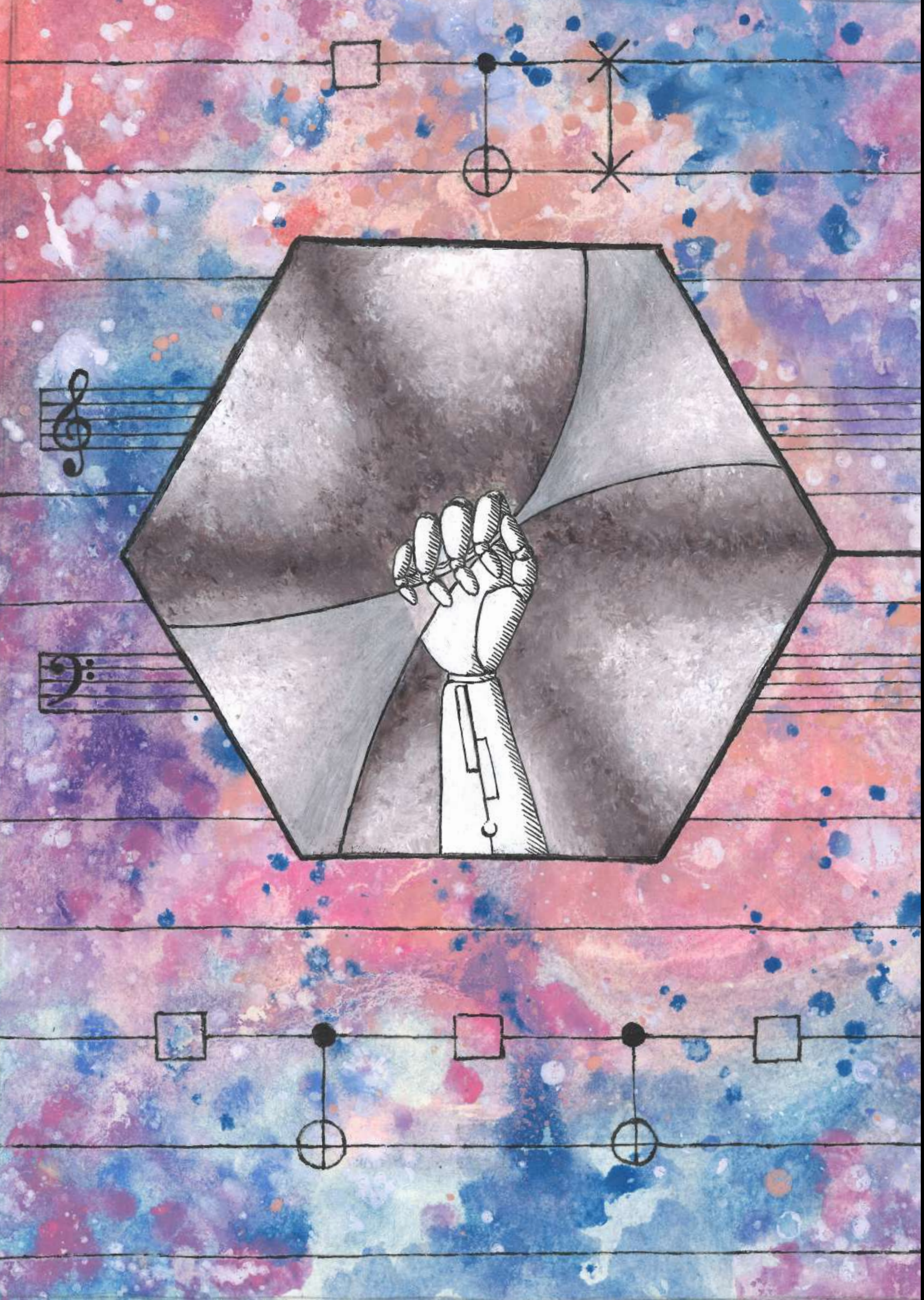
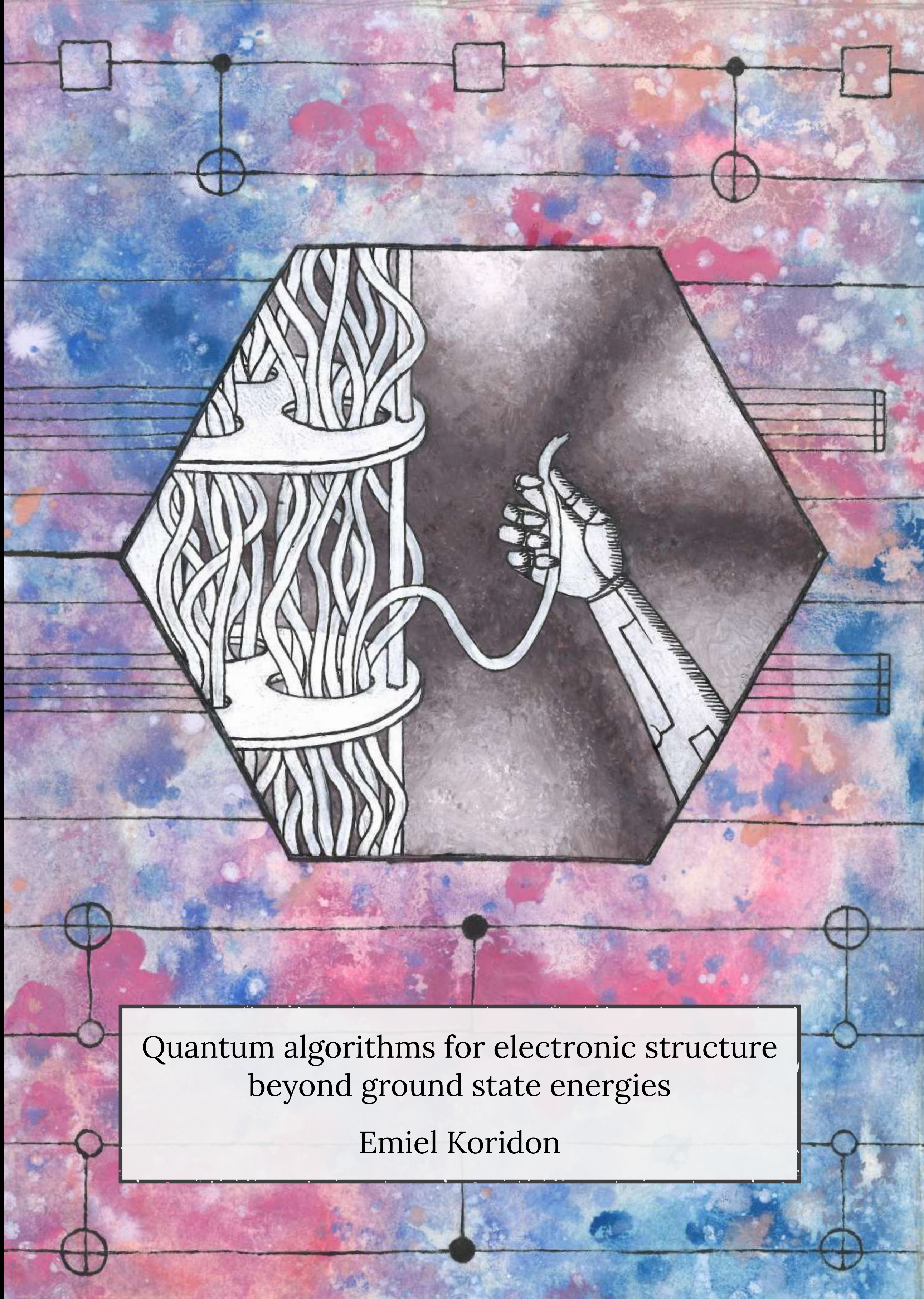




