing clusters upon inspection. Even though it is desirable to



(a) Formaldehyde; EPS = 0.6, $N_C = 5$, $N_{noise} = 1$ (removed)



(b) Formic Acid; EPS = 0.7, $N_C = 3$, $N_{noise} = 3$



(c) Formamide; EPS = 0.42, $N_C = 3$, $N_{noise} = 3$

Figure 5.15: ISOMAP and DBSCAN results for three carbonyl molecules. *Left:* trajectories in reduced coordinates. *Right:* product yields and corresponding average structures. Also given are the EPS parameters used in DBSCAN, and the number of clusters, N_C, and noisy points, N_{noise}, identified.

use a clustering method which is able to identify noise, it is clear one should always check those trajectories and confirm that they should indeed not be counted as valid product channels. This is evidently not always the case because of the fact that a cluster needs at least two elements, and hence forces any rare event represented by only a single trajectory to be labelled as noise.

Moving on to comments on the individual molecules, the outcome of the clustering is discussed, and the reduced coordinates identified by ISOMAP are related to corresponding nuclear DoFs. Note that this assignment is speculative and based entirely on educated guessing and intuition. It is, however, possible to confirm whether or not a given assignment is valid by looking at the correlation between a reduced coordinate and the assigned nuclear coordinate. An example of such a validation exercise is provided in Appendix D, section D.7, applied to the formic acid results, but will not be carried out for the other molecules. It is also pointed out that although the number of reduced coordinates was chosen to be 2 in all cases, it may sometimes be more appropriate to chose a higher number, in which case the assignments will be impacted.

Formaldehyde is discussed first. Three distinct products are found; C–H bond elongation, C–H bond dissociation, and formation of CO + H_2. Interpreting the ISOMAP plot in Figure 5.15a, it would be reasonable to assume that RC1 and RC2 correspond to a measure of the C–H bond length, and the H–C–H and/or O–C–H angles, respectively. Ignoring for now the cluster corresponding to CO + H_2, the fact that the remaining clusters are distributed across the entire plot indicates that there is much variation in the angle and bond length to which the reduced coordinates refer. This is also the reason for the "over-clustering" of the species with an elongated C–H bond, and hence why the clusters were manually merged. Finally, the CO + H_2 channel stands out from the other two since it cannot be characterised by a simple combination of normal modes, bond lengths, or angles. It's location on the ISOMAP plot is not very revealing, but this of course doesn't matter with respect to determining its relative yield.

The simulations of formic acid predict 4 distinct product channels; O–H dissociation, *trans*[†] C–OH dissociation, *cis* C–OH dissociation, and CO_2 + H_2. The first three are readily identified by DBSCAN, whereas the last one had only 1 representative trajectory (marked with "∗∗") and is therefore labelled as noise by DBSCAN.

[†] "*trans*" refers here to the hydrogen atom being opposite the carbonyl oxygen.

Interestingly, two different channels for C–O bond breaking are distinguished. It is possible that the *cis* channel represents the formation of hydroxyl radicals, whereas the *trans* dissociation is a pathway to the formation of H_2O. Notice how the OH fragment is oriented differently in both structures, and how the H–C–O angle opens up, allowing for H-abstraction by the OH radical which has the oxygen atom pointing in the right direction. Both the HCO + OH and CO + H_2O products have indeed been documented, as well as the production of H atoms.[358,359] Whether or not this analysis is ultimately correct, it is at least reassuring that ISOMAP and DBSCAN are sensitive to these two different pathways leading to the same fragments.

Another interesting observation related to these two fragmentation pathways is highlighted in the trajectory marked "∗". It is located at approximately equal distance from each cluster, at least in this 2-D plane. Upon inspection, it was found that this trajectory corresponds to the molecule undergoing nearly a full twist around the O–C–O–H dihedral, before the C–O bond breaks. Ultimately though, the OH fragment dissociates with hydrogen pointing away from the carbonyl oxygen (i.e. *trans*) and its cluster label is therefore justified.

The interpretation of the ISOMAP plot is more straightforward than for formaldehyde since the individual clusters are well separated. The RC1 coordinate clearly corresponds to the C–OH bond length, and RC2 is some measure of the O–C–O–H dihedral. Notice that neither one of the reduced coordinates represents the O–H distance, despite the fact that its dissociation accounts for over 75% of the total wavefunction. This fact indicates that dimensionality reduction does not necessarily map the 2 most important coordinates in terms of how active they are during a simulation, but rather those coordinates that are most important in order to differentiate trajectories. Such an insight is typically harder to obtain than merely knowing which modes are the most active, and can be quite valuable when, for example, aiming to build a model in which only a few important coordinates are included.

Lastly, formamide is seen to exhibit all three different kinds of N–H bond breaking in the amine group: *trans* N–H dissociation, *cis* N–H dissociation, and simultaneous N-2H dissociation. Two of those channels are found by clustering, but N-2H dissociation only occurs in a single trajectory and is therefore labelled as noise (marked with "∗∗"). Furthermore, the breaking of the *trans* N–H bond is observed in two separate clusters. Rather than merging them, it was decided that they are sufficiently different

since the N–H bond breaking occurs via two distinct mechanisms. One is a fast ballistic dissociation, while the other is slower and involves a twist of the O–C–N–H dihedral. Note that two noisy points are assigned to this second cluster as well, based on sufficient similarity. As far as the *cis* N–H dissociation, it appears to be a rapid ballistic dissociation.

Looking at the distribution of data in the ISOMAP plot, the points in the two clusters on the left are fairly spread out, presumably due to many subtle changes in various bonds, angles and/or dihedrals in the trajectories. This makes it more difficult to come up with a physical interpretation of the reduced coordinates. One guess would be that both RC1 and RC2 involve the length of the *cis* N–H bond, with RC1 also containing contributions from the O–C–N–H dihedral, and RC2 involving the *trans* N–H bond as well. Overall, it seems clear that during these simulations many different vibrational modes of formamide are excited, leading to various similar but slightly distinct ways of eventually dissociating into one of the four possible products observed here.

5.6.3 Flux

To put this analysis into perspective, an alternative way of identifying the most active/important nuclear DoFs is also demonstrated. This method is based on measuring the flux of the density along a coordinate (in this case, normal modes). Calculating the flux reveals how much of the total density goes into one particular channel. As such, it is particularly useful to determine the most important dissociation pathways, and was indeed used for that purpose in chapter 3 (see also section 3.2.3 in that chapter for a description of the theoretical background of flux analysis). For the three carbonyl molecules, the flux was measured along both directions of all normal modes, and the dividing surfaces were placed at $q = 15$ mass-frequency weighted normal mode units. Figure 5.16 plots the flux along those coordinates that resulted in non-negligible magnitudes. These plots either confirm that the main nuclear motions influenced by the dynamics are the same ones that were identified and discussed previously, or they provide additional and complementary information about relevant coordinates.

In the case of formaldehyde, the most flux is measured along Q_3 and Q_6, which correspond, respectively, to CH_2 scissoring and CH_2 stretching. This agrees exactly with the assignment of RC1 and RC2 in the ISOMAP plot being related to the C–

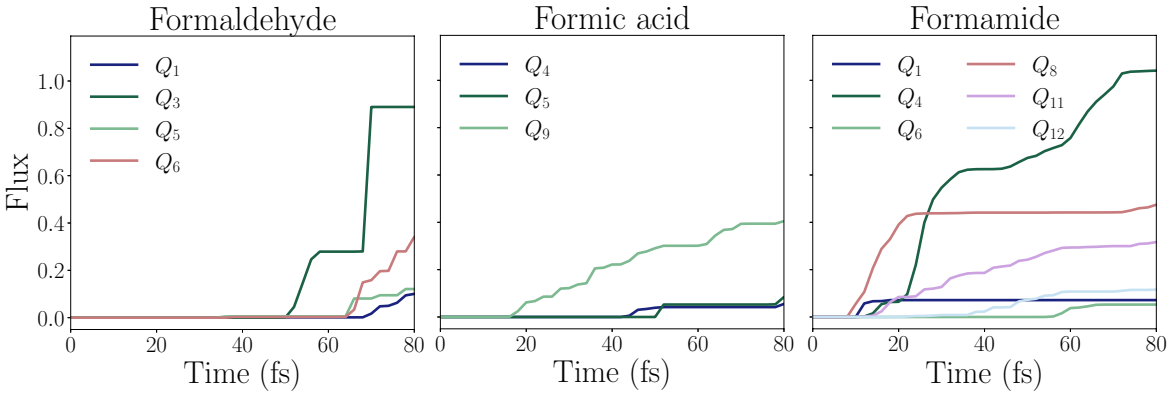


Figure 5.16: Integrated flux measured along key normal modes for three carbonyl molecules. See Appendix D, section D.6 for mode descriptions.

H bond length, and the C–H–C and/or O–C–H angles, respectively. In the case of formic acid however, such an agreement is only partially found. Modes Q_4 and Q_5 correspond to combinations of C–OH bond stretching and C–O–H angle bending, the same coordinates assigned to RC1 and RC2, but by far the most flux is found along Q_9; the O–H stretching mode. This is no surprise since, as commented before, dimensionality reduction and clustering identifies important modes based on how well they allow one to differentiate between products, whereas a flux analysis only reveals the magnitude of the density along a mode. Thus, in this particular case, the two methods are complimentary.

Finally, the formamide results confirm the assertion that many different modes are activated and play some role during the excited state dynamics. Most of the flux is however concentrated in modes Q_4 (C–N–H bend), Q_8 (NH_2 scissor), and Q_{11} (NH_2 stretch), which only partially matches with the RC1 and RC2 assignments made before. The two analyses would again seem to be complimentary.

5.6.4 Molecular complexity

As a final remark on this section covering the results of three similar carbonyl molecules, it is emphasised that although the analysis of non-adiabatic dynamics results can be highly system-specific and complex, the actual steps one might follow to arrive at an explanation remain relatively fixed. In that sense, the use of dimensionality reduction and clustering is proven to be a useful addition to the set of analysis tools by simplifying the task of identifying important features during QD simulations through

automation, which additionally removes some degree of bias.

Extending this idea of the transferability of a workflow from one molecule to another, it is hypothesised that the level of molecular complexity in a system should not massively impact the effort required to carry out a typical QD study. A quantitative measure of complexity was provided in Table 5.1, and despite the observed increase, the way in which the molecules were treated remained the same and valid conclusions could be drawn by applying the same techniques. Specifically, the use of automated procedures is key in reducing the effect of complexity. To see this, consider molecular complexity from a purely graph theoretic perspective as an increase in the number of potential product channels (e.g. bond dissociations). Although a chemist will have some intuition about what channels are physically more likely, such reasoning cannot be used reliably in the general case. Hence, a calculation in which only the realistically possible channels are automatically explored (DD-vMCG), and their ratios automatically calculated at the end (clustering), is relatively immune to the complexity of a given molecule.

On the other hand, it must be said that a rise in molecular complexity does pose problems in both the initialisation and propagation stages. In the former, the nature of different heteroatoms can complicate the electronic structure problem, and in the latter, an increasing system size leads to longer simulation times due to problematic scaling with respect to the number of atoms, as well as potential complications due to heavy atoms introducing relativistic effects. These can be seen as rather superficial issues however, which are both steadily improving over time and are expected to continue to do so.

5.7 Summary & concluding remarks

The present chapter differentiates itself from the others in this thesis by shifting the focus from obtaining accurate results or proposing more advanced ways of simulating certain systems, to providing a critical overview of the *procedures* necessary to perform non-adiabatic QD simulations. Specifically, the DD-vMCG method is taken as the central point of discussion, although some considerations reach beyond any one method and apply generally to the field of molecular simulation. By carrying out an analysis of the method's workflow, the idea is to first of all describe the steps re-

quired to successfully perform DD-vMCG calculations, something much easier said than done, and secondly, to identify features of the workflow that can be improved or at least simplified. The motivation behind these aims stems from a clear desire from the community of users and developers of non-adiabatic QD methods to establish more rigorous benchmarks, allowing for better comparison across methods (two specific examples of community-driven initiatives taken towards this goal are the recent prediction challenge, and the roadmap towards common benchmarks).[274,275]

In the first part of this chapter, the discussion of individual steps is accompanied and illustrated by examples of DD-vMCG simulations of 4 similar carbonyls. Despite their similarity, the increase in molecular complexity should be taken into account. Additionally, the photophysical and photochemical phenomena observed in these molecules are aligned with those identified by the recently developed roadmap as being of particular importance (specifically, dissociation and ISC). The simulation of these molecules in a standardised way represents a starting point towards establishing well documented procedures that are accessible and understood by a diverse[†] audience, and therefore invaluable in a benchmarking context.

A conceptual summary of the series of steps required during DD-vMCG simulations is presented in Table 5.4, which contains a big-picture definition of the tasks to be carried out on the left, simplified symbolic algorithms for achieving said tasks are laid out in the middle, and on the right are the required parameters that need to be chosen. The exact details of how to practically carry out each step are described in the relevant sections, referred to in the Table. In the initialisation stage, the wavepacket at $t = 0$ is generated from information calculated at a particular molecular geometry (usually the ground state minimum). This step is simplified by relying on an ideal vertical excitation out of a harmonic ground state, which avoids many practical difficulties related to initial conditions. However, the fact that the electronic structure problem needs to be considered at this stage makes it ultimately quite challenging.

Then, the wavepacket is propagated with an ever increasing number of nuclear basis functions, until a final, suitably converged simulation is obtained. Putting computational requirements aside, this propagation stage can be relatively easily facilitated by automatically detecting elements that might result in failed simulations, such as

[†] "diverse" meaning people familiar with non-adiabatic simulation methods, but not necessarily with specialised knowledge on any particular method.

Table 5.4: Overview of tasks contained in the DD-vMCG workflow.

Task	**Algorithm**	**Parameters**
	INITIALISATION	
Generate initial properties, input, and DB_0 (sec. 5.3.2)	**Input:** $\mathbf{x} = \{x_i, y_i, z_i\}$ **Goal:** $\mathbf{x} \rightarrow \Psi_0$ **Program:** $\mathbf{x} \rightarrow \{E_i, F_i, H_i, \Lambda_i\}_{i=1}^{S}$ $\mathbf{x}, \{E, F, H, \Lambda\} \rightarrow \Psi_0$	Molecule, electronic structure method, nr. of states (S)
	PROPAGATION	
Sequential propagations to build up (and prune) DB, including error checking (Secs. 5.3.3, 5.4.1)	**Input:** Ψ_0 **Goal:** $\Psi_0 \rightarrow \Psi^{N_{GWP}}$ **Program:** $\Psi_0 \rightarrow \Psi^n$: **while** $n < N_{GWP}$: check(Ψ^n) AND check(DB) **if** *check* =✓: $\Psi^n \rightarrow \Psi^{n+\Delta n}$ **else**: prune(DB)	Basis size (n), and nr. of additional functions (Δn), CAPs
	ANALYSIS	
Calculating expectation values	**Input:** Ψ, $\hat{A}$ **Goal:** $\Psi \rightarrow$ *observable* **Program:** $\langle\Psi\vert\,\hat{A}\,\vert\Psi\rangle \rightarrow \langle a\rangle$	Definition of operator $\hat{A}$
Dimensionality reduction and clustering of trajectories (sec. 5.4.2)	**Input:** $G = \{\mathbf{g}_i\}$, $W = \{GGP_i\}$ **Goal:** $G \rightarrow \Phi$ **Program:** $G \rightarrow \mathbf{D}_{Fréchet}$ $\mathbf{D}_{Fréchet} \rightarrow \{d\}$ *ISOMAP* $\mathbf{D}_{Fréchet} \rightarrow \{c\} = C$ *DBSCAN* $C \times W = \Phi$	Nr. of reduced dimensions (d), EPS clustering parameter

spurious energies, gradients or couplings. A suggestion is made for an automatic way of pruning a database during a series of intermediary simulations.

Finally, once a successful production run is obtained, the transformation from $\Psi \rightarrow$ *observables* is carried out by calculating various expectation values of interest, such as diabatic state population, flux, or change in nuclear position. Additionally, a novel methodology for the automatic determination of product ratios is described, based on dimensionality reduction techniques and clustering of trajectories.

Even though the DD-vMCG method is proven to be a flexible and relatively straightforward way to simulate complex excited state processes, there remain some questions about the ultimate accuracy of the method. The fact that reaching convergence is exceedingly difficult and not even well defined may limit its use in some cases. As usual, it will depend entirely on the scope of a particular investigation whether quantitatively accurate results are required and realistically attainable, or whether a qualitative description is sufficient. In any case, DD-vMCG could be a useful exploratory tool, used to obtain information on the PES in relevant parts of the configuration space, and using these surfaces as the starting point for more accurate grid-based dynamics for example. Thinking specifically about systems too large to be modelled in their entirety, identifying the most relevant DoFs is very important in the construction of reduced dimensionality models. In this respect, dimensionality reduction techniques may also be useful.

In the context of obtaining PESs, another tool that is emerging is the use of machine learning techniques. Apart from neural nets used to fit electronic structure data (section 5.4.3), these are not discussed explicitly here because the focus is on the dynamics, but since the quality of a QD result depends massively on the quality of the PES, it should not be ignored. In recent years, supervised machine learning algorithms (i.e. including a training phase) have been employed to address the electronic structure issue, both in terms of accelerating QC calculations and fitting the calculated points, with varying degrees of success.[273,360–364] Over time, such approaches could become the key to facilitating accurate QD simulations of larger and more complex systems by widening the electronic structure bottleneck.

On a related note, the role of increasingly powerful Artificial Intelligence (AI) agents (which are also based on machine learning) ought to be considered. One could imagine providing such an agent with the description of a particular molecule, and

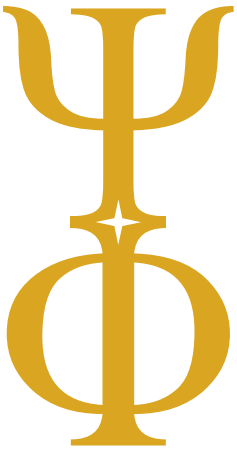
asking it to produce a prediction and explanation of any excited state processes that this molecule is likely to undergo. Presently, such tasks are beyond AI capabilities, arguably because there is no obvious way to derive the details of photophysical or photochemical phenomena from a molecular structure alone, and even if there was, a severe lack of training data hinders an AI's ability to accurately predict such processes. However, as more and more research is being carried out to better understand photochemical reactions, some simple design rules may start to emerge[365] that would allow a supervised learning algorithm to acquire enough "knowledge" to predict reactions of unstudied molecules, or indeed come up with new, better design rules that relate a molecular structure to its photochemical properties. The ability to give a detailed explanation may still be lacking though, and as emphasised in the introductory chapter, merely predicting outcomes without a sound explanation has limited scientific value. For these reasons, the machine learning tools used in this chapter are based on unsupervised learning (i.e. dimensionality reduction, and clustering) to avoid an expensive and possibly unreliable training phase.

As a final positive thought, consider the effect of molecular complexity on the applicability of the outlined workflow. In general, increasing complexity in terms of size and/or constituents only affects electronic structure calculations, and the scaling of the wavefunction propagation. Although often prohibitive, these types of issues are related to the amount of computing power, and the efficiency of available algorithms. These are not conceptual barriers, but are simply the result of a lack of resources and knowledge (both of which can be acquired). More importantly, increasing molecular complexity does not limit the use of the same general workflow, at least in principle. This fact is significant in the sense that it allows for the possibility of faster progress by widening participation and standardising general, non-system-specific procedures. Compare this with a time, not more than 20-30 years ago, where for each individual system a specific kinetic energy operator had to be built, and a set of suitable PESs obtained, both of which could very well take an entire PhD worth of work to complete. This surely is a thought that would make one optimistic about the future of the QD field.



CHAPTER 6

Conclusions



A fact which is not emphasised often is that professional academics are, alongside their research and/or teaching responsibilities, also professional writers.[366] This is true not only in Literature or Philosophy departments, but in STEM fields as well. A large part of being a research scientist involves communicating findings in ways that are compelling and appropriate for a given audience, while also being accurate and reflecting one's integrity. With that in mind, this thesis was written with the aim of presenting a coherent narrative about how computations based on first-principle approaches can be used to describe the intricate reactions of molecules in their excited states, at the quantum-mechanical level. Even though each chapter focuses on a specific topic, they are all connected by the overarching goal of understanding the pathways that are opened up once a molecule is electronically excited, and modelling those accurately. The resulting photophysical and photochemical reactions have long been of great academic and commercial interest, and yet their study remains challenging due to the wonderfully complex nature of matter at the smallest scales, where intuition often breaks down and one must rely on the outcome of complicated experiments and approximate simulations for insights into the relevant mechanisms.