Quantum algorithms for electronic structure beyond ground state energies

Emiel Koridon

Quantum algorithms for electronic structure beyond ground state energies - Emiel Koridon

VRIJE UNIVERSITEIT

Quantum algorithms for electronic structure beyond ground state energies

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
volgens besluit van de decaan
van de Faculteit der Bètawetenschappen
in het openbaar te verdedigen
op woensdag 7 mei 2025 om 15.45 uur
in de universiteit

door

Emiel Frans Koridon

geboren te Amsterdam

promotoren:	prof.dr. L. Visscher
	dr. F. Buda
promotiecommissie:	prof.dr. J. Rojo
	prof.dr. C.W.J. Beenakker
	prof.dr. K. Schoutens
	dr. K. Dolblhoff-Dier
	dr. S. Knecht

Voor mijn moeder, Ella

On the cover: *A vivid, abstract background suggestive of outer space, created with soft pastels blended in water for a fluid, vibrant texture. At its center is a biphenyl molecule, spanning front and back covers. Inside the front ring, a quantum computer is manipulated by a robotic hand; on the back, the hand grasps a conical intersection. Painted quantum circuits extend above and below the rings, while a musical stave weaves behind them. The contrasting materials – soft pastels, acrylic paint, and precise fineliners – as well as the colours and shapes, mirror the interplay between creative expression and scientific structure.* – by Eva Koridon

Contents

1 Introduction 1
- 1.1 Preface . . . 1
- 1.2 Electronic Structure Problem . . . 3
  - 1.2.1 Born-Oppenheimer approximation . . . 4
  - 1.2.2 Discretization . . . 4
  - 1.2.3 Canonical quantum chemistry methods . . . 8
- 1.3 Active Space Methods . . . 12
  - 1.3.1 Construction of the Hamiltonian . . . 12
  - 1.3.2 Choosing consistent orbitals . . . 14
  - 1.3.3 Static and dynamic correlations . . . 17
- 1.4 Quantum Algorithms for Chemistry . . . 19
  - 1.4.1 Paradigms of quantum computing . . . 20
  - 1.4.2 The Variational Quantum Eigensolver . . . 21
  - 1.4.3 Beyond near-term quantum algorithms . . . 26
- 1.5 Outline of the thesis . . . 27