Broadly speaking, there are two main themes to the research presented here; explaining how quantum dynamics simulations provide insights into molecular photoinduced processes by using suitable examples (i.e. methanol, uracil, and 4 carbonyl molecules), and secondly, discussing ways in which existing computational procedures can be extended and/or improved (i.e. implementation of a direct dynamics algorithm including explicit solvation, and a detailed description of the DD-vMCG workflow including suggested improvements). Taken together, they represent a rather typical journey of learning how to use specific tools, gaining an understanding of particular phenomena, building up experience, and eventually, identifying problematic areas and implementing possible solutions. This journey is told in 5 chapters, two of which contain relevant background information, and the other three containing newly obtained results. Each is briefly summarised to highlight the main conclusions, and some considerations of possible extensions to the work are given where appropriate.

Chapter 1

The thesis begins with an introduction to the general context in which the work is situated. Starting from the highest level, the idea of computational modelling is

explored and some thoughts are given about the role of mathematics in describing physical reality. Within that context, non-adiabatic quantum dynamics simulations are positioned as an accurate, but computationally demanding method of describing molecular motion on multiple electronic states. The field of photochemistry is then introduced as being ideally suited to be studied using the aforementioned simulations, and some key difficulties related to that goal are highlighted. Specifically, the challenge of describing photodissociation reactions is explained. Finally, the role of experimental measurements as a way to validate theoretical methods is discussed.

Chapter 2

The Theory chapter discusses all relevant computational methods, including some historical background, theoretical foundations, and useful applications. Since electronic excitations and their effects on a system are only properly described by quantum theory, the Time-Dependent Schrödinger Equation is taken as the starting point for all calculations. Its time-independent form serves as the basis for electronic structure methods used to obtain and characterise excited electronic states of a molecule at a given nuclear geometry. Many such methods exist, forming a spectrum of computational cost vs accuracy, where a trade-off is always made between the two. Most work described here lies towards the accurate but expensive end, with the electronic structure methods of choice being based on multi-configurational wavefunction approaches such as CASSCF, and its various extensions. These methods are necessary to describe the highly non-equilibrium phenomena encountered in this thesis.

The next part of the chapter is concerned with techniques to propagate a nuclear wavefunction in time, which are invaluable in the description of ultrafast processes. The methods of choice are again on the accurate but expensive side of the spectrum. The focus is put on those which have been variationally derived to provide numerically exact, albeit physically approximate, solutions to the TDSE. These are MCTDH, vMCG, and most importantly its *on-the-fly* extension, DD-vMCG. Acknowledging the fact that no exact solutions are attainable for realistic molecular systems, the various approximations made in the practical application of the aforementioned methods are outlined and evaluated. These can be summarised by first saying that, on the one hand, a fully coupled nuclear basis set is used which is able to describe the quantum mechanical nature of the phenomena of interest, but on the other hand, these approaches are

based on models of an idealised version of real physical systems, making comparison to experimental data challenging. Furthermore, the quality of dynamical simulations is only as good as the underlying PES, which constitutes a major bottleneck for the field as a whole.

Chapter 3

The photodissociation of methanol is used as a detailed example of the capabilities of the DD-vMCG method. Methanol is known to have different photofragmentation pathways, which can be properly described by this approach, provided the underlying electronic structure calculations are stable and accurate.

Detailed information about state populations, wavepacket trajectories and state-specific flux allows for a complete mechanistic picture to be built up, and quantitative branching ratios to be determined. Having said that, it must be noted that since the calculations are not converged, and probably cannot practically be converged, some questions remain about the ultimate quantitative accuracy of the numerical results. Possible avenues to obtain more accurate dynamics from the DD-vMCG ones include transforming the direct dynamics database into global analytical surfaces, in the appropriate sum-of-products form, and using those in grid-based QD methods such as MCTDH, which can be properly converged. Other ways to extend this work include looking at the effects of an environment on the photochemistry of methanol. It would be interesting to compare the gas-phase results with simulations in interstellar ice, since a pronounced difference in the branching ratios has been hypothesised.

Chapter 4

This chapter is about the development of a procedure, referred to as AQUATICS, allowing the simulation of explicitly solvated systems using a combination of DD-vMCG and classical molecular mechanics, known as force field quantum dynamics. Although several approaches exist which include solvent effects, none offer a balanced description of both the short-range intermolecular interactions between solvent and solute, including hydrogen bonding or changes in polarisation, and the long-range effect of bulk solvent. Because of this, there is still much room for improvement.

The method presented here involves a typical direct dynamics procedure, where the required electronic structure calculations are replaced by interacting force fields.

In terms of the dynamics, the Gaussian basis set inherent in the vMCG ansatz is partitioned to allow more flexibility to move on the familiar spectrum of cost and accuracy by introducing a hierarchy of approximations. Testing of this new methodology was done using Uracil, both in the gas- and condensed-phase, and preliminary results suggest that the effect of solvent molecules on population transfer processes can be captured using this approach. The development of this method is still ongoing, and requires much more testing on a variety of different systems.

Chapter 5

Here, a critical analysis of DD-vMCG is given. A discussion of all the separate steps required to perform a simulation using this method are outlined, from the choice of molecule at the beginning, to obtaining physical insights at the end. Being a comparatively young method, the systematic description of its workflow can be enlightening for anyone not familiar with the practical details, but more importantly, it allows one to reflect on where possible improvements could be made.

The key developments on the practical side of the methodology were related to pruning QC databases in order to manage the potential failure of electronic structure calculations, and using unsupervised machine learning tools (dimensionality reduction and clustering) to automatically analyse wavepacket trajectories. Perhaps more important however, is the hypothesis that a well defined workflow is not influenced by a particular molecule or its associated complexity. This was supported by the fact that 4 molecules could be studied in essentially the same way, regardless of their nature. A major weakness here is that the molecules were very similar to each other, and a more thorough investigation ought to be carried out on a more varied set of systems.

An overarching theme throughout this chapter is the idea of formalising the practical aspects of the field of non-adiabatic molecular dynamics in general. By agreeing on common best practices and important points of comparison across different methods, a systematic benchmarking initiative can be undertaken and will likely be beneficial to the field as a whole.

Now, to conclude this thesis a brief reflection on the future of the field of quantum dynamics is given. There are two fundamental factors which determine how quickly progress can be made in this field, measured for example by the accuracy or the system sizes that can be achieved. The first relates to available computing power, and

the second is algorithmic efficiency. Both are increasing over time, the former in a continuous linear fashion approximated by Moore's law,[367] and the latter in large but infrequent steps.[368] Even though Moore's law has a physical limit and cannot continue indefinitely, other forms of hardware will be developed that could effectively continue the trend of growing computing power (an obvious example is quantum computing, which has already found some applications in electronic structure calculations and QD).[369–372] Ultimately though, computation is a physical process and its fundamental limits are therefore governed by the laws of physics. On the other hand, there is no such physical limit when it comes to the development of more efficient algorithms. This depends entirely on the the continued growth of (human) knowledge, which is thought to be infinite.[373]

For these reasons, the most exciting developments in the field are expected to come in the form of new and creative ways of tackling the many challenges faced by computational chemists and physicists, some of which were outlined in the preceding chapters. In fact, the development work done in chapters 4 and 5 reflects this sentiment by considering ways in which the field can progress without relying on superior computer architectures. Of course, exploiting new hardware technologies is also essential for the continued development of accurate molecular models.

As a final thought, ask: what would it take for a niche, young field such as molecular quantum dynamics to become a mature scientific discipline? Almost certainly it will rely on slow, gradual progress and community-wide efforts to increase the fidelity of our simulations to the point that they can be used by experts and non-experts alike to predict and explain an increasing proportion of the enormous chemical complexity that enriches our world. This will require a continuous stream of enthusiastic scientists slowly progressing towards the common goal of *clarifying what is opaque about the bonds that matter make*, which also nicely summarises this thesis in just 10 words.



Bibliography

[1] Deutsch, D. *The Fabric of Reality*; Penguin Science, 1998.

[2] Goodstein, D. *Reference Manual on Scientific Evidence: Third Edition*; The National Academies Press: Washington, DC, 2011; pp 38–54.

[3] Popper, K. R. *Conjectures and Refutations: The Growth of Scientific Knowledge*; Routledge & Kegan Paul, 1963.

[4] Wigner, E. The Unreasonable Effectiveness of Mathematics in the Natural Sciences. *Commun. Pure Appl. Math.* **1960**, *13*, 1–14.

[5] Strocchi, F. *An Introduction to the Mathematical Structure of Quantum Mechanics*; Advanced series in mathematical physics; World Scientific Publishing, 2005; Vol. 27.

[6] Mukherjee, D. Role of mathematics in chemistry. *Curr. Sci.* **2005**, *88*, 371–373.

[7] Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpí, J. L. Molecular dynamics simulations: advances andapplications. *Adv Appl Bioinforma Chem* **2015**, *8*, 37–47.

[8] Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143.

[9] Schrödinger, E. Quantisierung als Eigenwertproblem. *Ann. Phys.* **1926**, *385*, 437–490.

[10] Dirac, P. A. M. Quantum Mechanics of Many-Electron Systems. *Proc. R. soc. Lond. Ser. A-Contain. Pap. Math. Phys. Character* **1929**, *123*, 714–733.

[11] Kimber, P.; Plasser, F. *Comprehensive Computational Chemistry*; Elsevier, 2024; pp 55–83.

[12] Olivucci, M.; Sinicropi, A. *Computational Photochemistry*; Elsevier, 2005; pp 1–33.

[13] Evans, R. C., Douglas, P., Burrow, H. D., Eds. *Applied Photochemistry*; Springer, 2013.

[14] Mai, S.; González, L. Molecular Photochemistry: Recent Developments inTheory. *Angew. Chem. Int. Ed.* **2020**, *59*, 16832–16846.

[15] Celani, P.; Bernardi, F.; Olivucci, M.; Robb, M. A. Excited-state reaction pathways for s-cis buta-1,3-diene. *J. Chem. Phys.* **1995**, *102*, 5733–5742.

[16] Garavelli, M.; Bernardi, F.; Olivucci, M.; Vreven, T.; Celani, P.; Robb, M. A. Potential-energy surfaces for ultrafast photochemistry Static and dynamic aspects. *Faraday Discuss.* **1998**, *110*, 51–70.

[17] Harvey, J. N. Spin-forbidden reactions: computational insight into mechanisms and kinetics. *WIREs Comput. Mol. Sci.* **2013**, *4*, 1–14.

[18] Harvey, J. N. Ab initio transition state theory for polar reactions in solution. *Faraday Discuss.* **2010**, *145*, 487–505.

[19] Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

[20] Marcus, R. A. Electrostatic Free Energy and Other Properties of States Having Nonequilibrium Polarization. I. *J. Chem. Phys* **1956**, *24*, 979.

[21] Förster, T. Diabatic and adiabatic processes in photochemistry. *Pure Appl. Chem.* **1970**, *24*.

[22] Bernardi, F.; De, S.; Olivucci, M.; Robb, M. A. Mechanism of Ground-State-Forbidden Photochemical Pericyclic Reactions : Evidence for Real Conical Intersections. *J. Am. Chem. Soc.* **1990**, *112*, 1737–1744.

[23] Bernardi, F.; Olivucci, M.; Robb, M. A. Predicting forbidden and allowed cycloaddition reactions: potential surface topology and its rationalization. *Acc. Chem. Res.* **1990**, *23*, 405–412.

[24] Meyer, H.-D. *Introduction To MCTDH*; 2011.

[25] Zewail, A. H. *Femtochemistry - Ultrafast Dynamics of the Chemical Bond*; World Scientific: Singapore, 1994.

[26] Chergui, M. In *Femtochemistry*; Chergui, M., Ed.; World Scientific: Singapore, 1996.

[27] Sherril, C. D. *Bond Breaking in Quantum Chemistry*; Elsevier, 2005; Vol. 1; Chapter 4, pp 45–56.

[28] Zewail, A. H. Femtochemistry: Atomic-Scale Dynamics of the Chemical Bond Using Ultrafast Lasers (Nobel Lecture). *Angew. Chem. Int. Ed.* **2000**, *39*, 2586–2631.

[29] Dantos, M.; Rosker, M. J.; Zewail, A. H. Femtosecond real-time probing of reactions. II. The dissociation reaction of ICN. *J. Chem. Phys.* **1988**, *89*, 6128–6140.

[30] Rose, T. S.; Rosker, M. J.; Zewail, A. H. Femtosecond real-time probing of reactions. IV. The reactions of alkali halides. *J. Chem. Phys.* **1989**, *91*, 7415–7436.

[31] Mata, R. A.; Suhm, M. A. Benchmarking Quantum Chemical Methods: Are We Heading in the Right Direction? *Angew. Chem. Int. Ed.* **2017**, *56*, 11011–11018.

[32] Popper, K. R. *The Logic of Scientific Discovery*; Hutchinson & Co., 1959.

[33] González, L.; Escudero, D.; Serrano-Andrés, L. Progress and Challenges in the Calculation of ElectronicExcited States. *ChemPhysChem* **2012**, *13*, 28–51.

[34] Ganoe, B.; Shee, J. On the notion of strong correlation in electronic structure theory. *Faraday Discuss.* **2024**, *254*, 53–75.

[35] Szenes, K.; Mörchen, M.; Fischill, P.; Reiher, M. Striking the right balance of encoding electron correlation in the Hamiltonian and the wavefunction ansatz. *Faraday Discuss.* **2024**, *254*, 359–381.

[36] Evangelista, F. A. Concluding remarks. *Faraday Discuss.* **2024**, *254*, 708–717.

[37] Benavides-Riveros, C. L.; Lathiotakis, N.; Marques, M. A. L. Towards a formal definition of static and dynamic electronic correlations. *Phys. Chem. Chem. Phys* **2017**, *19*, 12655–12664.

[38] Szalay, P. G.; Müller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.* **2012**, *112*, 108–181.

[39] Lischka, H.; Nachtigallová, D.; Aquino, A. J. A.; Szalay, P. G.; Plasser, F.; Machado, F. B. C.; Barbatti, M. Multireference Approaches for Excited States of Molecules. *Chem. Rev.* **2018**, *118*, 7293–7361.

[40] Matsika, S. Electronic Structure Methods for the Description of Nonadiabatic Effects and Conical Intersections. *Chem. Rev.* **2021**, *121*, 9407–9449.

[41] Yost, S. R.; Kowalczyk, T.; Van Voorhis, T. A multireference perturbation method using non-orthogonal Hartree-Fock determinants for ground and excited states. *J Chem. Phys.* **2013**, *139*, 174104.

[42] Small, D. W.; Sundstrom, E. J.; Head-Gordon, M. Restricted Hartree Fock using complex-valued orbitals: A long-known but neglected tool in electronic structure theory. *J. Chem. Phys.* **2015**, *142*, 024104.

[43] Sears, J. S.; Sherrill, C. D.; Krylov, A. I. A spin-complete version of the spin-flip approach to bond breaking:What is the impact of obtaining spin eigenfunctions? *J. Chem. Phys* **2003**, *118*, 9084.

[44] McLean, A. D.; Liu, B. Classification of configurations and the determination of interacting and noninteracting spaces in configuration interaction. *J. Chem. Phys.* **1973**, *58*, 1066–1078.

[45] Avery, J. *Creation and annihilation operators*; McGraw-Hill, 1976.

[46] Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley, 2000.

[47] Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a Systematic molecular-orbital theory for excited-states. *J. Chem. Phys* **1992**, *96*, 135.

[48] Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley, 2005; Chapter Including Electron Correlation in Molecular Orbital Theory, pp 203–248.

[49] Brillouin, L. Les problèmes de perturbations et les champs self-consistents. *J. Phys. Radium* **1932**, *3*, 373–389.

[50] Harrison, R.; Handy, N. Full ci calculations on bh, h2o, nh3, and hf. *Chem. Phys. Lett.* **1983**, *95*, 386–391.

[51] Knowles, P. J.; Schütz, M.; Werner, H.-J. In *Modern Methods and Algorithms of Quantum Chemistry*; Grotendorst, J., Ed.; NIC series; NIC, 2000; Vol. 1; Chapter Ab Initio Methods for Electron Correlationin Molecules, pp 61–151.

[52] Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

[53] Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.

[54] Musiał, M. In *Quantum Chemistry and Dynamics of Excited States*; Gonzalez, L., Lindh, R., Eds.; Wiley, 2021; Chapter 4, pp 77–103.

[55] Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: the Hitchhiker's guide to Fock space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433–462.

[56] Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method . A systematic biorthogonal approach to molecular excitation energies , transition probabilities , and excited state properties. *J. Chem. Phys.* **1993**, *98*, 7029–7039.

[57] Krylov, A. I.; Slipchenko, L. V.; Levchenko, S. V. *Breaking the Curse of the Non-Dynamical Correlation Problem : The Spin-Flip Method*; 2007; pp 89–102.

[58] Levchenko, S. V.; Krylov, A. I. Equation-of-motion spin-flip coupled-cluster model with single and double substitutions: Theory and application to cyclobutadiene. *J. Chem. Phys.* **2004**, *120*, 175–185.

[59] Casanova, D.; Krylov, A. I. Spin-flip methods in quantum chemistry. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4326–4342.

[60] Golubeva, A. A.; Nemukhin, A. V.; Klippenstein, S. J.; Harding, L. B.; Krylov, A. I. Performance of the Spin-Flip and Multireference Methods for Bond Breaking in Hydrocarbons:: A Benchmark Study. *J. Phys. Chem. A* **2007**, *111*, 13264–13271.

[61] Lyakh, D. I.; Musiał, M.; Lotrich, V. F.; Bartlett, R. J. Multireference Nature of Chemistry: The Coupled-Cluster View. *Chem. Rev.* **2012**, *112*, 182–243.

[62] Plasser, F.; Lischka, H. *Multi-Reference Configuration Interaction*; Wiley, 2020; Chapter 9, pp 277–297.

[63] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1989.

[64] Shepard, R. In *Advances in Chemical Physics: Ab Initio Methods in Quantum Chemistry Part 210.1002/9780470142943.ch2*; Lawley, K. P., Ed.; John Wiley & Sons, Ltd., 1987; Vol. 69; Chapter The Multiconfigurational Self-Consistent Field Method, pp 63–200.

[65] Werner, H.-J.; Meyer, W. A quadratically convergent multiconfiguration-self-consistent field method with simultaneous optimization of orbitals and CI coefficeients. *J. Chem. Phys.* **1980**, *73*, 2342–2356.

[66] Roos, B. O. The complete active space self consistent field method and its application to electronic structure calculations. *Adv. Chem. Phys.* **1987**, *87*, 399–466.

[67] Shavitt, I. *Methods of Electronic Structure Theory*; Springer US, 1977; pp 189–275.

[68] Schmidt, M. W.; Gordon, M. S. The construction and interpretation of MCSCF wavefunctions. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233–266.

[69] Werner, H.-J.; Meyer, W. A quadratically convergent MCSCF method for the simultaneous optimisation of several states. *J. Chem. Phys.* **1981**, *74*, 5794–5801.

[70] Levine, B. G.; Durden, A. S.; Esch, M. P.; Liang, F.; Shu, Y. CAS without SCF–Why to use CASCI and where to get the orbitals. *J. Chem. Phys.* **2021**, *154*, 090902.

[71] Szymczak, J. J.; Barbatti, M.; Lischka, H. Influence of the active space on CASSCF nonadiabatic dynamics simulations. *Int. J. Quantum Chem.* **2011**, *111*, 3307–3315.

[72] Tran, T.; Segarra-Martí, J.; Bearpark, M. J.; Robb, M. A. Molecular Vertical Excitation Energies Studied with First-Order RASSCF (RAS[1,1]): Balancing Covalent and Ionic Excited States. *J. Phys. Chem. A* **2019**, *123*, 5223–5230.

[73] Anderson, K.; Malmqvist, P.-Å.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-Order Perturbation Theory with a CASSCF Reference Function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

[74] Finley, J.; Malmqvist, P.-A.; Roos, B. O.; Serrano-andrés, L. The multi-state CASPT2 method. *Chem. Phys. Lett.* **1998**, *288*, 299–306.

[75] Møller, C.; Plesset, M. S. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

[76] Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

[77] Battaglia, S.; Fdez. Galván, I.; Lindh, R. *Theoretical and Computational Photochemistry*; Elsevier, 2023; pp 135–162.

[78] Olsen, J.; Roos, B. O.; Jørgensen, P.; Jensen, H. J. A. Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces. *J. Chem. Phys.* **1988**, *89*, 2185–2192.

[79] Malmqvist, P.-A.; Rendell, A.; Roos, B. O. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *J. Phys. Chem.* **1990**, *94*, 5477–5482.

[80] Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev. B* **1964**, *136*, 864–871.

[81] Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* **1965**, *140*, 1133–1138.

[82] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time-Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997–1000.

[83] Cremer, D. Density functional theory: coverage of dynamic and non-dynamic electron correlation effects. *Mol. Phys.* **2001**, *99*, 1899–1940.

[84] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289–320.

[85] Verma, P.; Truhlar, D. G. Status and Challenges of Density Functional Theory. *Trends Chem.* **2020**, *2*, 302–318.

[86] Taylor, J. T.; Tozer, D. J.; Curchod, B. F. E. On the Topological Phase around Conical Intersections with Tamm-Dancoff Linear-Response Time-Dependent Density Functional Theory. *J. Phys. Chem. A* **2024**, *128*, 5314–5320.

[87] Levine, B. G.; Ko, C.; Quenneville, J.; Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.* **2006**, *104*, 1039–1051.

[88] Herbert, J. M.; Zhang, X.; Morrison, A. F.; Liu, J. Beyond Time-Dependent Density Functional Theory Using Only Single Excitations: Methods for Computational Studies of Excited States in Complex Systems. *Acc. Chem. Res.* **2016**, *49*, 931–941.

[89] Subotnik, J. E.; Alguire, E. C.; Ou, Q.; Landry, B. R.; Fatehi, S. The Requisite Electronic Structure Theory To Describe Photoexcited Nonadiabatic Dynamics: Nonadiabatic Derivative Couplings and Diabatic Electronic Couplings. *Acc. Chem. Res.* **2015**, *48*, 1340–1350.

[90] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *84*, 457–484.

[91] Bernardi, F.; Olivucci, M.; Ragazos, I. N.; Robb, M. A. Origin of the nonstereospecificity in the ring opening of alkyl-substituted cyclobutenes. *J. Am. Chem. Soc.* **1992**, *114*, 2752–2754.

[92] Celani, P.; Bernardi, F.; Olivucci, M.; Robb, M. A. Conical Intersection Mechanism for Photochemical Ring Opening in Benzospiropyran Compounds. *J. Am. Chem. Soc.* **1997**, *119*, 10815–10820.

[93] Vreven, T.; Bernardi, F.; Garavelli, M.; Olivucci, M.; Robb, M. A.; Schlegel, H. B. *Ab initio* photoisomerisation dynamics of a simple retinal chromophore model. *J. Am. Chem. Soc.* **1997**, *119*, 12687–12688.

[94] Atchity, G. J.; Xantheas, S. S.; Ruedenberg, K. Potential energy surfaces near intersections. *J. Chem. Phys.* **1991**, *95*, 1862–1876.

[95] Yarkony, D. R. On the characterization of regions of avoided surface crossings using an analytic gradient based method. *J. Chem. Phys.* **1990**, *92*, 2457–2463.

[96] Migani, A.; Olivucci, M. *Conical Intersections*; World Scientific, 2004; Chapter 6, pp 271–320.

[97] Mead, C. A.; Truhlar, D. G. Conditions for the defintion of a strictly diabatic electronic basis for molecular systems. *J. Chem. Phys.* **1982**, *77*, 6090–6098.

[98] Von Neumann, J.; Wigner, E. *Quantum Chemistry*; World Scientific, 2000; Vol. 8; pp 25–31.

[99] Lewis, G. N.; Kasha, M. Phosphorescence in fluid media and the reverse process of singlet-triplet absorption. *J. Am. Chem. Soc.* **1945**, *67*, 994–1003.

[100] Kasha, M. Phosphorescence and the Role of the Triplet State in the Electronic Excitation of Complex Molecules. *Chem. Rev.* **1947**, *41*, 401–419.

[101] Harvey, J. N. On the accuracy of density functional theory in transition metal chemistry. *Annu. Rep. Prog. Chem., Sec. C* **2006**, *102*, 203.

[102] Li, J.; Yao, Q.; Wu, L.; Hu, Z.; Gao, B.; Wan, X.; Liu, Q. Designing light-element materials with large effective spin-orbit coupling. *Nat. Commun.* **2022**, *13*.

[103] Browne, A. J.; Krajewska, A.; Gibbs, A. S. Quantum materials with strong spin-orbit coupling: challenges and opportunities for materials chemists. *J. Mater. Chem. C* **2021**, *9*, 11640–11654.

[104] Goodenough, J. B. Spin-Orbit-Coupling Effects in Transition-Metal Compounds. *Phys. Rev.* **1968**, *171*, 466–479.

[105] Schnappinger, T.; Kölle, P.; Marazzi, M.; González, L.; de Vivie-Riedle, R. Ab initio molecular dynamics of thiophene: the interplay of internal conversion and intersystem crossing. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25662–25670.

[106] Valverde, D.; Mai, S.; Canuto, S.; Borin, A. C.; González, L. Ultrafast Intersystem Crossing Dynamics of 6-Selenoguanine in Water. *JACS Au* **2022**, *2*, 1699–1711.

[107] Daniel, C.; Heitz, M.-C.; Manz, J.; Ribbing, C. Spin-orbit induced radiationless transitions in organometallics: Quantum simulation of the $^1E \rightarrow {}^3A_1$ intersystem crossing process in $HCo(CO)_4$. *J. Chem. Phys.* **1995**, *102*, 905–912.

[108] Bellonzi, N.; Medders, G. R.; Epifanovsky, E.; Subotnik, J. E. Configuration interaction singles with spin-orbit coupling: Constructing spin-adiabatic states and their analytical nuclear gradients. *J. Chem. Phys.* **2019**, *150*, 014106.

[109] Bellonzi, N.; Alguire, E.; Fatehi, S.; Shao, Y.; Subotnik, J. E. TD-DFT spin-adiabats with analytic nonadiabatic derivative couplings. *J. Chem. Phys.* **2020**, *152*, 044112.

[110] Köppel, H.; Domcke, W.; Cederbaum, L. S. Multimode molecular Dynamics beyond the Born-Oppenheimer approximation. *Adv. Chem. Phys.* **1984**, *57*, 59–246.

[111] Worth, G. A. A model Hamiltonian to simulate the complex photochemistry of benzene. *J. Photochem. Photobiol. A* **2007**, *190*, 190–199.

[112] Franck, J.; Dymond, E. G. Elementary processes of photochemical reactions. *Trans. Faraday Soc.* **1926**, *21*, 536.

[113] Schinke, R. *Photodissociation Dynamics*; Cambridge University Press: Cambridge, 1993.

[114] Persico, M.; Granucci, G. An overview of nonadiabatic dynamics simulations methods , with focus on the direct approach versus the fitting of potential energy surfaces. *Theor. Chem. Acc.* **2014**, *133*, 1526.

[115] Suchan, J.; Hollas, D.; Curchod, B. F. E.; Slavíček, P. On the importance of initial conditions for excited-state dynamics. *Faraday Discuss.* **2018**, *212*, 307–330.

[116] Bonafé, F. P.; Aradi, B.; Hourahine, B.; Medrano, C. R.; Hernández, F. J.; Frauenheim, T.; Sánchez, C. G. A Real-Time Time-Dependent Density Functional Tight-Binding Implementation for Semiclassical Excited State Electron-Nuclear Dynamics and Pump-Probe Spectroscopy Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 4454–4469.

[117] Hernández, F. J.; Bonafé, F. P.; Aradi, B.; Frauenheim, C. G., T.and Sánchez Simulation of Impulsive Vibrational Spectroscopy. *J. Phys. Chem. A* **2019**, *123*, 2065–2072.

[118] Tavernelli, I.; Curchod, B. F. E.; Rothlisberger, U. Mixed quantum-classical dynamics with time-dependent external fields: A time-dependent density-functional-theory approach. *Phys. Rev. A* **2010**, *81*, 052508.

[119] Mignolet, B.; Curchod, B. F. E.; Martínez, T. J. Communication: XFAIMS–eXternal Field Ab Initio Multiple Spawning for electron-nuclear dynamics triggered by short laser pulses. *J. Chem. Phys.* **2016**, *145*, 191104.

[120] Penfold, T. J.; Pápai, M.; Møller, K. B.; Worth, G. A. Excited state dynamics initiated by an electromagnetic field within the Variational Multi-Configurational Gaussian (vMCG) method. *Comput. Theo. Chem.* **2019**, *1160*, 24–30.

[121] Domcke, W., Yarkony, D. R., Köppel, H., Eds. *Conical intersections: Electronic structure, dynamics and spectroscopy.*; World {S}cientific: Singapore, 2004.

[122] Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Conical intersections and their impact on molecular dynamics. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127–158.

[123] Worth, G. A.; Meyer, H.-D.; Köppel, H.; Cederbaum, L. S.; Burghardt, I. Using the MCTDH wavepacket propagation method to describe multimode non-adiabatic dynamics. *Int. Rev. Phys. Chem.* **2008**, *27*, 569–606.

[124] Meyer, H.-D.; Manthe, U.; Cederbaum, L. S. The Multi-Configurational Time-Dependent Hartree Approach. *Chem. Phys. Lett.* **1990**, *165*, 73–78.

[125] Manthe, U.; Meyer, H.-D.; Cederbaum, L. S. Wave-Packet Dynamics within the Multiconfiguration Hartree Framework: General Aspects and application to NOCl. *J. Chem. Phys* **1992**, *97*, 3199–3213.

[126] Beck, M. H.; Jäckle, A.; Worth, G. A.; Meyer, H.-D. The multiconfiguration time-dependent Hartree method: A highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **2000**, *324*, 1–105.

[127] Meyer, H.-D.; Worth, G. A. Quantum molecular dynamics: propagating wavepackets and density operators using the multiconfiguration time-dependent Hartree method. *Theo. Chem. Acc.* **2003**, *109*, 251–267.

[128] Meyer, H.-D.; Gatti, F.; Worth, G. A. *Multidimens. Quantum Dyn. MCTDH Theory Appl.*; Wiley-VCH: Weinheim, Germany, 2009.

[129] Gatti, F.; Lasorne, B.; Meyer, H.-D.; Nauts, A. *Applications of Quantum Dynamics in Chemistry*; LNC; Springer Nature: Switzerland, 2017; Vol. 98.

[130] Bonfanti, M.; Worth, G. A.; Burghardt, I. In *Quantum Chemistry and Dynamics of Excited States*; Gonzalez, L., Lindh, R., Eds.; Wiley, 2021; Chapter 12, pp 383–411.

[131] Light, J. C.; Hamilton, I. P.; Lill, J. V. Generalized discrete variable approximation in quantum mechanics. *J. Chem. Phys.* **1985**, *82*, 1400–1409.

[132] Light, J. C. Discrete Variable Representations in Quantum Dynamics. Time-Dependent Quantum Mol. Dyn. New York, 1992; pp 185–199.

[133] Burghardt, I.; Meyer, H.-D.; Cederbaum, L. S. Approaches to the approximate treatment of complex molecular systems by the multiconfiguration time-dependent Hartree method. *J. Chem. Phys.* **1999**, *111*, 2927–2938.

[134] Raab, A.; Worth, G. A.; Meyer, H.-D.; Cederbaum, L. S. Molecular Dynamics Of Pyrazine After Excitation To The S_2 Electronic State Using A Realistic 24-mode Model Hamiltonian. *J. Chem. Phys.* **1999**, *110*, 936–946.

[135] Vendrell, O.; Meyer, H.-D. Multilayer multiconfiguration time-dependent Hartree method: implementation and applications to a Henon-Heiles hamiltonian and to pyrazine. *J. Chem. Phys.* **2011**, *134*, 44135.

[136] Wang, H. Multilayer Multiconfiguration Time-Dependent Hartree Theory. *J. Phys. Chem. A* **2015**, *119*, 7951–7965.

[137] Römer, S.; Ruckenbauer, M.; Burghardt, I. Gaussian-based multiconfiguration time-dependent Hartree: A two-layer approach. I. Theory. *J. Chem. Phys.* **2013**, *138*, 64106–64113.

[138] Di Maiolo, F.; Worth, G. A.; Burghardt, I. Multi-layer Gaussian-based multiconfiguration time-dependent Hartree (ML-GMCTDH) simulations of ultrafast charge separation in a donor-acceptor complex. *J. Chem. Phys.* **2021**, *154*, 144106.

[139] Martínez, T. J.; Ben-Nun, M.; Ashkenazi, G. Classical / quantal method for multi-state dynamics: A computational study. *J. Phys. Chem.* **1996**, *100*, 2847–2856.

[140] Martínez, T. J.; Ben-Nun, M.; Levine, R. D. Multi-Electronic-State Molecular Dynamics: A Wave Function Approach with Applications. *J. Phys. Chem.* **1996**, *100*, 7884–7895.

[141] Quenneville, J.; Ben-Nun, M.; Martínez, T. J. Ab initio multiple spawning: photochemistry from first principles quantum molecular dynamics. *J. Phys. Chem. A* **2000**, *104*, 5161–5175.

[142] Shalashilin, D. V.; Child, M. S. Multidimensional quantum propagation with the help of coupled coherent states. *J. Chem. Phys.* **2001**, *115*, 5367–5375.

[143] Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys* **1990**, *93*, 1061–1071.

[144] Shalashilin, D. V. Nonadiabatic dynamics with the help of multiconfigurational Ehrenfest method: Improved theory and fully quantum 24D simulation of pyrazine. *J. Chem. Phys.* **2010**, *132*, 244111.

[145] Heller, E. J. Time-dependent approach to semiclassical dynamics. *J. Chem. Phys.* **1975**, *62*, 1544–1555.

[146] Heller, E. J. Frozen Gaussians: A very simple semiclassical approximation. *J. Chem. Phys.* **1981**, *75*, 2923–2931.

[147] Richings, G. W.; Worth, G. A. Multi-state non-adiabatic direct-dynamics on propagated diabatic potential energy surfaces. *Chem. Phys. Lett.* **2017**, *683*, 606–612.

[148] Worth, G. A.; Robb, M. A.; Burghardt, I. A novel algorithm for non-adiabatic direct dynamics using variational Gaussian wavepackets. *Faraday Discuss.* **2004**, *127*, 307–323.

[149] Worth, G. A.; Robb, M. A.; Lasorne, B. L. Solving the time-dependent Schrödinger equation for nuclear motion in one step: Direct dynamics of non-adiabatic systems. *Mol. Phys.* **2008**, *106*, 2077–2091.

[150] Spinlove, K. E.; Vacher, M.; Bearpark, M.; Robb, M. A.; Worth, G. A. Using quantum dynamics simulations to follow the competition between charge migration and charge transfer in polyatomic molecules. *Chem. Phys.* **2017**, *482*, 52–63.

[151] Worth, G. A.; Robb, M. A. Applying direct molecular dynamics to non-adiabatic systems. *Adv. Chem. Phys.* **2002**, *124*, 355–432.

[152] Richings, G. W.; Polyak, I.; Spinlove, K. E.; Worth, G. A.; Burghardt, I.; Lasorne, B. Quantum dynamics simulations using Gaussian wavepackets: the vMCG method. *Int. Rev. Phys. Chem.* **2015**, *34*, 269–308.

[153] Coonjobeeharry, J.; Spinlove, K. E.; Sanz Sanz, C.; Sapunar, M.; Došlić, N.; Worth, G. A. Mixed-quantum-classical or fully-quantized dynamics? A unified code to compare methods. *Philos. Trans. R. Soc. A* **2022**, *380*, 20200386.

[154] Richings, G. W.; Worth, G. A. A Practical Diabatisation Scheme for Use with the Direct-Dynamics Variational Multi-Configuration Gaussian Method. *J. Phys. Chem. A* **2015**, *119*, 12457–12470.

[155] Baer, M. Adiabatic and diabatic representations for atom-molecule collisions: Treatment of the collinear arrangement. *Chem. Phys. Lett.* **1975**, *35*, 112–118.

[156] Pacher, T.; Cederbaum, L. S.; Köppel, H. Adiabatic and quasidiabatic states in a gauge theoretical framework. *Adv. Chem. Phys.* **1993**, *84*, 293–391.

[157] Spinlove, K. E.; Richings, G. W.; Robb, M. A.; Worth, G. A. Curve Crossing in a Manifold of Coupled Electronic States: Direct Quantum Dynamics Simulations of Formamide. *Farad. Discuss.* **2018**, *212*, 191–215.

[158] Christopoulou, G.; Freibert, A.; Worth, G. A. Improved algorithm for the directdynamics variational multi-configurationalGaussian method. *J. Chem. Phys.* **2021**, *154*.

[159] Christopoulou, G.; Tran, T.; Worth, G. A. Direct nonadiabatic quantum dynamics simulations of the photodissociation of phenol. *Phys. Chem. Chem. Phys.* **2021**, *23*, 23684–23695.

[160] Frankcombe, T. J.; Collins, M. A.; Worth, G. A. Converged quantum dynamics with modified Shepard interpolation and Gaussian wave packets. *Chem. Phys. Lett.* **2010**, *489*, 242–247.

[161] Tully, J. C.; Preston, R. K. Trajectory Surface Approach to Nonadiabatic Molecular Collisions: The Reaction of Hˆ+ with D_2. *J. Chem. Phys.* **1971**, *55*, 562–572.

[162] Wigner, E. On the Quantum Correction for Thermodynamics Equilibrium. *Phys. Rev.* **1932**, *40*, 749–759.

[163] Janoš, J.; Slavíček, P.; Curchod, B. F. E. Selecting Initial Conditions for Trajectory-Based Nonadiabatic Simulations. *Acc. Chem. Res.* **2025**, *58*, 261–270.

[164] Prlj, A.; Hollas, D.; Curchod, B. F. E. Deciphering the Influence of Ground-State Distributions on the Calculation of Photolysis Observables. *J. Phys. Chem. A* **2023**, *127*, 7400–7409.

[165] Drukker, K. Basics of Surface Hopping in Mixed Quantum/Classical Simulations. *J. Comput. Phys.* **1999**, *153*, 225–272.

[166] Yonehara, T.; Hanasaki, K.; Takatsuka, K. Fundamental approaches to nonadiabaticity: Toward a chemical theory beyond the Born-Oppenheimer paradigm. *Chem. Rev.* **2012**, *112*, 499–542.

[167] Curchod, B. F. E.; Rothlisberger, U.; Tavernelli, I. Trajectory-based nonadiabatic dynamics with time-dependent density functional theory. *ChemPhysChem* **2013**, *14*, 1314–1340.

[168] Subotnik, J. E.; Amber Jain, A.; Landry, B.; Petit, A.; Ouyang, W.; Bellonzi, N. Understanding the Surface Hopping View of Electronic Transitions and Decoherence. *Annu. Rev. Phys. Chem.* **2016**, *67*, 387–417.

[169] Garavelli, M.; Bernardi, F.; Cembran, A. In *Computational Photochemistry*; Olivucci, M., Ed.; Elsevier, 2005; Vol. 16; Chapter VI, pp 191–223.

[170] Öberg, K. I. Photochemistry and Astrochemistry: Photochemical pathways to interstellar complex organic molecules. *Chem. Rev.* **2016**, *116*, 9631–9663.

[171] Öberg, K. I.; Garrod, R. T.; van Dishoeck, E. F.; Linnartz, H. Formation rates of complex organics in UV irradiated CH3OH-rich ices. *A&A* **2009**, *504*, 891–913.

[172] Satyapal, S.; Park, J.; Bersohn, R. Dissociation of methanol and ethanol activated by a chemical reaction or by light. *J. Chem. Phys.* **1989**, *91*, 6873–6879.

[173] Harich, S.; Lin, J. J.; Lee, Y. T.; Yang, X. Photodissociation Dynamics of Methanol ar 157 nm. *J. Phys. Chem.* **1999**, *103*, 10324–10332.

[174] Lucas, M.; Liu, Y.; Bryant, R.; Minor, J.; Zhang, J. Vacuum ultraviolet photodissociation dynamics of methanol at 121.6 nm. *Chem. Phys. Lett.* **2015**, *619*, 18–22.

[175] Bhattacharyya, D.; Zhang, Y.; Elles, C. G.; Bradforth, S. E. Electronic Structure of Liquid Methanol and Ethanol from Polarization-Dependent Two-Photon Absorption Spectroscopy. *J. Phys. Chem.* **2019**, *123*, 5789–5804.

[176] Lin, J. J.; Lee, X., Y. T. andYang Crossed molecular beam studies of the 0(1D)+CH4 reaction: Evidences for the CH2OH+H channel. *J. Chem. Phys.* **1998**, *109*, 2975–2978.

[177] Harich, S.; Lin, J. J.; Lee, Y. T.; Yang, X. Competing atomic and molecular hydrogen pathways in the photodissociation of methanol at 157 nm. *J. Chem. Phys.* **1999**, *111*, 5–9.

[178] Champenois, E. G.; Greenman, L.; Shivaram, N.; Cryan, J. P.; Larsen, K. A.; Rescigno, T. N.; McCurdy, C. W.; Belkacem, A.; Slaughter, D. M. Ultrafast photodissociation dynamics and nonadiabatic coupling betwen excited electronic states of methanol probed by time-resolved photoelectron spectroscopy. *J. Chem. Phys.* **2019**, *150*, 114301.