2 Orbital transformations to reduce the 1-norm of the electronic structure Hamiltonian 33
- 2.1 Introduction . . . 33
- 2.2 Theory . . . 34
  - 2.2.1 Electronic structure Hamiltonian in second quantization . . . 34
  - 2.2.2 The 1-Norm in quantum computing . . . 35
  - 2.2.3 Localized orbitals . . . 41
  - 2.2.4 1-Norm orbital-optimization . . . 44
- 2.3 Computational details . . . 45

2.4 Results . . . 45
2.4.1 Hydrogen and alkane chains: scaling of the 1-norm by increasing the number of atoms . . . 46
2.4.2 Benchmarking λ_Q for a variety of molecules and active spaces 47
2.4.3 Effect of increasing the size of the active space on λ_Q . . . 50
2.5 Conclusions . . . 52
Appendices . . . 54
2.A Electronic integrals transformation . . . 54
2.B Explicit form of λ_Q in terms of molecular integrals . . . 55

3 FragPT2: Multi-Fragment Wavefunction Embedding with Perturbative Interactions 61
3.1 Introduction . . . 61
3.2 FragPT2 method . . . 64
3.2.1 Construction of re-canonicalized intrinsic localized molecular orbitals . . . 65
3.2.2 Fragment embedding . . . 66
3.2.3 Multi-reference perturbation theory . . . 69
3.3 Numerical demonstration . . . 72
3.3.1 Numerical simulation details . . . 73
3.3.2 N_2 dimer . . . 73
3.3.3 Aromatic dimers . . . 73
3.3.4 Butadiene . . . 77
3.4 Outlook . . . 78
3.4.1 Computational efficiency . . . 78
3.4.2 Integration with quantum algorithms . . . 79
3.4.3 Further extensions . . . 80
3.5 Conclusion . . . 82
Appendices . . . 82
3.A Hamiltonian decomposition . . . 82
3.A.1 Charge-conserving terms . . . 83
3.A.2 Charge transfer terms . . . 84
3.B Fragment matrix elements . . . 86
3.B.1 Dispersion . . . 86
3.B.2 Single-charge transfer . . . 87
3.B.3 Double-charge transfer . . . 87
3.B.4 Triplet-triplet . . . 88

4 Analytical non-adiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver 91
4.1 Introduction . . . 91
4.2 Theory . . . 93
4.2.1 Quantum Chemistry for excited states . . . 93

4.2.2 Estimation of energies, analytical gradients and non-adiabatic coupling on a quantum computer . . . 97
4.3 Computational details . . . 107
4.4 Numerical results . . . 108
4.4.1 Illustration of the final state resolution in SA-OO-VQE . . . 108
4.4.2 Calculation of analytical gradients and non-adiabatic couplings . . . 110
4.4.3 Geometry optimization to locate formaldimine's conical intersection in the (α, ϕ) space . . . 113
4.4.4 Conical intersection optimization in the full geometry space 115
4.5 Conclusions and perspectives . . . 116
Appendices . . . 117
4.A Circuit gradient $\mathcal{G}^{\mathrm{C}}$ and Hessian $\mathcal{H}^{\mathrm{CC}}$ . . . 117
4.B Nuclear derivative of the electronic Hamiltonian operator . . . 120
4.C Analytical derivation of non-adiabatic couplings for SA-OO-VQE . 121

5 A hybrid quantum algorithm to detect conical intersections 125
5.1 Introduction . . . 125
5.2 Background . . . 127
5.2.1 Conical intersections . . . 127
5.2.2 Berry phases in real Hamiltonians . . . 128
5.2.3 Measuring Berry phase with a variational wavefunction . . 129
5.3 Methods . . . 130
5.3.1 Fixing the gauge with a real ansatz . . . 130
5.3.2 Avoiding full optimization via Newton-Raphson steps . . . 131
5.3.3 Regularization and backtracking . . . 132
5.3.4 Measuring the final overlap . . . 134
5.3.5 Overview of the algorithm . . . 135
5.4 Error analysis and bounding . . . 136
5.4.1 Bounding the NR error . . . 137
5.4.2 Bounding the sampling noise . . . 138
5.4.3 Scaling of the total cost . . . 138
5.5 Adapting to an orbital-optimized PQC ansatz . . . 139
5.5.1 An OO-PQC ansatz with geometric continuity . . . 139
5.5.2 Measuring boundary terms with the OO-PQC ansatz . . . 141
5.5.3 Newton-Raphson updates of the OO-PQC ansatz . . . 141
5.6 Numerical results . . . 142
5.6.1 Numerical simulation details . . . 143
5.6.2 Minimal model with an degeneracy-free ansatz . . . 143
5.6.3 Sampling noise . . . 145
5.6.4 Larger basis and active space . . . 146
5.7 Conclusion and outlook . . . 148
5.7.1 Paths towards improving convergence . . . 148
5.7.2 Potential applications . . . 150