[179] Cheng, B.-M.; Bahou, M.; Chen, W.-C.; Yui, C.-H.; Lee, Y.-P.; Lee, L. C. Experimental and theoretical studies on vacuum ultraviolet absorption cross sections and photodissociation of CH3OH, CH3OD, CD3OH and CD3OD. *J. Chem. Phys.* **2002**, *117*, 1633–1640.

[180] Ashfold, M. N. R.; King, G. A.; Murdock, D.; Nix, M. G. D.; Oliver, T. A. A.; Sage, A. G. $\pi\sigma^*$ excited states in molecular photochemistry. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1218–1238.

[181] Hockey, E. K.; McLane, N.; Martí, C.; Duckett, L.; Osborn, D. L.; Dodson, L. G. Direct Observation of Gas-Phase Hydroxymethylene: Photoionization and Kinetics Resulting from Methanol Photodissociation. *J. Am. Chem. Soc.* **2024**, *146*, 14416–14421.

[182] Lee, S.-H.; Lee, H.-I.; Lee, Y. T. Distributions of angular anisotropy and kinetic energy of productsfrom the photodissociation of methanol at 157 nm. *J. Chem. Phys.* **2004**, *121*, 11053–11059.

[183] Chen, Z.; Eppink, A. T. J. B.; Jiang, B.; Groenenboom, G. C.; Yang, X.; H., P. D. Product pair correlation in CH_3OH photodissociation at 157 nm: the OH + CH_3 channel. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2350–2355.

[184] Hutchison, J. M.; Holiday, R. J.; Bach, A.; Hsieh, S.; Crim, F. F. Action spectroscopy and photodissociation of vibrationally excited methanol. *J. Phys. Chem.* **2004**, *108*, 8115–8118.

[185] Kayanuma, M.; Shoji, M.; Furuya, K.; Aikawa, Y.; Umemura, M.; Shigeta, Y. Theoretical study of the photodissociation reaction of methanol. *Chem. Phys. Lett.* **2019**, *714*, 137–142.

[186] Qu, C.; Bowman, J. M. Full-Dimensional, ab initio potential energy surface for CH3OH to CH3+OH. *Mol. Phys.* **2013**, *111*, 1964–1971.

[187] Houston, P. L.; Qu, C.; Fu, B.; Bowman, J. M. Calculations of Dissociation Dynamics of CH3OH on a Global Potential Energy Surface Reveal the Mechanism for the Formation of HCOH; Roaming Plays a Role. *J. Phys. Chem. Lett.* **2024**, *15*, 9994–10000.

[188] Buenker, R. J.; Olbrich, G.; Schuchmann, H.-P.; L., S. B.; von Sonntag, C. Photolysis of methanol at 185 nm. Quantum Mechanical calculations and product study. *J. Am. Chem. Soc.* **1984**, *106*, 4362–4368.

[189] Shao, K.; Fu, B.; Zhang, D. H. A global full-dimensional potential energy surface and quasiclassical trajectory study of the O(1D) + CH4 multichannel reaction. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24098–24107.

[190] Yang, J.; Shao, K.; Zhang, D.; Shuai, Q.; Fu, B.; Zhang, D. H.; Yang, X. Trapped Abstraction in the O(1D) + CHD3 → OH + CD3 Reaction. *J. Phys. Chem. Lett.* **2014**, *5*, 3106–3111.

[191] Cigrang, L. L. E.; Worth, G. A. Modeling Photodissociation: Quantum Dynamics Simulations of Methanol. *J. Phys. Chem. A* **2024**, *128*, 7546–7557.

[192] Li Manni, G.; Fdez. Galván, I.; Alavi, A.; Aleotti, F.; Aquilante, F.; Autschbach, J.; Avagliano, D.; Baiardi, A.; Bao, J. J.; Battaglia, S.; Birnoschi, L.; Blanco-González, A.; Bokarev, S. I.; Broer, R.; Cacciari, R.; Calio, P. B.; Carlson, R. K.; Carvalho Couto, R.; Cerdán, L.; Chibotaru, L. F.; Chilton, N. F.; Church, J. R.; Conti, I.; Coriani, S.; Cuéllar-Zuquin, J.; Daoud, R. E.; Dattani, N.; Decleva, P.; de Graaf, C.; Delcey, M. G.; De Vico, L.; Dobrautz, W.; Dong, S. S.; Feng, R.; Ferré, N.; Filatov(Gulak), M.; Gagliardi, L.; Garavelli, M.; González, L.; Guan, Y.; Guo, M.; Hennefarth, M. R.; Hermes, M. R.; Hoyer, C. E.; Huix-Rotllant, M.; Jaiswal, V. K.; Kaiser, A.; Kaliakin, D. S.; Khamesian, M.; King, D. S.; Kochetov, V.; Krośnicki, M.; Kumaar, A. A.; Larsson, E. D.; Lehtola, S.; Lepetit, M.-B.; Lischka, H.; López Ríos, P.; Lundberg, M.; Ma, D.; Mai, S.; Marquetand, P.; Merritt, I. C. D.; Montorsi, F.; Märchen, M.; Nenov, A.; Nguyen, V. H. A.; Nishimoto, Y.; Oakley, M. S.; Olivucci, M.; Oppel, M.; Padula, D.; Pandharkar, R.; Phung, Q. M.; Plasser, F.; Raggi, G.; Rebolini, E.; Reiher, M.; Rivalta, I.; Roca-Sanjuán, D.; Romig, T.; Safari, A. A.; Sánchez-Mansilla, A.; Sand, A. M.; Schapiro, I.; Scott, T. R.; Segarra-Martí, J.; Segatta, F.; Sergentu, D.-C.; Sharma, P.; Shepard, R.; Shu, Y.; Staab, J. K.; Straatsma, T. P.; Sørensen, L. K.; Tenorio, B. N. C.; Truhlar, D. G.; Ungur, L.; Vacher, M.; Veryazov, V.; Voß, T. A.; Weser, O.; Wu, D.; Yang, X.; Yarkony, D.; Zhou, C.; Zobel, J. P.; Lindh, R. The OpenMolcas Web: A Community-Driven Approach to Advancing Computational Chemistry. *J. Chem. Theory Comput.* **2023**, *19*, 6933–6991, PMID: 37216210.

[193] Markmann, A.; Worth, G. A.; Cederbaum, L. S. Allene and pentatetraene cations as models for intramolecular charge transfer: vibronic coupling Hamiltonian and conical intersections. *J. Chem. Phys.* **2005**, *122*, 144320.

[194] Worth, G. A.; Welch, G.; Paterson, M. J. Wavepacket dynamics study of $Cr(CO)_5$ after formation by photodissociation: relaxation through an $(E \oplus A) \otimes E$ Jahn-Teller conical intersection. *Mol. Phys.* **2006**, *104*, 1095–1105.

[195] Wang, F.; Neville, S. P.; Wang, R.; Worth, G. A. Quantum Dynamics Study of Photoexcited Aniline. *J. Phys. Chem. A* **2013**, *117*, 7298–7307.

[196] Neville, S. P.; Worth, G. A. A reinterpretation of the electronic spectrum of pyrrole: A quantum dynamics study. *J. Chem. Phys.* **2014**, *140*, 034317–13.

[197] Bray, J. K.; Weiss, D. R.; Levitt, M. Optimized Torsion-Angle Normal Modes Reproduce Conformational Changes More Accurately Than Cartesian Modes. *Biophys. J.* **2011**, *101*, 2966–2969.

[198] Prlj, A.; Ibele, L. M.; Marsili, E.; Curchod, B. F. E. On the Theoretical Determination of Photolysis Properties for Atmospheric Volatile Organic Compounds. *J. Phys. Chem. Lett.* **2020**, *11*, 5418–5425.

[199] Nikoobakht, B.; Köppel, H.; Gindensperger, E.; Cederbaum, L. S. Efficient computation of adiabatic electronic populations in multi-mode vibronic systems: Theory, implementation, and application. *J. Chem. Phys.* **2012**, *137*, 114110.

[200] Allan, C. S. M.; Lasorne, B.; Worth, G. A.; Robb, M. A. A straightforward method of analysis for direct quantum dynamics: Application to the photochemistry of a model cyanine. *J. Phys. Chem. A* **2010**, *114*, 8713–8729.

[201] Roos, B. O.; Andersson, K. Multiconfigurational perturbation theory with level shift - the Cr_2 potential revisited. *Chem. Phys. Lett.* **1995**, *245*, 215–223.

[202] Kozma, B.; Tajti, A.; Demoulin, B.; Izsák, R.; Nooijen, M.; Szalay, P. G. A New Benchmark Set for Excitation Energy of Charge Transfer States: Systematic Investigation of Coupled Cluster Type Methods. *J. Chem. Theory Comput.* **2020**, *16*, 4213–4225.

[203] Rishi, V.; Perera, A.; Nooijen, M.; Bartlett, R. J. Excited states from modified coupled cluster methods: Are they any better than EOM CCSD? *J. Chem. Phys.* **2017**, *146*, 144104.

[204] Battaglia, S.; Lindh, R. On the role of symmetry in XDW-CASPT2. *J. Chem. Phys.* **2021**, *154*, 034102.

[205] Shiozaki, T.; Győrffy, W.; Celani, P.; Werner, H.-J. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *J. Chem. Phys.* **2011**, *135*, 081106.

[206] Battaglia, S.; Lindh, R. Extended Dynamically Weighted CASPT2: The Best of Two Worlds. *J. Chem. Theory Comput.* **2020**, *16*, 1555–1567.

[207] Nakatani, N. *Matrix Product States and Density Matrix Renormalization Group Algorithm*; Elsevier, 2018.

[208] Hammond, B. L.; Lester, W. A.; Reynolds, P. J. *Monte Carlo Methods in Ab Initio Quantum Chemistry*; WORLD SCIENTIFIC, 1994.

[209] Toulouse, J.; Assaraf, R.; Umrigar, C. J. *Electron Correlation in Molecules - ab initio Beyond Gaussian Quantum Chemistry*; Elsevier, 2016; pp 285–314.

[210] Izsák, R.; Ivanov, A. V.; Blunt, N. S.; Holzmann, N.; Neese, F. Measuring Electron Correlation. The Impactof Symmetry and Orbital Transformations. *J. Chem. Theory Comput.* **2023**, *19*, 703–2720.

[211] Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference Character for 3d Transition-Metal-ContainingMolecules. *J. Chem. Theory Comput.* **2012**, *8*, 460–468.

[212] Ruedenberg, K.; Atchity, G. J. A quantum chemical determination of diabatic states. *J. Chem. Phys.* **1993**, *99*, 3799–3803.

[213] Atchity, G. J.; Ruedenberg, K. Determination of diabatic states through enforcement of configurational uniformity. *Theo. Chem. Acc.* **1997**, *97*, 47–58.

[214] Nakamura, H.; Truhlar, D. G. The direct calculation of diabatic states based on configurational uniformity. *Chem. Phys.* **2001**, *115*, 10353–10372.

[215] Nakamura, H.; Truhlar, D. G. Direct diabatization of electronic states by the fourfold way. II. Dynamical correlation and rearrangement processes. *J. Chem. Phys.* **2002**, *117*, 5576–5593.

[216] Nakamura, H.; Truhlar, D. G. Extension of the fourfold way for calculation of global diabatic potential energy surfaces of complex, multiarrangement, non-Born-Oppenheimer systems: Application to HNCO (S0,S1)). *J. Chem. Phys.* **2003**, *118*, 6816–6829.

[217] Avila Ferrer, F. J.; Cerezo, J.; Stendardo, E.; Improta, R.; Santoro, F. Insights for an Accurate Comparison of Computational Data to Experimental Absorption and Emission Spectra: Beyond the Vertical Transition Approximation. *J. Chem. Theory Comput.* **2013**, *9*, 2072–2082.

[218] Cigrang, L. L. E.; Green, J. A.; Gómez, S.; Cerezo, J.; Improta, R.; Prampolini, G.; Santoro, F.; Worth, G. A. Non-adiabatic direct quantum dynamics using force fields: Toward solvation. *J. Chem. Phys.* **2024**, *160*, 174120.

[219] Shi, G.; Song, J.; Tian, P. Theoretical study of the reactions of triplet oxygen atom with 1-propanol and 2-propanol. *Chem. Phys. Lett.* **2021**, *781*, 138990.

[220] Rigby, M.; Montzka, S. A.; Prinn, R. G.; White, J. W. C.; Young, D.; O'Doherty, S.; Lunt, M. F.; Ganesan, A. L.; Manning, A. J.; Simmonds, P. G.; Salameh, P. K.; Harth, C. M.; Mühle, J.; Weiss, R. F.; Fraser, P. J.; Steele, L. P.; Krummel, P. B.; McCulloch, A.; Park, S. Role of atmospheric oxidation in recent methane growth. *Proc. Natl. Acad. Sci.* **2017**, *114*, 5373–5377.

[221] Cheng, C.-H.; Redfern, S. A. T. Impact of interannual and multidecadal trends on methane-climate feedbacks and sensitivity. *Nat. Commun.* **2022**, *13*, 3592.

[222] Bates, K. H.; Jacob, D. J.; Wang, S.; Hornbrook, R. S.; Apel, E. C.; Kim, M. J.; Millet, D. B.; Wells, K. C.; Chen, X.; Brewer, J. F.; Ray, E. A.; Commane, R.; Diskin, G. S.; Wofsy, S. C. The Global Budget of Atmospheric Methanol: New Constraints on Secondary, Oceanic, and Terrestrial Sources. *J. Geophys. Res. Atmos.* **2021**, *126*, 1–23.

[223] Satta, M.; Catone, D.; Castrovilli, M. C.; Nicolanti, F.; Cartoni, A. $CH_3OH^{\bullet+}$ + CH4 Reaction: Mechanistic Insights and Reaction Rates for Astrochemical and Atmospheric Environments. *Molecules* **2025**, *30*, 1029.

[224] Curchod, B. F. E.; Orr-Ewing, A. J. Perspective on Theoretical and Experimental Advances in Atmospheric Photochemistry. *J. Phys. Chem. A* **2024**, *128*, 6613–6635.

[225] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*.

[226] Mennucci, B. Polarizable continuum model. *Comput. Mol. Sci.* **2012**, *2*.

[227] Caricato, M. Coupled cluster theory with the polarizable continuum model of solvation. *Int J Quantum Chem.* **2019**, *119*, e25710.

[228] Cerezo, J.; Liu, Y.; Lin, N.; Zhao, X.; Improta, R.; Santoro, F. Mixed Quantum/Classical Method for Nonadiabatic Quantum Dynamics in Explicit Solvent Models: The $\pi\pi^*/n\pi^*$ Decay of Thymine in Water as a Test Case. *J. Chem. Theory Comput.* **2018**, *14*, 820–832.

[229] Green, J. A.; Gómez, S.; Worth, G.; Santoro, F.; Improta, R. Solvent Effects on Ultrafast Charge Transfer Population: Insights from the Quantum Dynamics of Guanine-Cytosine in Chloroform. *Chem. Eur. J.* **2022**, *28*, e202201731.

[230] Segalina, A.; Aranda, D.; Green, J. A.; Cristino, V.; Caramori, S.; Prampolini, G.; Pastore, M.; Santoro, F. How the Interplay among Conformational Disorder, Solvation, Local, and Charge-Transfer Excitations Affects the Absorption Spectrum and Photoinduced Dynamics of Perylene Diimide Dimers: A Molecular Dynamics/Quantum Vibronic Approach. *J. Chem Theory Comput.* **2022**, *18*, 3718–3736, PMID: 35377648.

[231] Honig, B.; Karplus, M. Implications of Torsional Potential of Retinal Isomers for Visual Excitation. *Nature* **1971**, *229*, 558–560.

[232] Warshel, A.; Karplus, M. Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization. *J. Am. Chem. Soc.* **1972**, *94*, 5612–5625.

[233] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227.

[234] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

[235] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678–5796, PMID: 25853797.

[236] Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Adding Explicit Solvent Molecules to Continuum Solvent Calculations for the Calculation of Aqueous Acid Dissociation Constants. *J. Phys. Chem. A* **2006**, *110*, 2493–2499, PMID: 16480309.

[237] Pliego Jr, J. R.; Riveros, J. M. Hybrid discrete-continuum solvation methods. *WIREs Comp. Mol. Sci.* **2020**, *10*, e1440.

[238] Santoro, F.; Green, J. A.; Martinez-Fernandez, L.; Cerezo, J.; Improta, R. Quantum and semiclassical dynamical studies of nonadiabatic processes in solution: achievements and perspectives. *Phys. Chem. Chem. Phys.* **2021**, *23*, 8181–8199.

[239] Lee, M. K.; Huo, P.; Coker, D. F. Semiclassical Path Integral Dynamics: Photosynthetic Energy Transfer with Realistic Environment Interactions. *Annu. Rev. Phys. Chem.* **2016**, *67*, 639–668.

[240] Boggio-Pasqua, M.; Robb, M. A.; Groenhof, G. Hydrogen Bonding Controls Excited-State Decay of the Photoactive Yellow Protein Chromophore. *J. Am. Chem. Soc.* **2009**, *131*, 13580–13581.

[241] Ruckenbauer, M.; Barbatti, M.; Sellner, B.; Muller, T.; Lischka, H. Azomethane: Nonadiabatic Photodynamical Simulations in Solution. *J. Phys. Chem. A* **2010**, *114*, 12585–12590.

[242] Guo, X.; Zhao, Y.; Cao, Z. A QM/MM MD insight into photodynamics of hypoxanthine: distinct nonadiabatic decay behaviors between keto-N7H and keto-N9H tautomers in aqueous solution. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15381–15388.

[243] Cusati, T.; Granucci, G.; Persico, M. Photodynamics and Time-Resolved Fluorescence of Azobenzene in Solution: A Mixed Quantum-Classical Simulation. *J. Am. Chem. Soc.* **2011**, *133*, 5109–5123.

[244] Ibele, L. M.; Sánchez-Murcia, P. A.; Mai, S.; Nogueira, J. J.; González, L. Excimer Intermediates en Route to Long-Lived Charge-Transfer States in Single-Stranded Adenine DNA as Revealed by Nonadiabatic Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 7483–7488.

[245] Polonius, S.; Zhuravel, O.; Bachmair, B.; Mai, S. LVC/MM: A Hybrid Linear Vibronic Coupling/Molecular Mechanics Model with Distributed Multipole-Based Electrostatic Embedding for Highly Efficient Surface Hopping Dynamics in Solution. *J. Chem. Theory Comput.* **2023**, *19*, 7171–7186.

[246] Vandaele, E.; Mališ, M.; Luber, S. The ΔSCF method for non-adiabatic dynamics of systems in the liquid phase. *J. Chem. Phys.* **2022**, *156*, 130901.

[247] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

[248] Middleton, C. T.; de La Harpe, K.; Su, C.; Law, Y. K.; Crespo-Hernández, C. E.; Kohler, B. DNA excited-state dynamics: from single bases to the double helix. *Annu. Rev. Phys. Chem.* **2009**, *60*, 217–239.

[249] Barbatti, M.; Borin, A.; Ullrich, S. *Photoinduced Phenomena in Nucleic Acids I: Nucleobases in the Gas Phase and in Solvents*; Topics in Current Chemistry; Springer International Publishing, 2015.

[250] Improta, R.; Santoro, F.; Blancafort, L. Quantum Mechanical Studies on the Photophysics and the Photochemistry of Nucleic Acids and Nucleobases. *Chem. Rev.* **2016**, *116*, 3540–3593, PMID: 26928320.

[251] Chakraborty, P.; Liu, Y.; Weinacht, T.; Matsika, S. Effect of dynamic correlation on the ultrafast relaxation of uracil in the gas phase. *Faraday Discuss.* **2021**, *228*, 266–285.

[252] Improta, R.; Barone, V.; Lami, A.; Santoro, F. Quantum Dynamics of the Ultrafast $\pi\pi*/n\pi*$ Population Transfer in Uracil and 5-Fluoro-Uracil in Water and Acetonitrile. *J. Phys. Chem. B* **2009**, *113*, 14491–14503, PMID: 19780519.

[253] Bezabih, M. S.; Kaliakin, D. S.; Blanco-González, A.; Barneschi, L.; Tarnovsky, A. N.; Olivucci, M. Comparative Study of Uracil Excited-State Photophysics in Water and Acetonitrile via RMS-CASPT2-Driven Quantum-Classical Trajectories. *J. Phys. Chem. B* **2023**, *127*, 10871–10879, PMID: 38055701.

[254] Cacelli, I.; Prampolini, G. Parametrization and Validation of Intramolecular Force Fields Derived from DFT Calculations. *J. Chem. Theory Comput.* **2007**, *3*, 1803.

[255] Cerezo, J.; Prampolini, G.; Cacelli, I. Developing accurate intramolecular force fields for conjugated systems through explicit coupling terms. *Theor. Chem. Acc.* **2018**, *137*, 80.

[256] Barone, V.; Cacelli, I.; De Mitri, N.; Licari, D.; Monti, S.; Prampolini, G. Joyce and Ulysses: integrated and user-friendly tools for the parameterization of intramolecular force fields from quantum mechanical data. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3736–3751.

[257] Shalashilin, D. V.; Burghardt, I. Gaussian-based techniques for quantum propagation from the time-dependent variational principle: Formulation in terms of trajectories of coupled classical and quantum variables. *J. Chem. Phys.* **2008**, *129*, 84104–84109.

[258] Mathews, J.; Walker, R. *Mathematical Methods of Physics*; W. A. Benjamin: New York, 1990.

[259] Martínez, T. J.; Levine, R. D. Non-adiabatic molecular dynamics: Split-operator multiple spawning with applications to photodissociation. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 941–947.

[260] Ben-Nun, M.; Martínez, T. J. *Ab Initio* Quantum Molecular Dynamics. *Adv. Chem. Phys.* **2002**, *121*, 439–512.

[261] Zurek, W. H. Preferred States, Predictability, Classicality and the Environment-Induced Decoherence. *Prog. Theor. Phys.* **1993**, *89*, 281–312.

[262] Prezhdo, O. V.; Rossky, P. J. Relationship between Quantum Decoherence Times and Solvation Dynamics in Condensed Phase Chemical Systems. *Phys. Rev. Lett.* **1998**, *81*, 5294–5297.

[263] Schwartz, B. J.; Bittner, E. R.; Prezhdo, O. V.; Rossky, P. J. Quantum decoherence and the isotope effect in condensed phase nonadiabatic molecular dynamics simulations. *J. Chem. Phys.* **1996**, *104*, 5942–5955.

[264] Xie, W.; Sapunar, M.; Došlić, N.; Sala, M.; Domcke, W. Assessing the performance of trajectory surface hopping methods: Ultrafast internal conversion in pyrazine. *J. Chem. Phys.* **2019**, *150*, 154119.

[265] Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.

[266] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

[267] Santoro, F.; Barone, V.; Gustavsson, T.; Improta, R. Solvent Effect on the Singlet Excited-State Lifetimes of Nucleic Acid Bases: A Computational Study of 5-Fluorouracil and Uracil in Acetonitrile and Water. *J. Am. Chem. Soc.* **2006**, *128*, 16312–16322.

[268] Kroes, G.-J.; Díaz, C. Quantum and classical dynamics of reactive scattering of H_2 from metal surfaces. *Chem. Soc. Rev.* **2016**, *45*, 3658–3700.

[269] Hernández, F. J.; Crespo-Otero, R. Modeling Excited States of Molecular Organic Aggregates for Optoelectronics. *Annu. Rev. Phys. Chem.* **2023**, *74*, 547–571.

[270] do Casal, T. M.; Toldo, J. M.; Pinheiro Jr, M.; Barbatti, M. Fewest switches surface hopping with Baeck-An couplings. *Open Research Europe* **2022**, *1*, 49.

[271] Zhao, X.; Merritt, I. C. D.; Lei, R.; Shu, Y.; Jacquemin, D.; Zhang, L.; Xu, X.; Vacher, M.; Truhlar, D. G. Nonadiabatic Coupling in Trajectory Surface Hop-

ping: Accurate Time Derivative Couplings by the Curvature-Driven Approximation. *J. Chem. Theory Comput.* **2023**, *19*, 6577–6588.

[272] Shu, Y.; Zhang, L.; Chen, X.; Sun, S.; Huang, Y.; Truhlar, D. G. Nonadiabatic Dynamics Algorithms with Only Potential Energies and Gradients: Curvature-Driven Coherent Switching with Decay of Mixing and Curvature-Driven Trajectory Surface Hopping. *J. Chem. Theory Comput.* **2022**, *18*, 1320–1328.

[273] Li, Z.; Hernández, F. J.; Salguero, C.; Lopez, S. A.; Crespo-Otero, R.; Li, J. Machine learning photodynamics decode multiple singlet fission channels in pentacene crystal. *Nat. Commun.* **2025**, *16*, 1194.

[274] Prediction Challenge: Cyclobutanone Photochemistry. `https://pubs.aip.org/jcp/collection/16531/Prediction-Challenge-Cyclobutanone-Photochemistry`, Accessed: 2025-01-17.

[275] Cigrang, L. L. E.; Curchod, B. F. E.; Ingle, R. A.; Kelly, A.; Mannouch, J. R.; Accomasso, D.; Alijah, A.; Barbatti, M.; Chebbi, W.; Došlić, N.; Eklund, E. C.; Fernandez-Alberti, S.; Freibert, A.; González, L.; Granucci, G.; Hernández, F. J.; Hernández-Rodríguez, J.; Jain, A.; Janoš, J.; Kassal, I.; Kirrander, A.; Lan, Z.; Larsson, H. R.; Lauvergnat, D.; Dé, B. L.; Lee, Y.; Maitra, N. T.; Min, S. K.; Peláez, D.; Picconi, D.; Raucci, U.; Qiu, Z.; Robertson, P.; Gil, E. S.; Sapunar, M.; Schürger, P.; Sinnott, P.; Tretiak, S.; Tikku, A.; Vindel-Zandbergen, P.; Worth, G. A.; Agostini, F.; Gómez, S.; Ibele, L. M.; Prlj, A. Roadmap for Molecular Benchmarks in Nonadiabatic Dynamics. *J. Phys. Chem.* **2025**, Under review (manuscript available at `https://arxiv.org/abs/2502.14569`).

[276] Ibele, L. M.; Curchod, B. F. E. A molecular perspective on Tully models for nonadiabatic dynamics. *Phys. Chem. Chem. Phys.* **2020**, *22*, 15183–15196.

[277] Oprea, T. I.; Bologa, C. Molecular Complexity: You Know It When You See It. *J. Med. Chem.* **2023**, *66*, 12710–12714.

[278] Bertz, S. H. The first general index of molecular complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599–3601.

[279] Speight, J. G. *Biomass Processes and Chemicals*; Elsevier, 2022; Chapter 9, pp 293–353.

[280] Okoye-Chine, C. G.; Otun, K.; Shiba, N.; Rashama, C.; Ugwu, S. N.; Onyeaka, H.; Okeke, C. T. Conversion of carbon dioxide into fuels-A review. *J. CO2 Util.* **2022**, *62*, 102099.

[281] Farrell, R. E. *RNA Methodologies*; Elsevier, 2005; Chapter 14, pp 317–332.

[282] Wu, Y.; Huo, J.; Yang, G.; Wang, Y.; Wang, L.; Wu, S.; Yao, L.; Fu, Q.; Wang, L. Measurement report: Production and loss of atmospheric formaldehyde at a suburban site of Shanghai in summertime. *Atmos. Chem. Phys.* **2023**, *23*, 2997–3014.

[283] de Gouw, J.; Farmer, D. Cloud droplets aid the production of formic acid in the atmosphere. *Nature* **2021**, *593*, 198–199.

[284] Solomon, P. M. Interstellar molecules. *Physics Today* **1973**, *26*, 32–40.

[285] Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823–823.

[286] Ruckenbauer, M.; Barbatti, M.; Müller, T.; Lischka, H. Nonadiabatic Photodynamics of a Retinal Model in Polar and Nonpolar Environment. *J. Phys. Chem. A* **2013**, *117*, 2790–2799.

[287] Bourne-Worster, S.; Worth, G. A. Quantum dynamics of excited state proton transfer in green fluorescent protein. *J. Chem. Phys.* **2024**, *160*, 065102.

[288] Janoš, J.; Slavíček, P. What Controls the Quality of Photodynamical Simulations? Electronic Structure Versus Nonadiabatic Algorithm. *J. Chem. Theory Comput.* **2023**, *19*, 8273–8284.

[289] Mai, S.; Atkins, A. J.; Plasser, F.; González, L. The Influence of the Electronic Structure Method on IntersystemCrossing Dynamics. The Case of Thioformaldehyde. *J. Chem. Theory Comput.* **2019**, *15*, 3470–3480.

[290] Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L.; Knizia, G. Automated Construction of Molecular Active Spaces from Atomic Valence Orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 4063–4078.

[291] Kaufold, B. W.; Chintala, N.; Pandeya, P.; Dong, S. S. Automated Active Space Selection with Dipole Moments. *J. Chem. Theory Comput.* **2023**, *19*, 2469–2483.

[292] Curtin, G.; Thomas, M.; Pieri, E. Autopylot: Pragmatic Excited State Electronic Structure Benchmarking. `https://chemrxiv.org/engage/chemrxiv/article-details/681b6a7c927d1c2e66607785`, 2025.

[293] Collins, M. A. Molecular potential-energy surfaces for chemical reaction dynamics. *Theor. Chem. Acc.* **2002**, *108*, 313–324.

[294] Godsi, O.; Evenhuis, C. R.; Collins, M. A. Interpolation of multidimensional diabatic potential energy matrices. *J. Chem. Phys.* **2006**, *125*, 104105.

[295] Godsi, O.; Collins, M. A.; Peskin, U. Quantum grow–a quantum dynamics sampling approach for growing potential energy surfaces and nonadiabatic couplings. *J. Chem. Phys.* **2010**, *132*, 124106.

[296] Kjønstad, E. F.; Koch, H. Understanding failures in electronic structure methods arising from the geometric phase effect. `https://doi.org/10.48550/arXiv.2411.08209`, 2024.

[297] Park, J. W. Single-State Single-Reference and Multistate Multireference Zeroth-Order Hamiltonians in MS-CASPT2 and Conical Intersections. *J. Chem. Theory Comput.* **2019**, *15*, 3960–3973.

[298] Sen, S.; Schapiro, I. A comprehensive benchmark of the XMS-CASPT2 method for the photochemistry of a retinal chromophore model. *Mol. Phys.* **2018**, *116*, 2571–2582.

[299] Tran, T.; Worth, G. A.; Robb, M. A. Coherent Excitation of the CH Stretching Vibrations in $C_2H_4^+$: The Role of the Derivative Coupling Studied by the Quantum Ehrenfest Method. *J. Comput. Chem.* **2025**, *46*.

[300] Barbatti, M.; Bondanza, M.; Crespo-Otero, R.; Demoulin, B.; Dral, P. O.; Granucci, G.; Kossoski, F.; Lischka, H.; Mennucci, B.; Mukherjee, S.; Pederzoli, M.; Persico, M.; Pinheiro Jr, M.; Pittner, J.; Plasser, F.; Sangiogo Gil, E.; Stojanovic, L. Newton-X Platform: New Software Developments for Surface

Hopping and Nuclear Ensembles. *J. Chem. Theory Comput.* **2022**, *18*, 6851–6865.

[301] Pinheiro, M.; de Oliveira Bispo, M.; Mattos, R. S.; Telles do Casal, M.; Chandra Garain, B.; Toldo, J. M.; Mukherjee, S.; Barbatti, M. ULaMDyn: enhancing excited-state dynamics analysis through streamlined unsupervised learning. *Digit. Discov.* **2025**, *4*, 666–682.

[302] Sittel, F.; Stock, G. Robust Density-Based Clustering To Identify Metastable Conformational States of Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 2426–2435.

[303] Sheong, F. K.; Silva, D.-A.; Meng, L.; Zhao, Y.; Huang, X. Automatic State Partitioning for Multibody Systems (APM): An Efficient Algorithm for Constructing Markov State Models To Elucidate Conformational Dynamics of Multibody Systems. *J. Chem. Theory Comput.* **2014**, *11*, 17–27.

[304] Schatz, G. C.; Kuppermann, A. Quantum mechanical reactive scattering for three-dimensional atom plus diatom systems. II. Accurate cross sections for H + H_2. *J. Chem. Phys.* **1976**, *65*, 4668–4692.

[305] Webster, F.; Light, J. C. Reactive scattering of H+H_2 and its isotopic analogs. *J. Chem. Phys* **1986**, *85*, 4744.

[306] Jäckle, A.; Meyer, H.-D. Reactive scattering using the multiconfiguration time-dependent Hartree approximation: General aspects and application to the collinear H+H_2 → H_2+H reaction. *J. Chem. Phys* **1995**, *102*, 5605–5615.

[307] Atkins, A. J.; González, L. Trajectory Surface-Hopping Dynamics Including Intersystem Crossing in $[Ru(bpy)3]^{2+}$. *J. Phys. Chem. Lett.* **2017**, *8*, 3840–3845.

[308] Capano, G.; Penfold, T. J.; Chergui, M.; Tavernelli, I. Photophysics of a copper phenanthroline elucidated by trajectory and wavepacket-based quantum dynamics: a synergetic approach. *Phys. Chem. Chem. Phys.* **2017**, *19*, 19590–19600.

[309] Virshup, A. M.; Chen, J.; Martínez, T. J. Nonlinear dimensionality reduction for nonadiabatic dynamics: The influence of conical intersection topography on population transfer rates. *J. Chem. Phys.* **2012**, *137*, 22A519.

[310] Belyaev, A. K.; Domcke, W.; Lasser, C.; Trigila, G. Nonadiabatic nuclear dynamics of the ammonia cation studied by surface hopping classical trajectory calculations. *J. Chem. Phys.* **2015**, *142*, 104307.

[311] Li, X.; Hu, D.; Xie, Y.; Lan, Z. Analysis of trajectory similarity and configuration similarity in on-the-fly surface-hopping simulation on multi-channel nonadiabatic photoisomerization dynamics. *J. Chem. Phys.* **2018**, *149*, 244104–17.

[312] Richings, G. W.; Habershon, S. Analyzing Grid-Based Direct Quantum Molecular Dynamics Using Non-Linear Dimensionality Reduction. *Molecules* **2021**, *26*, 7418.

[313] Jolliffe, I. *Principal Component Analysis*; Springer Series in Statistics; Springer, 2002.

[314] Coifman, R. R.; Lafon, S.; Lee, A. B.; Maggioni, M.; Nadler, B.; Warner, F.; Zucker, S. W. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7426–7431.

[315] Coifman, R. R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30.

[316] Balasubramanian, M.; Schwartz, E. L. The Isomap Algorithm and Topological Stability. *Science* **2002**, *295*, 7–7.

[317] Tenenbaum, J. B.; de Silva, V.; Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **2000**, *290*, 2319–2323.

[318] Borg, I.; Groenen, P. J. F. *Modern Multidimensional Scaling: Theory and Applications*; Springer Series in Statistics; Springer New York, 2005.

[319] van den Maaten, L.; Postma, E. O.; van den Herik, H. J. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 13.

[320] Chong, B. K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science* **2021**, *4*, 37–40.

[321] Steinley, D. K-means clustering: A half-century synthesis. *Br. J. Math. Stat. Psychol.* **2006**, *59*, 1–34.

[322] Liu, J.; Han, J. *Data clustering*; Chapman and Hall/CRC, 2018; pp 177–200.

[323] Tokuda, E. K.; Comin, C. H.; Costa, L. d. F. Revisiting agglomerative clustering. *Physica A* **2022**, *585*, 126433.

[324] Lukasová, A. Hierarchical agglomerative clustering procedure. *Pattern Recognit.* **1979**, *11*, 365–381.

[325] Rehman, S. U.; Asghar, S.; Fong, S.; Sarasvady, S. DBSCAN: Past, present and future. The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014). 2014; pp 232–238.

[326] Stewart , G.; Al-Khassaweneh, M. An Implementation of the HDBSCAN* Clustering Algorithm. *Appl. Sci.* **2022**, *12*, 2405.

[327] Liu, S.; Cao, S.; Suarez, M.; Goonetillek, E. C.; Huang, X. Multi-Level DBSCAN: A Hierarchical Density-Based Clustering Method for Analyzing Molecular Dynamics Simulation Trajectories. `https://www.biorxiv.org/content/early/2021/06/10/2021.06.09.447666`, 2021.

[328] González-Alemán, R.; Platero-Rochart, D.; Rodríguez-Serradet, A.; Hernández-Rodríguez, E. W.; Caballero, J.; Leclerc, F.; Montero-Cabrera, L. MDSCAN: RMSD-based HDBSCAN clustering of long molecular dynamics. *Bioinformatics* **2022**, *38*, 5191–5198.

[329] Acheson, K.; Kirrander, A. Automatic Clustering of Excited-State Trajectories: Application to Photoexcited Dynamics. *J. Chem. Theory Comput.* **2023**, *19*, 6126–6138.

[330] Kára, J.; Acheson, K.; Kirrander, A. DONKEY: A Flexible and Accurate Algorithm for Clustering. *J. Chem. Theory Comput.* **2025**, Preprint available at `https://doi.org/10.1021/acs.jctc.4c01750`.

[331] Eiter, T.; Mannila, H. *Computing Discrete Fréchet Distance*; techreport CD-TR 94/64, 1994.

[332] Alt, H.; M., G. Computing the Fréchet distance between two polygonal curves. *Int. J. Comput. Geom. Appl.* **1995**, *05*, 75–91.

[333] Seyler, S. L.; Kumar, A.; Thorpe, M. F.; Beckstein, O. Path Similarity Analysis: A Method for Quantifying Macromolecular Pathways. *PLoS Comput. Biol.* **2015**, *11*, e1004568.

[334] Brown, A.; Pradhan, E. Fitting potential energy surfaces to sum-of-products form with neural networks using exponential neurons. *J. Theo. Comp. Chem.* **2017**, *16*, 1730001.

[335] Bishop, C. *Pattern Recognition and Machine Learning*; Springer New York, 2006; Chapter 5, pp 225–290.

[336] Manzhos, S.; Carrington, T. Using neural networks to represent potential surfaces as sums of products. *J. Chem. Phys.* **2006**, *125*, 194105.

[337] Minns, R. S.; Parker, D. S. N.; Penfold, T. J.; Worth, G. A.; Fielding, H. H. Competing ultrafast intersystem crossing and internal conversion in the "channel 3" region of benzene. *Phys. Chem. Chem. Phys.* **2010**, *12*, 15607–15615.

[338] Reichardt, C.; Vogt, R. A.; Crespo-Hernández, C. E. On the origin of ultrafast nonradiative transitions in nitro-polycyclic aromatic hydrocarbons: Excited-state dynamics in 1-nitronaphthalene. *J. Chem. Phys.* **2009**, *131*, 224518.

[339] Marian, C. M. Spin-orbit coupling and intersystem crossing in molecules. *WIREs Comput. Mol. Sci.* **2012**, *2*, 187–203.

[340] Li, J.; Lopez, S. A. *Comprehensive Computational Chemistry*; Elsevier, 2024; pp 658–698.

[341] Gilbert, A.; Baggott, J. *Essentials of Molecular Photochemistry*; Blackwell Scientific Publications, 1991.

[342] Huang, T.-T.; Li, E. Y. Enhanced spin-orbit coupling driven by state mixing in organic molecules for OLED applications. *Org. Electron.* **2016**, *39*, 311–317.

[343] Mai, S.; Pollum, M.; Martínez-Fernández, L.; Dunn, N.; Marquetand, P.; Corral, I.; Crespo-Hernández, C. E.; González, L. The origin of efficient triplet state population in sulfur-substituted nucleobases. *Nat. Commun.* **2016**, *7*, 1:8.

[344] El-Sayed, M. A. Triplet state: Its radiative and nonradiative properties. *Acc. Chem. Res.* **1968**, *1*, 8–16.

[345] Zhang, L.; Shu, Y.; Sun, S.; Truhlar, D. G. Direct coherent switching with decay of mixing for intersystem crossing dynamics of thioformaldehyde: The effect of decoherence. *J. Chem. Phys.* **2021**, *154*, 094310.

[346] Guan, Y.; Xie, C.; Guo, H.; Yarkony, D. R. Toward a Unified Analytical Description of Internal Conversion and Intersystem Crossing in the Photodissociation of Thioformaldehyde. I. Diabatic Singlet States. *J. Chem. Theory Comput.* **2023**, *19*, 6414–6424.

[347] Moule, D. C.; Lim, E. C. Highly Varying Photophysical Properties of Thiocarbonyls: Validation of Fundamental Theoretical Concepts of Electronic Radiationless Transitions. *J. Phys. Chem. A* **2002**, *106*, 3072–3076.

[348] Calio, P. B.; Truhlar, D. G.; Gagliardi, L. Nonadiabatic Molecular Dynamics by Multiconfiguration Pair-Density Functional Theory. *J. Chem. Theory Comput.* **2022**, *18*, 614–622.

[349] Curchod, B. F.; Rauer, C.; Marquetand, P.; González, L.; Martínez, T. J. Communication: GAIMS - Generalized Ab Initio Multiple Spawning for both internal conversion and intersystem crossing processes. *J. Chem. Phys.* **2016**, *144*, 101102–6.