5.7.3 Outlook . . . 151
Appendices . . . 151
5.A Bounding overlaps by change in ansatz parameters . . . 152
5.B Bounding the norm of energy derivatives . . . 153
5.C Analytical orbital gradient and Hessian . . . 155
5.D Bounding the cumulative error due to Newton-Raphson updates . . . 157
5.E Bounding the sampling cost . . . 160

6 Machine Learning Density Functionals from Noisy Quantum Data 163
6.1 Introduction . . . 163
6.2 Background . . . 166
6.2.1 Density Functional Theory . . . 166
6.2.2 The Fermi-Hubbard model . . . 167
6.2.3 The learning task . . . 168
6.3 The dataset . . . 168
6.3.1 Expectation value estimation . . . 169
6.3.2 Variational quantum eigensolver . . . 170
6.4 Learning the functional . . . 172
6.4.1 Method . . . 172
6.4.2 Alternative model selection . . . 173
6.5 Results . . . 175
6.5.1 Energy prediction . . . 176
6.5.2 Density optimization . . . 179
6.6 Discussion and outlook . . . 181

Bibliography 183

Summary 217

Acknowledgments 221

List of publications 223

List of Abbreviations 225

CHAPTER 1

Introduction

1.1 Preface

Quantum chemistry plays a crucial role in understanding and predicting the behavior of molecular systems. The rise of high-performance computing has revolutionized this field by enabling the development of computational methods to explore molecular ground state energies, excited states, and reaction dynamics. At its core is the challenge of solving the electronic structure problem, which focuses on determining the energies and properties of electrons in molecules. These insights are critical across industries such as pharmaceuticals, materials science, and renewable energy [1]. Approximate quantum chemistry algorithms for classical computers have been widely used in these areas, the most important being Density Functional Theory (DFT), also called the workhorse of quantum chemistry. It is especially useful for studying large systems due to its balance between accuracy and computational efficiency [2]. However, DFT is not a universal solution to every problem, and it can fail to model strong electron correlations in molecules or to model excited states. For these cases, more accurate methods like coupled-cluster, tensor networks and active space methods have been used to some success [3]. These approximate methods usually have some parameter to adjust their accuracy. However, the challenge is that for some systems, dialing up the accuracy to an acceptable level may exceed our computational budget.

The problem lies in the complexity of solving the electronic structure problem exactly, which grows exponentially with system size when using classical computers. This computational bottleneck limits the ability to accurately model larger or more complex molecules, especially those that exhibit strong electronic correlations. Quantum computers offer the potential to break through some of these limitations by efficiently solving certain tasks that are intractable for their classical counterparts. The concept of *quantum advantage* refers to the point at which a quantum computer can outperform classical computers for a specific task. Current-day quantum devices are imperfect: they have low qubit counts and are plagued by noise. Despite this, there has been rapid progress in the field, and quantum advantage already has been reached for the problem of sampling from probability distributions that are hard for classical computers [4–6]. Nevertheless, these problems are, to the best of our knowledge, completely useless; making this milestone an important one but just a starting point to the goal of *quantum utility*. This could be defined as the point where quantum can outperform classical on specific, practical tasks with real-world value. *Quantum computational chemistry* is a term that is used in the community that indicates using quantum computers to solve problems in quantum chemistry [7].

Open problems in quantum chemistry that quantum computers could potentially address include understanding biological nitrogen fixation [8, 9], assessing chemical mechanisms of reactivity for pharmaceuticals [10] and transition metal catalysts for carbon capture [11]. While these applications could revolutionize drug discovery, materials science, and renewable energy, they are still far from the scope of near-term quantum devices, as the noise degrades the final result beyond the high precision that is required. In this context, achieving quantum advantage requires both identifying suitable computational targets and optimizing quantum algorithms that are more robust to noise and work within the constraints of (near-term) quantum devices. This thesis tries to work on both sides, by devising schemes to reduce both space- and time-complexity of quantum algorithms, and proposing noise-robust targets beyond ground state energies that are more difficult for classical algorithms.

This introduction presents fundamental concepts from both quantum chemistry and quantum computing to provide a foundation for the subsequent chapters, which represent the primary contributions of this thesis. In doing so, it seeks to unify the chapters into a cohesive entity. It is structured as follows: Section 1.2 provides an overview of the electronic structure problem in quantum chemistry, including the Born-Oppenheimer approximation, orbitals, basis sets, and Hartree-Fock and DFT methods, focused on ground states. Section 1.3 discusses the construction of an *active space*, a central concept in many chapters of this thesis, that enables the treatment of correlated systems. It goes into depth on orbital optimization and static and dynamical correlation. Section 1.4 provides a concise introduction into the emerging field of near-term quantum simulation of chemistry, including the Noisy Intermediate-Scale Quantum (NISQ) era and variational quantum algorithms. Finally, Section 1.5 outlines the contributions of this thesis and the structure of the subsequent chapters.

1.2 Electronic Structure Problem

A molecule is a quantum system consisting of a collection of nuclei and electrons that interact with each other through the electromagnetic force. To construct a physical description of the system, a general approach is to consider the energy operator of the system, called the *Hamiltonian*. To construct the Hamiltonian of the molecule, we have to consider all the different interactions between the nuclei and electrons. This includes their kinetic energies (T_e and T_N); nuclear-nuclear (U_NN) and electron-electron (U_ee) repulsion; and nuclear-electron (U_Ne) attraction. The total molecular Hamiltonian can thus be written as:

$$\begin{aligned} H_\mathrm{mol} = & \overbrace{-\sum_{i=1}^{\eta} \frac{1}{2}\nabla_i^2}^{T_\mathrm{e}} \quad \overbrace{-\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2}^{T_\mathrm{N}} \quad \overbrace{-\sum_{i=1}^{\eta}\sum_{A=1}^{M} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}}^{U_\mathrm{Ne}} \\ & \overbrace{+\sum_{i<j}^{\eta} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}^{U_\mathrm{ee}} \quad \overbrace{+\sum_{A<B}^{M} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}}^{U_\mathrm{NN}} \end{aligned} \tag{1.1}$$

where M_A is the mass of nucleus A, η is the number of electrons, M is the number of nuclei, Z_A is the charge of nucleus A, and $\mathbf{r}_i \in \mathbb{R}^3$ and $\mathbf{R}_A \in \mathbb{R}^3$ are the position vectors of electron i and nucleus A, respectively. In this thesis, they are often written in shorthand as one big vector $\mathbf{r} \in \mathbb{R}^{3\eta}$ and $\mathbf{R} \in \mathbb{R}^{3M}$. In Eq. (1.1) we are using *atomic units* that we will use throughout this work; in this system of natural units the electron mass, the elementary unit of charge, the reduced Planck constant and the Coulomb force constant are all normalized as: $m_e = e = \hbar = \frac{1}{4\pi\epsilon_0} = 1$. This results in the energy having atomic units of Hartree, which is equivalent to about 27.2 electronvolt (eV).

To retrieve the molecular energy levels and their eigenstates, we solve the time-independent[1] Schrödinger equation:

$$H\left|\Psi_k\right\rangle = E_k\left|\Psi_k\right\rangle, \tag{1.2}$$

where $\left|\Psi_k\right\rangle$ and E_k its k-th eigenstate and eigenvalue. If one manages to find these, there are many interesting properties that we can extract. The energies give essential information for understanding chemical reactions and determining reaction rates between the reactants and the products. Other quantities like vibrational frequencies, dipole moments and polarizabilities provide valuable insights into the behavior of molecular systems. In the following section, we will isolate the Hamiltonian for the electrons and show how to approximate the ground state with classical methods.

---

[1] This is not the only Schrödinger equation that is important, as many quantum chemistry methods also rely on the time-dependent version, especially when simulating dynamics.

1.2.1 Born-Oppenheimer approximation

There is no known analytical solution to the Schrödinger equation (1.2) for this interacting system, already for any element heavier than hydrogen. However, we can at this point invoke a trick: we isolate the dynamics of the electrons. This is the *Born-Oppenheimer approximation*, introduced by the two famous physicists in 1927 [12], and has been the cornerstone of any quantum chemistry method. As the nuclei are several orders of magnitude heavier than the electrons, we allow the nuclei to be treated as classical particles with fixed positions given by the now classical variables $\mathbf{R}$. This approximation simplifies the molecular Hamiltonian by removing the kinetic energy terms of the nuclei and treating them as fixed points in space. The electronic Hamiltonian in the Born-Oppenheimer approximation is given by:

$$H_{\mathrm{e}}(\mathbf{R}) = E_{\mathrm{NN}}(\mathbf{R}) + T_{\mathrm{e}} + U_{\mathrm{ee}} + U_{\mathrm{Ne}}(\mathbf{R}) \tag{1.3}$$

where the nuclear repulsion term $E_{\mathrm{NN}}(\mathbf{R}) = \sum_{A<B}^{M} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}$ is now a constant, that is often omitted in this thesis as it can be added without extra computational cost.