[350] Richings, G. W.; Habershon, S. Direct Grid-Based Nonadiabatic Dynamics on Machine-Learned Potential Energy Surfaces: Application to Spin-Forbidden Processes. *J. Phys. Chem. A* **2020**, *124*, 9299–9313.

[351] Malmqvist, P. Å.; Roos, B. O.; Schimmelpfennig, B. The restricted active space (RAS) state interaction approach with spin-orbit coupling. *Chem. Phys. Lett.* **2002**, *357*, 230–240.

[352] Clouthier, D. J.; Ramsay, D. A. The Spectroscopy of Formaldehyde and Thioformaldehyde. *Annu. Rev. Phys. Chem.* **1983**, *34*, 31–58.

[353] Kawasaki, M.; Kasatani, K.; Ogawa, Y.; Sato, H. Spectra and emission lifetimes of $H_2CS(\tilde{A}^1\ A^2)$. *Chem. Phys.* **1983**, *74*, 83–88.

[354] Bearpark, M. J.; Robb, M. A.; Schlegel, H. B. Direct method for the location of the lowest energy point on a potential surface crossing. *Chem. Phys. Lett.* **1994**, *223*, 269.

[355] Clouthier, D. J.; Huang, G.; Merer, A. J. A spectroscopic view of internal conversion in a small polyatomic molecule: Sub-Doppler intracavity dye laser spectroscopy of thioformaldehyde. *J. Chem. Phys.* **1992**, *97*, 1630–1637.

[356] Hachey, M. R. J.; Grein, F. The Spectroscopy of H_2CS: Out-of-Plane Potentials for Singlet States and Geometry Optimization of the $^1(\pi\pi^*)$, $^1(\sigma\pi^*)$,$^1(n^0\pi^{*2})$, and Rydberg States. *J. Mol. Spectrosc.* **1995**, *172*, 384–406.

[357] Cigrang, L. L. E.; Saha, S.; Worth, G. Including spin-orbit coupling in direct wavepacket dynamics. *In preparation*

[358] Su, H.; He, Y.; Kong, F.; Fang, W.; Liu, R. Photodissociation of formic acid. *J. Chem. Phys.* **2000**, *113*, 1891–1897.

[359] Lee, K. W.; Lee, K.-S.; Jung, K.-H.; Volpp, H.-R. The 212.8-nm photodissociation of formic acid: Degenerate four-wave mixing spectroscopy of the nascent $OH(X^2\Pi_i)$ radicals. *J. Chem. Phys.* **2002**, *117*, 9266–9274.

[360] de la Cerda, J.; Triana, J. F. Machine-learning potential energy surfaces implications in photodissociation process. *Chaos* **2025**, *35*, 043102.

[361] Li, S.; Xie, B.-B.; Yin, B.-W.; Liu, L.; Shen, L.; Fang, W.-H. Construction of Highly Accurate Machine Learning Potential Energy Surfaces for Excited-State Dynamics Simulations Based on Low-Level Data Sets. *J. Phys. Chem. A* **2024**, *128*, 5516–5524.

[362] Han, B.; Yu, K. Refining potential energy surface through dynamical properties via differentiable molecular simulation. *Nat. Commun.* **2025**, *16*, 816.

[363] Hu, D.; Xie, Y.; Li, X.; Li, L.; Lan, Z. Inclusion of Machine Learning Kernel Ridge Regression Potential Energy Surfaces in On-the-Fly Nonadiabatic Molecular Dynamics Simulation. *J. Phys. Chem. Lett.* **2018**, *9*, 2725–2732.

[364] Schmitz, G.; Godtliebsen, I. H.; Christiansen, O. Machine learning for potential energy surfaces: An extensive database and assessment of methods. *J. Chem. Phys.* **2019**, *150*, 244113.

[365] Girija, A. V.; Zeng, W.; Myers, W. K.; Kilbride, R. C.; Toolan, D. T. W.; Zhong, C.; Plasser, F.; Rao, A.; Bronstein, H. Singlet Fission in Pechmann Dyes: Planar Chromophore Design and Understanding. *J. Am. Chem. Soc.* **2024**, *146*, 18253–18261.

[366] Goran, M. Scientists are also writers. *J. Chem. Educ.* **1962**, *39*, 479.

[367] Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **1965**, *38*, 4.

[368] Sherry, Y.; Thompson, N. C. How Fast Do Algorithms Improve? [Point of View]. *Proc. IEEE* **2021**, *109*, 1768–1777.

[369] Whitfield, J. D.; Biamonte, J.; Aspuru-Guzik, A. Simulation of electronic structure Hamiltonians using quantum computers. *Mol. Phys.* **2011**, *109*, 735–750.

[370] Motta, M.; Rice, J. E. Emerging quantum computing algorithms for quantum chemistry. *WIREs Comput. Mol. Sci.* **2021**, *12*, e1580.

[371] Ollitrault, P. J.; Mazzola, G.; Tavernelli, I. Nonadiabatic Molecular Quantum Dynamics with Quantum Computers. *Phys. Rev. Lett.* **2020**, *125*, 260511.

[372] MacDonell, R. J.; Dickerson, C. E.; Birch, C. J. T.; Kumar, A.; Edmunds, C. L.; Biercuk, M. J.; Hempel, C.; Kassal, I. Analog quantum simulation of chemical dynamics. *Chem. Sci.* **2021**, *12*, 9794–9805.

[373] Deutsch, D. *The Beginning of Infinity*; Penguin Science, 2011.

Appendices

APPENDIX A

Modelling photodissociation

A.1 Coordinates & Frequencies

Table A.1: Cartesian coordinates (in Å) for methanol used in this work (optimised using CCSD// aug-cc-pVDZ).

Label	**X**	**Y**	**Z**
C1	-0.67980632	-0.02010217	0.00000040
O2	0.75677758	0.12639138	0.00000008
H3	-1.08527514	0.98693443	-0.00000046
H4	-1.02289486	-0.54602403	0.89222483
H5	-1.02289427	-0.54602363	-0.89222447
H6	1.15320100	-0.75360198	-0.00000038

Table A.2: Frequencies (in cm^{-1}) of the 12 normal modes of Methanol (CCSD//aug-cc-pVDZ).

Mode	**Symm.**	**Frequency**	**Description**
Q_1	A″	289.0272	Torsion
Q_2	A′	1026.9026	C–O stretch
Q_3	A′	1045.9601	CH_3 + OH6 rocking
Q_4	A″	1131.3388	H4/5–C–O angle bend
Q_5	A′	1372.1551	C–O–H6 angle bend
Q_6	A′	1437.1821	CH_3 symm. bend
Q_7	A″	1474.1371	H3 rocking
Q_8	A′	1487.4185	H4–C–H5 scissoring
Q_9	A′	3154.5375	CH_3 symm. stretch
Q_{10}	A″	3230.1449	C–H4/5 antisymm. stretch
Q_{11}	A′	3287.6298	C–H3 stretch
Q_{12}	A′	3838.8837	O–H6 stretch

A.2 CASPT2 intruder states

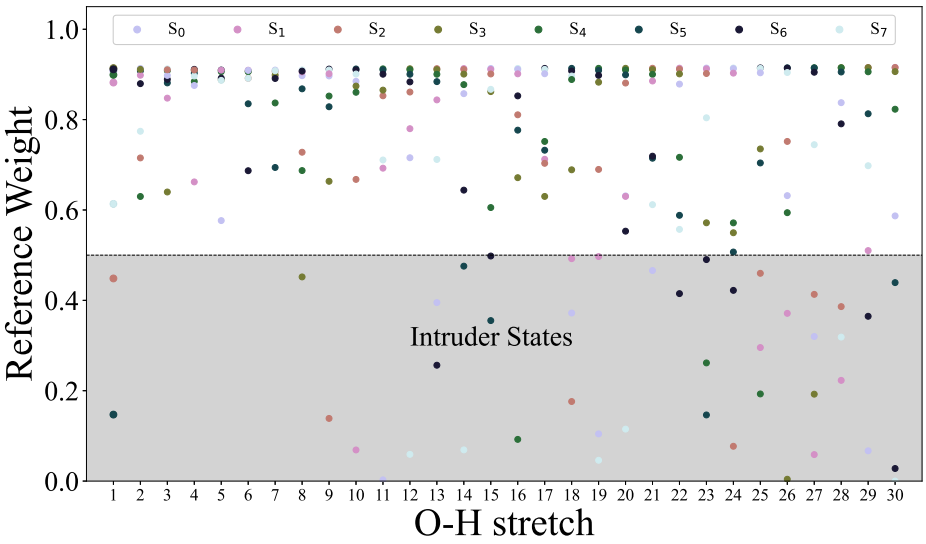


Figure A.1: The CASPT2 reference weights of each electronic state along the O-H stretching coordinate.

A.3 Local minimum of $\tilde{F}$ state

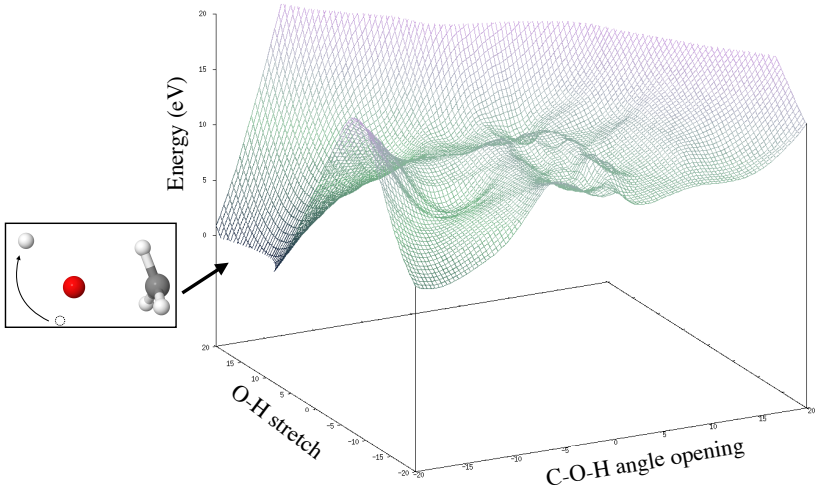


Figure A.2: PES of the diabatic $\tilde{F}$ state. The local minimum is indicated, along with the corresponding geometry.

APPENDIX B

Force Field Quantum Dynamics

B.1 Geometry and Frequencies

Table B.1: Cartesian coordinates (in Å) for Ura used in this work.

Label	**X**	**Y**	**Z**
C2	-1.23106	0.36160	0.00000
C6	0.05566	-1.70289	0.00000
C5	1.23697	-1.06563	0.00000
C4	1.27366	0.39276	0.00000
N3	0.00000	0.98501	0.00000
H6	-0.02805	-2.78354	0.00000
H5	2.17748	-1.59751	0.00000
O4	2.27178	1.08248	0.00000
O2	-2.29119	0.94775	0.00000
N1	-1.13733	-1.02276	0.00000
H1	-2.02185	-1.50914	0.00000
H3	-0.02233	1.99762	0.00000

Table B.2: Normal modes[a] of Uracil for the ground electronic state, point group = C_s. A' modes are in-plane whereas the A" modes correspond to out-of-plane motions. Frequencies in cm^{-1}

A' (*in-plane*)			A" (*out-of-plane*)		
Label	**Freq.**	**Description**	**Label**	**Freq.**	**Description**
Q_3	391.45	O_2–O_4 scissoring	Q_1	152.29	N_3 bend
Q_5	524.44	C_5–C_4–N_3 scissoring	Q_2	170.92	N_1–C_5 antisymm. bend
Q_6	549.63	C_5–H_5 wag	Q_4	401.02	C_6–H_6 wag
Q_7	566.70	N_1–H_{N1} wag	Q_8	569.46	N_1–H_{N1} wag
Q_{12}	786.44	breathing mode	Q_9	692.77	N_3–H_{N2} wag
Q_{14}	986.20	N_3–H_{N2} wag	Q_{10}	740.34	C_5–H_5 wag
Q_{16}	1003.99	ring bending	Q_{11}	772.79	C_2 bend
Q_{17}	1109.34	C_5–H_5 wag	Q_{13}	827.49	H_5–H_6 symm. wag
Q_{18}	1225.09	C_6–H_6 wag	Q_{15}	993.94	H_5–H_6 antisymm. wag
Q_{19}	1255.87	C_5–H_5 wag			
Q_{20}	1409.11	N_3–H_{N2} wag			
Q_{21}	1438.53	H_{N2}–H_5 scissoring			
Q_{22}	1448.46	N_1–H_{N1} wag			
Q_{23}	1536.60	N_1–H_{N1} wag			
Q_{24}	1732.47	C_6–C_5 stretch			
Q_{25}	1856.35	C_4–O_4 stretch			
Q_{26}	1886.58	C_2–O_2 stretch			
Q_{27}	3256.28	C_6–H_6 stretch			
Q_{28}	3294.66	C_5–H_5 stretch			
Q_{29}	3636.26	N_3–H_{N3}			
Q_{30}	3672.59	N_1–H_{N1}			

[a] Note that since a specific FF for each electronic state is derived, normal modes and frequencies are different for S_0, $n\pi^*$ and $\pi\pi^*$.

B.2 QVC_0 Hamiltonian Frequencies and Linear Parameters

Table B.3: Frequencies (A′ = in-plane, A″ = out-of-plane) and linear parameters of the 2-state vibronic coupling Hamiltonian of Uracil. n in $\kappa^{(n)}$ and $\lambda^{(n,n)}$ refers to the electronic state where 1 = $n\pi^*$ and 2 = $\pi\pi^*$. Highlighted in red are the 10 system modes used in the 10-D + 20-D partition of the nuclear basis. All units are eV.

Mode	Symmetry	ω	$\kappa^{(1)}$	$\kappa^{(2)}$	$\lambda^{(1,2)}$
Q_3	A′	0.0189	-0.0048	0.0363	
Q_5	A′	0.0212	0.0301	0.0300	
Q_6	A′	0.0497	0.0411	0.0097	
Q_7	A′	0.0706	0.0217	0.0846	
Q_{12}	A′	0.0859	-0.0202	0.0489	
Q_{14}	A′	0.0918	-0.0475	-0.0617	
Q_{16}	A′	0.0958	-0.0039	0.0709	
Q_{17}	A′	0.1026	-0.0194	-0.0032	
Q_{18}	A′	0.1232	-0.0188	-0.0966	
Q_{19}	A′	0.0485	-0.0281	-0.0982	
Q_{20}	A′	0.0650	-0.0026	-0.0931	
Q_{21}	A′	0.0681	-0.0515	0.0148	
Q_{22}	A′	0.0703	0.1289	0.0554	
Q_{23}	A′	0.0975	-0.1117	-0.0745	
Q_{24}	A′	0.1223	0.0480	0.1304	
Q_{25}	A′	0.1245	0.3388	0.1047	
Q_{26}	A′	0.1375	-0.0991	-0.0858	
Q_{27}	A′	0.1519	-0.0395	-0.0562	
Q_{28}	A′	0.1557	0.0006	0.0005	
Q_{29}	A′	0.1747	0.0320	0.0135	
Q_{30}	A′	0.1784	-0.0010	0.0255	
Q_1	A″	0.1796			0.0150
Q_2	A″	0.1905			-0.0060
Q_4	A″	0.2148			-0.0021
Q_8	A″	0.2302			0.0095
Q_9	A″	0.2339			0.0048
Q_{10}	A″	0.4037			-0.0110
Q_{11}	A″	0.4085			-0.0112
Q_{13}	A″	0.4508			0.0126
Q_{15}	A″	0.4553			0.0435

B.3 Intermolecular QMD-FF parameters

Table B.4: Intermolecular parameters obtained for Ura in different electronic states: σ (Å) and ϵ (kJ/mol) are the Lennard-Jones parameters. q (e^-) are the partial charges. Atom types are shown in Figure 4.5.

Atom type	σ	ϵ	q^{S_0}	q^{S_1}	q^{S_2}
C_2	3.400	0.3598	0.3878	0.3440	0.3925
C_6	3.400	0.3598	-0.1532	-0.4149	-0.2616
C_5	3.400	0.3598	-0.2183	-0.1532	-0.0733
C_4	3.400	0.3598	0.4345	0.1480	0.2721
N_3	3.250	0.7113	-0.2209	-0.1899	-0.2308
H_6	2.511	0.0628	0.2220	0.2414	0.2220
H_5	2.600	0.0628	0.1547	0.2069	0.1526
O_4	2.960	0.8786	-0.5273	-0.2114	-0.5335
O_2	2.960	0.8786	-0.5197	-0.5291	-0.4907
N_1	3.250	0.7113	-0.1525	-0.0577	-0.0571
H_1	1.069	0.0657	0.3077	0.2839	0.3154
H_3	1.069	0.0657	0.2851	0.3320	0.2925

APPENDIX C

AQUATICS documentation

1 Introduction

This guide describes the implementation and workflow of the AQUATICS procedure for performing Force Field Quantum Dynamics (FFQD). The method is implemented in the QUANTICS suite of quantum dynamics programs, and a familiarity with this package is assumed.

2 Prerequisites

2.1 Software

If QUANTICS is already installed, the version which contains the interface to GROMACS can be found in the "gromacs" branch of the QUANTICS git repository:

https://gitlab.com/quantics/quantics.git

To install this version, use:

`git pull origin gromacs`

The version of GROMACS that is currently used is the 2021.3 double precision version:

`gromacs/2021.3-dbl`

2.2 Database Files

The following named files are necessary to include in the database directory before initiating a DD-vMCG calculation using GROMACS on-the-fly:

- GROMACS topology files for each of the excited electronic states of the solute, named following the convention `S_`n`.top`, where n is the electronic state number.
- GROMACS input files named `input_md.mdp` and `input_nm.mdp`. `input_md.mdp` uses `integrator = md` and is used to calculate energies and forces, whilst `input_nm.mdp` uses `integrator = nm` and is used to calculate the Hessian. The `.mdp` files can be modified by the user as long as the integrator is the correct one and the keyword `nstfout = 1` is used in `input_md.mdp`.
- A `template.dat` file containing the energy gaps between the S_0 and S_n minima, and the vibronic couplings between the electronic states, written in Cartesian basis.
- A Gaussian *.log* file containing a ground state frequency calculation.

3 Tutorial

This section will outline the procedure that can be applied in principle to any molecule but in this case, Uracil is taken as an example. For a detailed discussion of the theory, methodology and Uracil results, see the paper *"Non-adiabatic direct*

quantum dynamics using force fields: Toward solvation", published in *JCP*. [1] All data can be found on Hartree in the following directory: `/home/uccalci/uracil/QD/`. A README file in this directory explains the contents.

3.1 Generating the Topology

Before starting the dynamics, one must generate appropriate GROMACS topology files for each electronic state of interest. This topology file contains all the specific interactions between the different types of atoms and includes the parameters for the various potential functions. To obtain these parameters with Quantum Mechanical accuracy, the program *Joyce* can be used to generate the `.top` files from *Gaussian* calculations. The documentation for *Joyce* can be found here: https://nextcloud.area.fi.cnr.it/index.php/s/pk7QB885FdwzXjW.

Part of the topology file for the first excited state of Uracil is shown in Fig. 1.

When the topology files are generated correctly from *Joyce*, the top part of the file should look similar to Fig. 1. Here are some important things to note:

- The names in the `[ atomtypes ]` section define the specific *type* of an atom and its associated parameters, they are not a unique atom *label*. For example, notice that `H6` and `H5` are both referring to hydrogen atoms but they have slightly different sigma parameters because they are not chemically equivalent in the molecule and hence are assigned a different type.
- The `[ moleculetype ]` section defines the name of the residue, by which GROMACS means the name of the molecule. In this case "UNK" refers to Unknown. This is a convention in GROMACS and will make it clear that the molecule is unknown to GROMACS so it should not try to find existing parameters for it. This is a default in *Joyce* and is also used in QUANTICS to generate new geometries so it is recommended to always use "UNK" as the residue name.
- In the `[ atoms ]` section, the residue name is listed again along with the atomtypes. Here, the atom *labels* are also defined in the `atom` column. These are unique for every atom and in this case, correspond to the `atomtypes` but this is not a requirement.

The rest of the topology file lists all the intramolecular interactions.

3.2 Setting up Quantics input

After the topology files are created, one needs to set up an appropriate QUANTICS input file. Each section is briefly discussed.