Solving the Schrödinger equation for the electronic Hamiltonian in Eq. (1.3) is what's meant with the *electronic structure problem*. In contrast to the nuclei, the gap between the ground and excited state of the electrons is usually on the order of several eV. This is much higher than the typical nuclear kinetic energy or the thermal energy at room temperature, which is about 25 meV. This makes the Born-Oppenheimer an extremely reasonable approximation, as well as making the electronic ground state $|\Psi_0\rangle$ and its energy $E_0(\mathbf{R})$ of particular interest, as the electrons prefer to be in their ground state in typical reaction environments. The dynamics of the nuclei can be reintroduced, after one has solved Eq. (1.3), by letting them move over the *potential energy surface* (PES) $E_0(\mathbf{R})$ generated by the electrons. In practice, this PES is extremely useful to study molecular reactions; its local minima define equilibrium geometries. A reaction can be described by going from some reactant state located in the local minimum $\mathbf{R}_{\mathrm{R}}$ to some product state at some other $\mathbf{R}_{\mathrm{P}}$. The fastest way from $\mathbf{R}_{\mathrm{R}} \to \mathbf{R}_{\mathrm{P}}$ travels through a saddle point $\mathbf{R}_{\mathrm{T}}$ representing the *transition state*. The height of the barrier ΔE at the transition state will determine the *rate* Γ of the reaction. The empirical formula of the Arrhenius equation unveils that this reaction rate depends exponentially on this barrier: $\Gamma \propto e^{-\Delta E}$. This implies that to accurately predict rates we need a high-precision electronic structure method.

1.2.2 Discretization

Orbital basis

Solving the Schrödinger equation for the Hamiltonian in its bare form of Eq. (1.3) amounts to a series of continuous differential equations that are very hard to solve. To make the electronic structure problem manageable for our digital computers, we have to discretize the space. In quantum chemistry, this entails the introduction of *basis sets*, a discrete set of *single-particle basis functions* $\{\chi_\mu\}$ that are centered around the nuclei and are eigenfunctions of some one-electron Hamiltonian. The simplest version of the electronic Hamiltonian, which is the Hydrogen atom with one nucleus and one electron, can be solved analytically. We can use these exact solutions as effective basis functions, called *atomic orbitals*, and with some adjustments, they work well for describing other nuclei too. To calculate expectation values of (local one-body) observables, we need them to be square integrable functions on $\mathbb{R}^3$, i.e. part of $L^2(\mathbb{R}^3)$, such that

$$\langle O \rangle = \int d\mathbf{r} \chi_\mu^*(\mathbf{r}) O(\mathbf{r}) \chi_\mu(\mathbf{r}) < \infty \tag{1.4}$$

We ideally want these atomic orbitals to satisfy three requirements:

- They should have a systematic way of improving their accuracy towards *completeness*, i.e. the ability to approximate any function on $\mathbb{R}^3$ with a linear combination.
- They should be able to represent any atomic and molecular state and electron density with only a few terms in a linear combination.
- Their integrals should be efficiently computable.

It is impossible to find a general set of functions that fulfill all three requirements for all systems. Thus, we need to make compromises depending on the problem and system we want to solve. While a general basis of the space like a plane wave basis would be sufficient for the first and last point, it has very slow convergence to represent a compact molecule. Hydrogenic orbitals fulfill the first and second point, but they are difficult to analytically integrate. As a common compromise, quantum chemists use either Slater-type orbitals (STOs) or Gaussian-type Orbitals (GTOs). Out of these approximations of hydrogenic orbitals, STOs are more accurate (modelling correctly the nuclear-electron cusp) but GTOs are more computationally efficient (having a closed-form expression for the integrals). The general form of a GTO is:

$$\chi_\mu(\mathbf{r}) \propto x^k y^l z^m e^{-\alpha r^2} \tag{1.5}$$

To approximate the cusp, contracted Gaussians are used that consist of a linear combination of GTOs. The coefficients can be optimized for specific atoms by optimizing the energy with an electronic structure method.

To build a multi-particle wavefunction of the molecule, we need a compact basis set for the molecule. Directly using the atomic orbital basis will not do, as they do not take into account the interactions between neighbouring atoms. Neither do they comprise an orthonormal set. As the non-orthogonal atomic orbitals come close together, they like to combine into linear combinations and form *molecular bonds*. We can build *molecular orbitals* (MOs) that span over multiple atoms, represented mathematically as a linear combination of atomic orbitals:

$$\phi_p(\mathbf{R}, \mathbf{r}) = \sum_{\mu} \chi_\mu(\mathbf{r}, \mathbf{R}) C_{\mu p}(\mathbf{R}). \tag{1.6}$$

The set of N MOs $\{\phi_p\}$ are spatial orbitals. Electrons are fermions, thus we have to specify their spin as well as their position. We can take an orthonormal set of spin functions σ, and build the following basis of spin-orbitals:

$$\psi_{p\sigma}(\mathbf{x}) = \phi_p(\mathbf{r})\sigma(m_\mathrm{s}) \tag{1.7}$$

where we explicitly use the spin label σ on the left, indicating the spin-function, and $m_\mathrm{s} \in \{\uparrow, \downarrow\}$, and we used the shorthand composite label $\mathbf{x} = (\mathbf{r}, m_\mathrm{s})$. This new MO basis will be a more efficient compact basis of the electrons in the molecule, that are orthonormal:

$$\int d\mathbf{x} \psi^*_{p\sigma}(\mathbf{x}) \psi_{q\tau}(\mathbf{x}) = \delta_{pq}\delta_{\sigma\tau} \tag{1.8}$$

Notice that each pair of spin-orbitals has corresponds to one unique spatial orbital. This is also called the *restricted* formalism, which we will use throughout this thesis. Formulating spin-orbitals in this manner makes no approximation for building exact ground states, as the Hamiltonian commutes with total (S^2) and projected (S_z) spins. However, approximate methods produce states that are not necessarily eigenstates of the Hamiltonian, and thus might not possess the same symmetries. Therefore, restricting spin-orbitals in this way, effectively forcing the symmetry on the state, might be suboptimal for some approximate methods. In the *unrestricted* formalism, different spatial orbitals can be used for different spins. This is commonly used in approximate methods like Hartree-Fock (see Section 1.2.3) whenever the ground state is suspected to contain unpaired spins (open shell).

Slater determinants

To subsequently build a multi-particle state for η electrons out of the spin-orbital basis, we can take a product of single-particle functions. However, we are constrained by the fact that electrons are fermions; the total wavefunction should be anti-symmetric under a swap of the position-spin variables of any two electrons: $\Phi(\mathbf{x}_1, \mathbf{x}_2) = -\Phi(\mathbf{x}_2, \mathbf{x}_1)$. To enforce the anti-symmetric property in all variables, we construct the multi-particle basis by anti-symmetrizing the product state. This results in the *Slater determinant*:

$$\Phi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\eta) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \ldots & \psi_\eta(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \ldots & \psi_\eta(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_\eta) & \psi_2(\mathbf{x}_\eta) & \ldots & \psi_\eta(\mathbf{x}_\eta) \end{vmatrix}. \tag{1.9}$$

A closer look at Eq. (1.9) tells us that, given a fixed set of $2N$ spin-orbitals, the only degree of freedom we now have is to choose which of them are occupied (the ones to include in the Slater determinant), and which are unoccupied. Therefore, we can denote a Slater determinant as $\Phi(\{\psi_i\})$, where $\{\psi_i\}$ are the set of spin-orbitals that are occupied. The amount of Slater determinants we can make from putting η electrons in $2N$ spin-orbitals is exactly $\binom{2N}{\eta}$ – disregarding symmetry restrictions.

Second quantization

This inspires us to go to the paradigm of second quantization. In second quantization, we move from *what electron occupies which orbitals* to *what orbitals are occupied*. The basis states live in an abstract linear vector space called the *Fock space*, and each Slater determinant is described by an *occupation number vector* (ONV) $|\mathbf{f}\rangle$. In this vector, each spin-orbital is enumerated and specified to be empty (0) or filled (1). The operators that can change these ONVs into one another are the fermionic creation and annihilation operators $a^\dagger_{p\sigma}$ and $a_{p\sigma}$ respectively, which create or remove a particle in mode p with spin σ. They have the following commutation properties:

$$\left[a^\dagger_{p\sigma}, a_{q\tau}\right] = \delta_{pq}\delta_{\sigma\tau} \quad \left[a_{p\sigma}, a_{q\tau}\right] = \left[a^\dagger_{p\sigma}, a^\dagger_{q\tau}\right] = 0 \tag{1.10}$$

A general ONV can be defined by creating electrons in the specified spin-orbitals by acting with creation operators onto the normalized vacuum state $|\Omega\rangle$:

$$|\mathbf{f}\rangle = |f_1, f_2, \ldots, f_{2N}\rangle = {a^\dagger_{1\uparrow}}^{f_1} {a^\dagger_{1\downarrow}}^{f_2} \ldots {a^\dagger_{N\downarrow}}^{f_{2N}} |\Omega\rangle \,. \tag{1.11}$$

To obtain the electronic structure Hamiltonian of Eq. (1.3) in second quantization, we can project it onto the finite basis of occupation number vectors (1.11). Because the Hamiltonian does not depend on spin, the spin variable can be integrated out. For a concise formulation, consider the following spin-summed excitation operators:

$$E_{pq} = \sum_\sigma a^\dagger_{p\sigma} a_{q\sigma} \tag{1.12}$$

$$e_{pqrs} = \sum_{\sigma\tau} a^\dagger_{p\sigma} a^\dagger_{r\tau} a_{s\tau} a_{q\sigma} = E_{pq}E_{rs} - E_{ps}\delta_{qr} \tag{1.13}$$

As the Hamiltonian in Eq. (1.3) only couples two determinants that are connected by at most two excitations, these form a complete operator basis for the Hamiltonian. After the projection, the second quantized Hamiltonian reads:

$$H = \sum_{pq} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} e_{pqrs} \tag{1.14}$$

Where the one- and two-electron integrals are given by integrating spatial orbitals as:

$$h_{pq} = \int d\mathbf{r} \phi_p^*(\mathbf{r}) \left(\frac{1}{2} \nabla^2 - \sum_{A=1}^{M} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right) \phi_q(\mathbf{r}) \tag{1.15}$$

$$g_{pqrs} = \iint d\mathbf{r}_1 d\mathbf{r}_2 \frac{\phi_p^*(\mathbf{r}_1)\phi_q(\mathbf{r}_1)\phi_r^*(\mathbf{r}_2)\phi_s(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \tag{1.16}$$

1.2.3 Canonical quantum chemistry methods

Levels of approximation

Every molecule is a quantum system; however, one does not have to solve the Schrödinger equation (1.2) exactly to sufficiently capture the physics of the system. In this thesis, we will differentiate between two types of methods used in quantum chemistry: those designed for classical computers (*classical methods*) and those for quantum computers (*quantum methods*). Classical methods for solving for ground states of Eq. (1.14) have found success in a plethora of real-world problems [1]. In some cases, just treating electrostatic interactions with classical force fields is sufficient to describe their behavior. Examples of these systems are simple salts, ionic crystals or large proteins and biomolecules like enzymes and DNA [13].