3.2.1 RUN-SECTION

This section is essentially the same as any other direct dynamics run, making sure that the propagation is done in normal modes (i.e. `direct = nmodes`).

```
 [ atomtypes ]
; name   mass       charge   ptype    sigma       epsilon
   C2  12.01100    0.0       A    3.39967e-01  3.59824e-01   ; FF: GAFF  c
   C6  12.01100    0.0       A    3.39967e-01  3.59824e-01   ; FF: GAFF  cc
   C5  12.01100    0.0       A    3.39967e-01  3.59824e-01   ; FF: GAFF  cd
   C4  12.01100    0.0       A    3.39967e-01  3.59824e-01   ; FF: GAFF  c
   N3  14.00670    0.0       A    3.25000e-01  7.11280e-01   ; FF: GAFF  n
   H6   1.00800    0.0       A    2.51055e-01  6.27600e-02   ; FF: GAFF  h4
   H5   1.00800    0.0       A    2.59964e-01  6.27600e-02   ; FF: GAFF  ha
   O4  15.99940    0.0       A    2.95992e-01  8.78640e-01   ; FF: GAFF  o
   O2  15.99940    0.0       A    2.95992e-01  8.78640e-01   ; FF: GAFF  o
   N1  14.00670    0.0       A    3.25000e-01  7.11280e-01   ; FF: GAFF  n
   Hn1  1.00800    0.0       A    1.06908e-01  6.56888e-02   ; FF: GAFF  hn
   Hn2  1.00800    0.0       A    1.06908e-01  6.56888e-02   ; FF: GAFF  hn

 [ moleculetype ]
; Name      nrexcl
  UNK         6

 [ atoms ]
;   nr   type  resnr residue  atom  cgnr charge  mass
    1     C2     1     UNK    C2      1   0.0     12.01100
    2     C6     1     UNK    C6      2   0.0     12.01100
    3     C5     1     UNK    C5      3   0.0     12.01100
    4     C4     1     UNK    C4      4   0.0     12.01100
    5     N3     1     UNK    N3      5   0.0     14.00670
    6     H6     1     UNK    H6      6   0.0      1.00800
    7     H5     1     UNK    H5      7   0.0      1.00800
    8     O4     1     UNK    O4      8   0.0     15.99940
    9     O2     1     UNK    O2      9   0.0     15.99940
    10    N1     1     UNK    N1      10  0.0     14.00670
    11    Hn1    1     UNK    Hn1     11  0.0      1.00800
    12    Hn2    1     UNK    Hn2     12  0.0      1.00800
```

Figure 1: The `[ atomtypes ]`, `[ moleculetype ]` and `[ atoms ]` section of the `.top` file for the S1 state of Uracil.

3.2.2 INITIAL-GEOMETRY-SECTION

The cartesian coordinates are given in Angstroms. The atom labels need to be defined as follows: `"element":"gmx_label"` where `"element"` is simply the elemental symbol of the atom, and `"gmx_label"` corresponds to the labels in the `[ atoms ]` section of the topology file. The atoms should be listed in the same order as in the topology file. For example:

```
INITIAL-GEOMETRY-SECTION
nstates = 2    init_state = 2
cartesian = angst
 C:C2                    -1.23106    0.36160    0.00000
 C:C6                     0.05566   -1.70289    0.00000
 C:C5                     1.23697   -1.06563    0.00000
...
```

3.2.3 DIRDYN-SECTION

This section is also not very different from a typical direct dynamics simulation except that `qcprogram` should be set to "gromacs" and `method` should be "lvc". The `dd_diab` keyword should also be set to "none". For the `subcmd` keyword, the path should point to the bash script that runs all the necessary GROMACS commands. This script is found in `$QUANTICS_DIR/bin/run_gromacs`. The script and all the commands are discussed in Sec. 4.2.

3.3 Hybrid calculations

To set up *hybrid* calculations, where different sets of GWPs will be propagated according to different equations of motion, two sections in the QUANTICS input need to be altered. Firstly, a `SBASIS-SECTION` is required, where one can partition the normal modes of the system in as many different sets as desired. For example, the 2-set calculation used for Uracil:

```
SBASIS-SECTION
v12,v13,v15,v07,v23,v25,v01,v20,v19,v24= 61
v03,v05,v02,v08,v22,v10,v14,v16,v04,v09,v17,v18,v06,v11,v21,v26,v27,v28,v29,v30= 41
gwp_type=frozen,normal
end-sbasis-section
```

Next, the `INTEGRATOR-SECTION` can be used to specify which set should obey which equation of motion. There are currently three options:

- `vmcg`: The fully coupled, variational EoMs are used:

$$i\dot{\mathbf{A}} = \mathbf{S}^{-1}(\mathbf{H} - i\boldsymbol{\tau})\mathbf{A} \tag{1}$$

$$i\dot{\boldsymbol{\Lambda}} = \mathbf{X} + \mathbf{C}^{-1}\mathbf{Y}_R \tag{2}$$

- `classical`: Higher order terms are omitted in the EoM for the linear Gaussian parameters, leaving only the classical terms:

$$i\dot{\boldsymbol{\xi}} = \mathbf{X} \tag{3}$$

- `saddle`: The classical approximation is used but additionally, the Saddle-point approximation is used to evaluate the following integrals:

$$\langle G_i | \hat{H}^{el} | G_j \rangle \approx \hat{H}^{el} \langle G_i | G_j \rangle \tag{4}$$

These options are used in combination with the `gwp_model` keyword, where `@m1` refers to the first set of modes, as per the `SBASIS-SECTION`. Here an example of a hybrid vMCG + classical calculations:

```
INTEGRATOR-SECTION
 VMF
 RK5 = 1.0d-7
 gwp_model@m1 = vmcg
 gwp_model@m2 = classical
end-integrator-section
```

3.4 Setting up the Database directory

Along with an appropriate input file, a database directory needs to be set up before starting the calculations. This directory needs to contain the following files:

1. input_md.mdp
2. input_nm.mdp
3. S_n.top (1 for every state n)
4. Gaussian *.log* file containing a ground-state frequency calculation
5. template.dat (see "template.dat file" section)

3.5 template.dat file

The `template.dat` file contains two sections:

1. The energy gaps (in a.u.) between the S_0 and S_n minima, which are written below the keyword `ener_offset`, one per line for each of the N states. For example:

```
    ener_offset
    0.167410945
    0.194406706

```

   The S_n minima are those utilised by Joyce to construct the excited state force fields.

2. The vibronic couplings (in a.u.) between electronic states n and m, in the Cartesian basis, written below a `coupling` keyword. Each line contains an x, y and z component for each atom, with the atomic order as specified in the QUANTICS input file geometry section. For multiple states, e.g. if $N = 4$,

then the couplings would be written in subsequent lines following the order S_1:S_2, S_1:S_3, S_1:S_4, S_2:S_3, S_2:S_4, S_3:S_4. In uracil (where there is only the S_1:S_2 coupling) this section is written as:

```
coupling
0.000000000000     0.000000000000    -0.003652510287
0.000000000000     0.000000000000    -0.007300153644
0.000000000000     0.000000000000     0.002978475657
...

```

3.5.1 Obtaining Vibronic Couplings

The vibronic couplings required in part 2 of the `template.dat` file described in the previous section were obtained with the Overdia[2] code, whose manual may be found here: www.iccom.cnr.it/en/overdia-en/. Of course other methods e.g. VCHam may be used to compute these couplings. As Overdia computes the couplings in the normal mode basis, a transformation to the Cartesian basis was required. This was achieved using the python script `nmgrad2cartgrad.py`. This script requires the following named files as input:

- `nmfreq.dat`: A list of the vibrational frequencies (in cm-1).
- `masses`: A list of the atomic masses (in amu), in the order of the geometry specification (n.b. this file is created automatically by the `prep.fc.e` utility code from Overdia).
- `Tmat.dat`: The normal mode matrix $T(i,j)$ in Cartesian coordinates (not mass weighted), where i indexes the Cartesian coordinate for each atom ($i = 1, 3 \times N_{\text{atom}}$), and j indexes the normal mode ($j = 1, 3 \times N_{\text{atom}} - 6$). The file is written as a column vector in which j is the faster index, i.e. the first $3 \times N_{\text{atom}} - 6$ entries are for the x coordinate of atom 1. This file can be obtained from the '`state`' file created by the `prep.fc.e` utility code in Overdia, by removing the first $3 \times N_{\text{atom}}$ entries (which correspond to the geometry) and the final $3 \times N_{\text{atom}} - 6$ entries (which correspond to the vibrational frequencies. Equivalently, it could be created directly from the frequency output of an electronic frequency calculation (e.g. in Gaussian it is where the '`Frequencies`' are defined).

Additionally, a file containing the vibronic couplings in the normal mode basis is required, and taken as a command line argument. The script then prints the vibronic couplings in Cartesian basis to the screen, which can be piped to an output file, e.g.

```
python nmgrad2cartgrad.py lambda12.inp > lambda12_cartgrad.out
```

3.6 Running Quantics

When all the steps described above are complete, a QUANTICS test can be run using the following command:

`quantics -mnd -t [input file]`

where the `-t` option specifies that a test run is to be performed. It should perform the GROMACS calculations once per state and then terminate. The database directory should now contain a `database.sql` file, containing the outputs of the GROMACS calculation, i.e. the energies, gradients and Hessians. There should also be one directory for every state, containing the raw GROMACS outputs. The database files can be checked to make sure everything ran correctly. If everything is working properly, QUANTICS can now be run to do a full Direct Dynamics calculation using the same command as before (but without `-t`):

`quantics -mnd [input file]`

4 Notes on Implementation

4.1 Quantics - Gromacs Interface

The interface is written in the `source/opfuncs/funcqchemmod.F90` as with the other quantum chemistry interfaces, with `ddgromacs`, `wrgromacs` and `rdgromacs` routines whose function is as follows:

- `ddgromacs`: Main controller of the GROMACS interface, calls the `wrgromacs` and `rdgromacs` routines. It also reads from the `template.dat` file, adding the energetic offset and inputting the vibronic couplings.
- `wrgromacs`: Writes a *.g96* structure file for input to GROMACS.
- `rdgromacs`: Reads the energies, frequencies and hessians output by GROMACS.

The `wrgromacs` and `rdgromacs` routines feature calls to the utility file `gromacs_io.f90` in the directory `source/lib/utilities/`

Note: Currently, the input *.g96* file is hard-coded in the `wrgromacs` routine to have as a residue name 'UNK', whilst the box size is hard-coded as 10 nm^3 in the `gromacs_io.f90` utility. The latter will be corrected when solvent is included to take the default box provided by GROMACS.

4.2 GROMACS run script

The bash script that runs all the necessary GROMACS commands and extracts the data in a way that is then readable by QUANTICS is structured as follows:

1. `module load` is used to load the correct version of GROMACS. If available, the double precision version should be used.

2. The filename variables are set up, as defined in QUANTICS and the number of atoms is read from the structure file.

3. The energy of the current geometry (as defined in the *.g96* structure file) is calculated using the following commands:

   `gmx_d grompp -f input_md.mdp -p [.top] -c [.g96] -o [.tpr]`

   used to pre-process the `.top` and `.g96` files and create the `.tpr` file which contains all the information for the actual calculation that is started using `mdrun`:

   `gmx_d mdrun -v -s [.tpr] -deffnm ener`

   which generates the output in a file called *ener.edr* from which the energy is read using:

   `gmx_d energy -f ener.edr -o ener.xvg`

   From the *.xvg* file, the energy is stored in *ener.dat* with units of $kJmol^{-1}$.

4. The forces can be extracted by dumping the contents of the *ener.trr* file (created in the previous step) and searching for the relevant data with *grep*:

   `gmx_d dump -f ener.trr | grep f`

   The forces are then stored in the file *force.dat* with units of $kJmol^{-1}$ nm^{-1}.

5. To get the hessian, a different calculation needs to be performed and hence another *.mdp* file is used for the pre-processing:

   `gmx_d grompp -f input_nm.mdp -p [.top] -c [.g96] -o [.tpr]`

   again, a *.tpr* is created used by `mdrun`:

   `gmx_d mdrun -v -s [.tpr] -deffnm hess -nt 1`

   The hessian is stored in the binary file *hess.mtx* and can be extracted by using dump:

   `gmx_d dump -mtx hess.mtx`

   The matrix is stored in the file *full_hess.out* with units of $kJmol^{-1}$ nm^{-2}.

6. The output file *gromacs.out* is written using *ener.dat* and *force.dat*.

The full script is written here:

```
#! /bin/bash

module load gromacs/2021.3-dbl

GRO_FILE=$1
TOP_FILE=$2
cd $3

sed '2q;d' $GRO_FILE > natm.dat
NATM=`cat natm.dat`
rm natm.dat

```

```
13 ################################################################
14 #       Get energy (integrator = md) in ener.dat               #
15 ################################################################
16
17         gmx_d grompp -f ../input_md.mdp -p $TOP_FILE -c $GRO_FILE -o ener.tpr -
       quiet
18
19         gmx_d mdrun -v -s ener.tpr -deffnm ener -quiet
20
21         printf "Potential\n0" > enerinp.dat
22         gmx_d energy -f ener.edr -o ener.xvg -quiet < enerinp.dat
23         rm enerinp.dat
24
25         tail -n -1 ener.xvg | awk '{print $2}' > ener.dat
26
27 ################################################################
28 #            Get forces (integrator = md) in force.dat         #
29 ################################################################
30
31         gmx_d dump -f ener.trr    | grep f > one.dat
32
33                 sed -i 1,2d one.dat
34
35                 cut -d{ -f 2 one.dat > two.dat
36                 cut -d} -f 1 two.dat  > force.dat
37                 rm one.dat
38                 rm two.dat
39
40         sed -i 's/,/ /g' force.dat
41
42 ################################################################
43 #       Get Hessian (integrator = nm) in hess.mtx              #
44 ################################################################
45
46         gmx_d grompp -f ../input_nm.mdp -p $TOP_FILE -c $GRO_FILE -o hess.tpr -
       quiet
47
48         gmx_d mdrun -v -s hess.tpr -deffnm hess -nt 1 -quiet
49
50         gmx_d dump -mtx hess.mtx -quiet > full_hess.out
51
52         sed -i 1,2d full_hess.out
53
54 # Write output
55
56 cat <<EOF> out
57 echo " GROMACS output: "
58 echo " "
59 echo " Energy (in kJ mol-1): "
60 echo "ener:"
61 cat ener.dat
62 echo " "
63 echo " Force (in kJ mol-1 nm-1, conversion in Quantics): "
64 echo "force:"
65 cat force.dat
66 echo " "
67 echo " Hessian: "
68 echo " "
69 echo "Full hessian in full_hess.out file (read by quantics) "
70 EOF
71 bash out > gromacs.out
72 rm out
73
74 rm ener.*
75 rm mdout.mdp
76 rm hess.*
```

References

[1] L. L. E. Cigrang, J. A. Green, S. Gómez, J. Cerezo, R. Improta, G. Prampolini, F. Santoro, and G. A. Worth, "Non-adiabatic direct quantum dynamics using force fields: Toward solvation", *J. Chem. Phys.* **160**, pp. 174120 (2024).

[2] "Overdia1.0, a Fortran 90 code for parametrization of model hamiltonians based on a maximum-overlap diabatisation, available free of charge at http://www.iccom.cnr.it/overdia-en, author = Santoro, F. and Green, A. J., year = 2022,".

APPENDIX D

Towards accessible quantum dynamics workflows

D.1 Thioformaldehyde active space

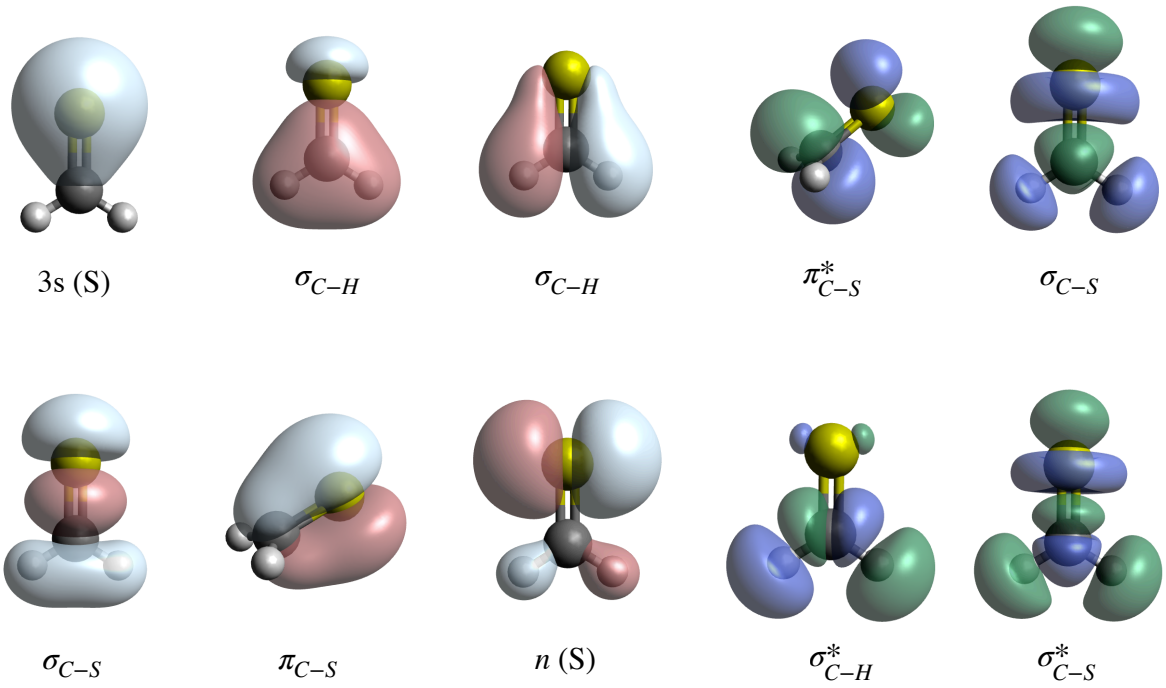
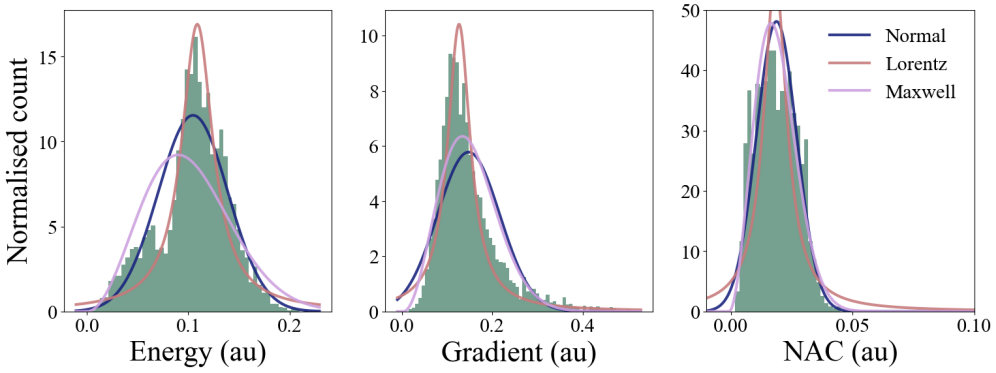


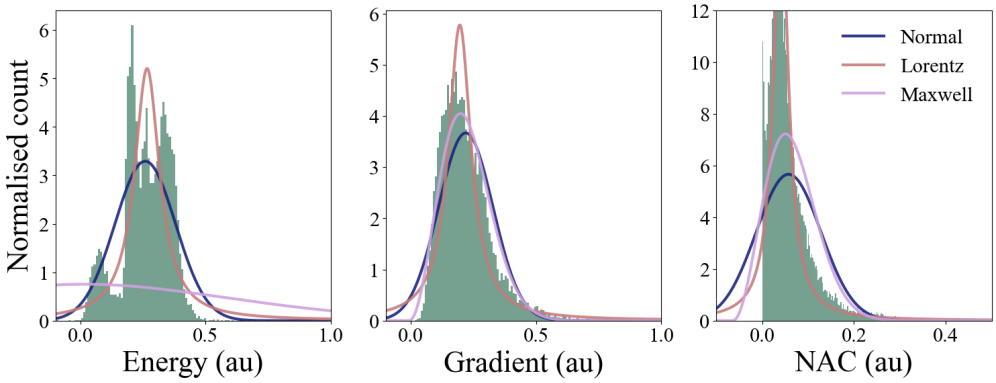
Figure D.1: Orbitals included in the (12,10) active space of thioformaldehyde. Red and light blue orbitals are initially doubly occupied, whereas green and dark blue orbitals are unoccupied.

D.2 Probability distributions of QC databases

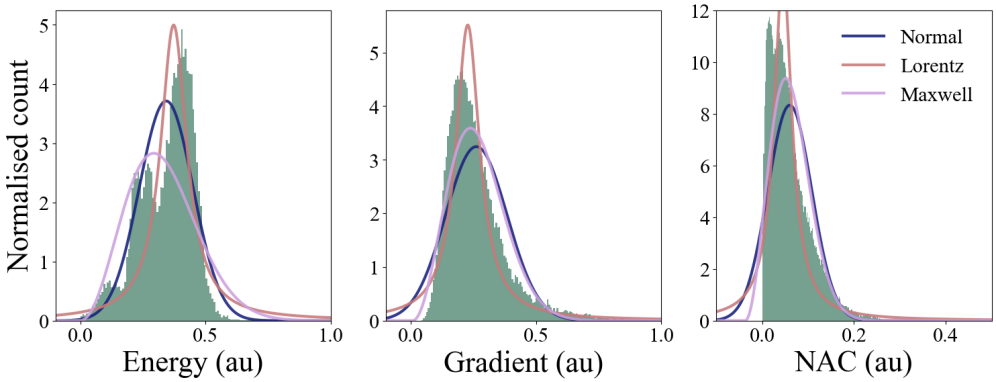
As stated in the main text of Chapter 5, notice that in all cases a normal distribution best describes the distribution of energies, whereas a Maxwell-Boltzmann distribution is more appropriate for gradients and non-adiabatic couplings.



(a) Thioformaldehyde, $N_{rec} = 1741$



(b) Formic acid, $N_{rec} = 6389$



(c) Formamide, $N_{rec} = 8746$

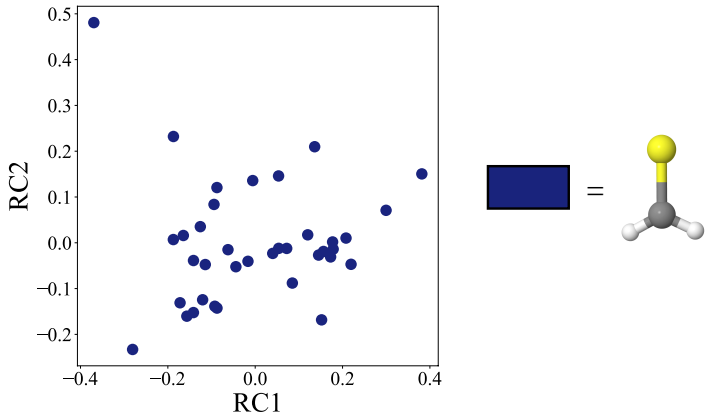
Figure D.2: Probability distributions fitted to (from left to right) energies, gradients and couplings (all in atomic units) for (a) Thioformaldehyde, (b) Formic acid, and (c) Formamide.

D.3 Thioformaldehyde S_1 excitation

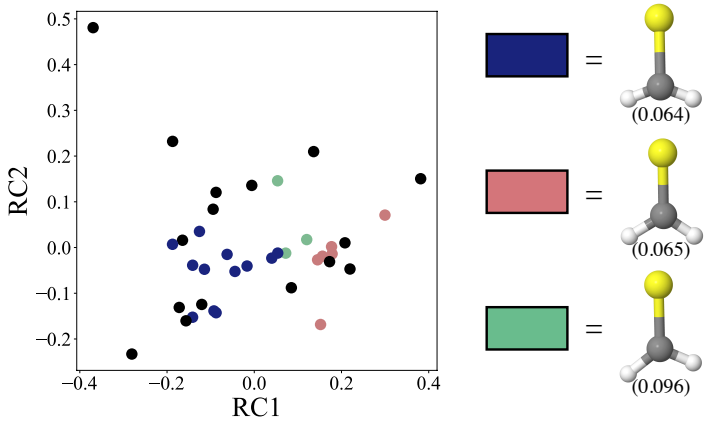
The trajectories resulting from a DD-vMCG simulation of thioformaldehyde, initiated on S_1, are plotted in Figure D.3. In total, 37 GWPs were used for a total simulation time of 500 fs. The two plots show the clustering results, using both DBSCAN and HDBSCAN. Additionally, the sensitivity analysis for DBSCAN is shown in Table D.1, from which it is clear that an EPS value of 0.5 is a sensible choice. Alternatively, one may also consider trying 0.4, and manually inspect the one noisy trajectory.

Table D.1: Sensitivity analysis with respect to EPS. N_C are the number of clusters generated, and N_{noise} the number of unclustered trajectories.

EPS	N_C	N_{noise}
0.10	0	37
0.20	1	20
0.30	1	3
0.40	1	1
0.50	1	0
0.60	1	0
0.70	1	0
0.80	1	0
0.90	1	0
1.00	1	0



(a) DBSCAN (EPS = 0.5)



(b) HDBSCAN

Figure D.3: Clustered trajectories of thioformaldehyde, along with the average structure of each cluster, using both DBSCAN and HDBSCAN. In parentheses are the RMSD values of the HDBSCAN structures, compared to the average DBSCAN geometry.

D.4 Comparison of dimensionality reduction & clustering techniques

Different algorithms to perform dimensionality reduction and clustering are compared in Figure D.4. Notice that instead of K-Means, the K-Medoids method[1,2] has been used, the reason being that K-Means is based on the arithmetic mean of the data and therefore not compatible with the use of a pairwise distance matrix between data points. Hence, it cannot be applied to the Fréchet distance matrix. K-Medoids does allow a distance matrix to be used, and is related to K-Means in the sense that it too defines clusters based on the distance away of the point perceived as the centre of that cluster. Looking at the figure, this method performs surprisingly well, but it is clear that DBSCAN is ultimately the most useful since it identifies noise and doesn't require the number of clusters to be identified beforehand.

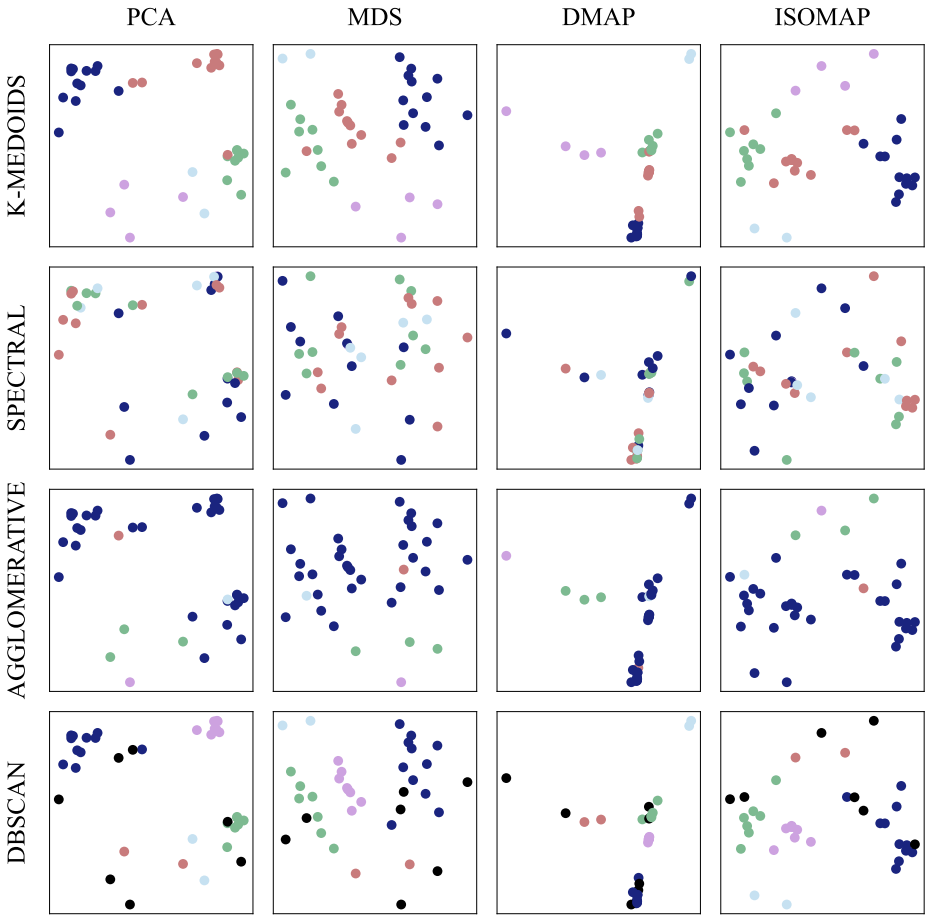
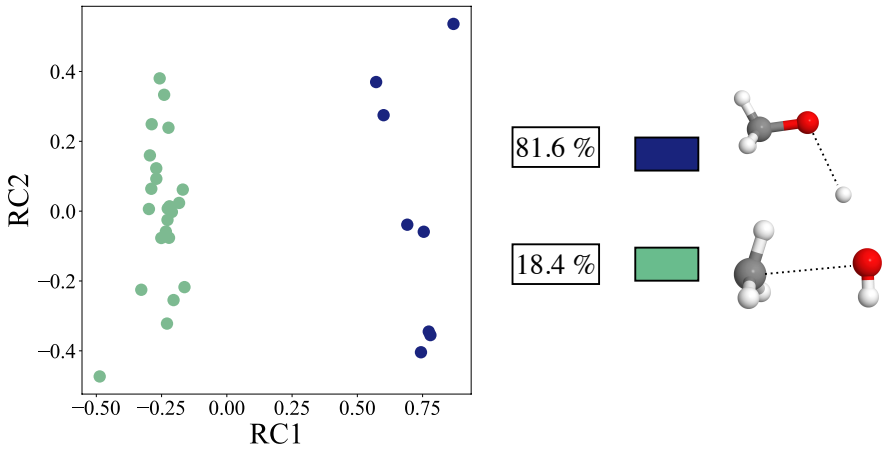


Figure D.4: Comparison of clustering (rows) and dimensionality reduction (columns) algorithms, applied to 37 formaldehyde trajectories.

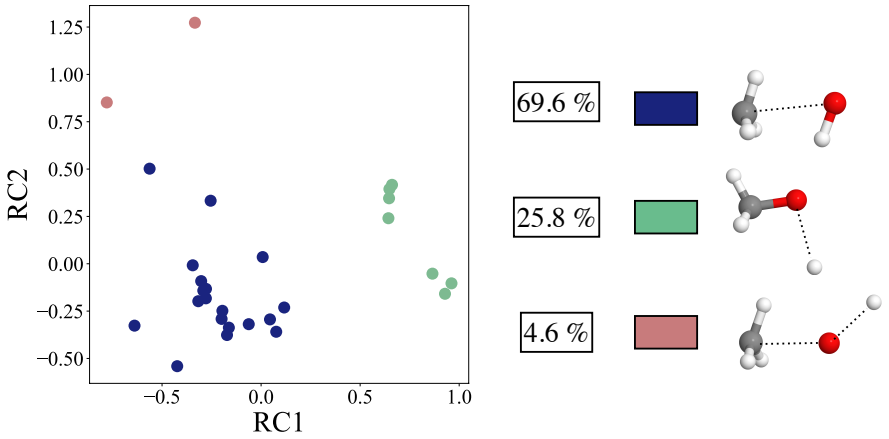
D.5 Automatic calculation of branching ratios in methanol

The data presented in Figure D.5 is the result of automatically determining the product ratios from the DD-vMCG simulations of methanol. The procedure itself is outlined in Section 5.4.2, whereas the three methanol simulations are described in Chapter 3.

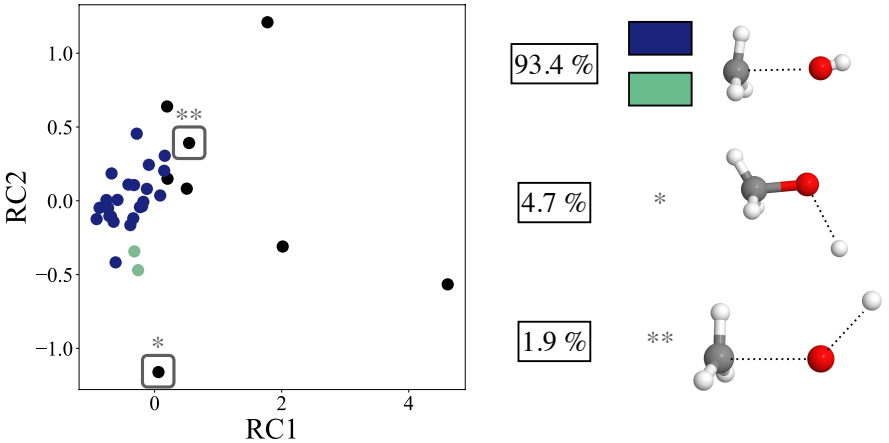
The products resulting from the first two simulations ($\tilde{A} \leftarrow \tilde{X}$ and $\tilde{B} \leftarrow \tilde{X}$) are captured perfectly by DBSCAN. In terms of reduced coordinates, in Figure D.5a they are easily assigned as RC1 = O-H bond length, and RC2 = C-O-H angle. This same assignment seems to be valid for Figure D.5b as well. However, the third simulation ($\tilde{C} \leftarrow \tilde{X}$) is more problematic. The most likely reason is that there is essentially one big cluster of 26 trajectories (out of a total of 33) in which C-O dissociation occurs, and only a single trajectory that contributes to either of the other two possible products. Clearly, density-based clustering algorithms struggle with situations like that, and hence it becomes necessary to perform some manual inspection of the noisy trajectories to identify the relevant products.



(a) $\tilde{A} \leftarrow \tilde{X}$; EPS $= 0.5$, $N_C = 2$, $N_{noise} = 2$ (removed)



(b) $\tilde{B} \leftarrow \tilde{X}$; EPS $= 0.5$, $N_C = 3$, $N_{noise} = 5$ (removed)



(c) $\tilde{C} \leftarrow \tilde{X}$; EPS $= 0.4$, $N_C = 2$, $N_{noise} = 8$

Figure D.5: ISOMAP and DBSCAN results for methanol simulations. *Left:* trajectories in reduced coordinates. *Right:* product yields and corresponding average structures. Also given are the EPS parameters used in DBSCAN, and the number of clusters, N_C, and noisy points, N_{noise}, identified.

D.6 Ground state geometries & frequencies

D.6.1 Formaldehyde

Table D.2: Cartesian coordinates (in Å) for formaldehyde (optimised using CCSD//aug-cc-pVDZ).

Atom	X	Y	Z
C	0.000000	0.000000	-0.534090
H	0.000000	0.937176	-1.122035
H	0.000000	-0.937176	-1.122035
O	0.000000	0.000000	0.681076

Table D.3: Frequencies (in cm^{-1}) of formaldehyde (CCSD//aug-cc-pVDZ).

Label	Frequency	Description
Q_1	1202.9348	CH_2 oop bend
Q_2	1296.3098	CH_2 rock
Q_3	1578.6981	C–O stretch
Q_4	1831.3703	CH_2 scissor
Q_5	2983.3521	CH_2 symmetric stretch
Q_6	3050.4994	CH_2 asymmetric stretch

D.6.2 Thioformaldehyde

Table D.4: Cartesian coordinates (in Å) for thioformaldehyde (optimised using CCSD//aug-cc-pVDZ).

Atom	X	Y	Z
C	0.000000	0.000000	-1.031512
H	0.000000	0.924833	-1.611352
H	0.000000	-0.924833	-1.611352
S	0.000000	0.000000	0.588236

Table D.5: Frequencies (in cm^{-1}) of thioformaldehyde (CCSD//aug-cc-pVDZ).

Label	Frequency	Description
Q_1	1028.0884	CH_2 oop bend
Q_2	1044.8179	CH_2 rock
Q_3	1100.2962	C–S stretch
Q_4	1557.2362	CH_2 scissor
Q_5	3134.9501	CH_2 symmetric stretch
Q_6	3222.7867	CH_2 asymmetric stretch

D.6.3 Formic acid

Table D.6: Cartesian coordinates (in Å) for formic acid (optimised using CCSD//aug-cc-pVDZ).

Atom	**X**	**Y**	**Z**
C	0.0000000	0.4243560	0.0000000
H	-0.3788400	1.4618700	0.0000000
O	-1.0381500	-0.4455490	0.0000000
H	-0.6570160	-1.3390360	0.0000000
O	1.1676320	0.1119280	0.0000000

Table D.7: Frequencies (in cm^{-1}) of formic acid (CCSD//aug-cc-pVDZ).

Label	**Frequency**	**Description**
Q_1	628.8741	CO_2 scissor
Q_2	661.2947	O–H oop bend
Q_3	1059.3448	C–H oop bend
Q_4	1137.8974	C–OH stretch + C–O–H bend
Q_5	1322.5978	C–O–H bend
Q_6	1409.3436	H–C–O bend
Q_7	1818.8071	C–O carbonyl stretch
Q_8	3119.5383	C–H stretch
Q_9	3771.9297	O–H stretch

D.6.4 Formamide

Table D.8: Cartesian coordinates (in Å) for formamide (optimised using CCSD//aug-cc-pVDZ).

Atom	X	Y	Z
C	0.000000	0.401721	-0.096031
N	0.000000	-0.162504	1.145371
H	0.000000	1.541283	-0.009698
H	0.000000	0.374796	1.979555
O	0.000000	-0.207339	-1.131411
H	0.000000	-1.153754	1.217289

Table D.9: Frequencies (in cm^{-1}) of formamide (CCSD//aug-cc-pVDZ).

Label	Frequency	Description
Q_1	-518.4278	NH_2 oop bend
Q_2	487.9620	NH_2 oop twist
Q_3	502.5994	O–C–N scissor
Q_4	1039.3102	C–N–H bend
Q_5	1124.0515	C–H oop bend
Q_6	1295.2009	C–N stretch
Q_7	1494.3722	C–H wag
Q_8	1627.0117	NH_2 scissor
Q_9	1957.2915	C–O stretch
Q_{10}	2702.5836	C–H stretch
Q_{11}	3829.6820	NH_2 symmetric stretch
Q_{12}	3967.9268	NH_2 asymmetric stretch

D.7 Validating assignment of reduced coordinates

When assigning a nuclear DoF to one of the reduced coordinates obtained by, in this case, ISOMAP, a simple validation exercise can be carried out. The two coordinates should of course be correlated, which can be determined by plotting them against each other, for each trajectory. Such plots are shown in Figure D.6, with the data taken from the formic acid results (see Figure 5.15b). RC1 was assumed to be related to the C-OH bond length, which is confirmed by seeing the separation of trajectories with a short C-OH bond and low RC1 value (blue cluster), and those leading to dissociation of the C-OH bond and correspondingly high RC1 values (red and green clusters). Similarly, the plot of RC2 and normal mode Q_2, which is related to the O-C-O-H dihedral angle, shows a clear correlation between the two coordinates.

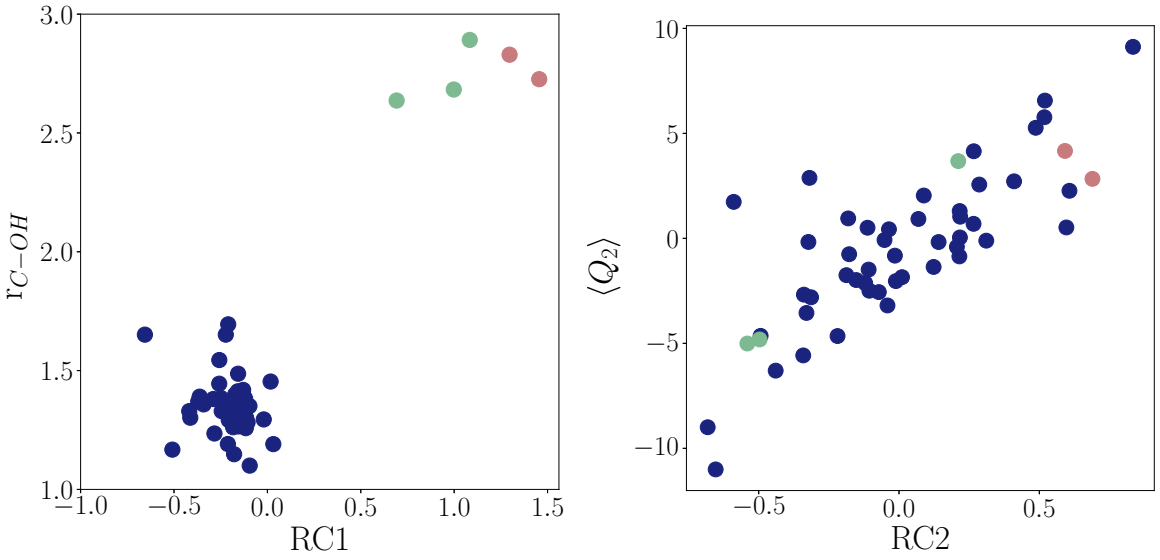


Figure D.6: Correlation plots for reduced coordinates of formic acid. *Left*: RC1 against the C-OH bond distance. *Right*: RC2 against the expectation value of normal mode Q_2.

References

[1] Arora, P.; D.; Varshney, S. Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia Comput. Sci.* **2016**, *78*, 507–512.

[2] Madhulatha, T. S. Advances in Computing and Information Technology; Springer Berlin Heidelberg, 2011; pp 472–481.