In other cases, treating the electrons with a single Slater determinant (or its density, as in DFT), is sufficient to capture the electronic structure of the molecule. These type of methods we will call *single-configurational methods*. They can be used for many small organic molecules like ethane or benzene, but also bigger closed-shell molecules. In these cases, properties like equilibrium molecular geometries, reaction energies and vibrational frequencies are accurately described by single-configurational classical methods [14].

There are also cases where a single-configurational description breaks down completely [14]. These class of systems are called *strongly correlated systems*. For us, these are of particular interest, as they challenge even the highest accuracy classical methods, suggesting the application of quantum methods. For a discussion about strongly correlated molecules and the types of correlations involved, see Section 1.3.3. Not all types of correlation demand quantum methods; some can be managed with classical methods. In fact, an efficient approach requires *multi-scale modeling* – a layered strategy that uses different levels of theory for different scales of the problem. For example, the environment surrounding a correlated molecule can be treated at a lower theoretical level. Within the molecule, the strongly correlated regions, suitable for a quantum method, can be isolated and *embedded* within a mean-field environment, treated with a classical method. For further details on embedding techniques, see Chapter 3.

Hartree-Fock

A common single-configurational method that is also a prerequisite for many higher-level calculations is Hartree-Fock. In this section, we outline Hartree-Fock in a restricted orbital basis, also called Restricted Hartree-Fock (RHF). Hartree-Fock optimizes the coefficients $C_{\mu p}$ in Eq. (1.6), given a set of N initial orbitals[2] $\{\phi_p^i\}$ and η electrons to return a new set of orbitals $\{\phi_p^{\mathrm{HF}}\}$. Occupying η of these orbitals gives the *lowest energy Slater determinant*. To find these orbitals, we define the so-called Fock operator:

$$f = \sum_{pq} f_{pq} E_{pq} \tag{1.17}$$

$$f_{pq} = \left(h_{pq} + \sum_i (2g_{pqii} - g_{piiq}) \right) \tag{1.18}$$

The coefficients of the Fock operator f_{pq} form the Fock matrix. Note that the sum over i only goes over the specified *occupied* orbitals. This one-body operator describes a system of electrons that not only have kinetic energy and feel a nuclear potential, but also feel all other *occupied* electrons through a Coulomb interaction (g_{pqii}) and an exchange interaction (g_{piiq}). Thus, the Fock operator explicitly depends on the occupied orbitals; each independent particle feels the mean field of the rest of the electrons. Its single-particle eigenfunctions can be found by diagonalizing $f_{pq} = \sum_r U_{pr} \varepsilon_r U_{rq}^\dagger$, obtaining a new set of orbitals $\phi_p \leftarrow \sum_q \phi_q U_{qp}$. In this basis, the single-particle states are eigenstates of the Fock operator with energy ε_k: $f a_{p\sigma}^\dagger |\Omega\rangle = \varepsilon_p a_{p\sigma}^\dagger |\Omega\rangle$. With these new orbitals, we can once again calculate a new Fock matrix using Eq. (1.18), that can be diagonalized once again. Doing this until convergence will yield the lowest energy Slater determinant $\Phi(\{\phi_p\})$, with energy:

$$E_{\mathrm{HF}} \left(\{\phi_p\} \right) = 2 \sum_i h_{ii} + \sum_{ij} (2g_{iijj} - g_{ijji}) \tag{1.19}$$

This cycle of diagonalizing the Fock matrix and calculating orbitals is called Hartree-Fock *Self Consistent Field* (HF-SCF).

The output of the HF procedure is a set of unique orbitals known as *canonical Molecular Orbitals* (CMOs). They have a definite energy ordering, i.e. $\varepsilon_k \geq \varepsilon_{k-1}$.

---

[2] This can be for example a set of orthonormalized atomic orbitals, but can also be something more intricate like eigenfunctions of the one-body part of the Hamiltonian.

Of these orbitals, $\eta/2$ are doubly occupied (given an even amount of electrons). From these orbitals, valuable quantities like the electron density, and properties like excitation energies and electron affinities can be estimated. Each occupied spin-orbital represents an individual electron, which is particularly useful in the chemical interpretation of the result, for example in studying electron transfer reactions. The low-energy orbitals are often core orbitals of the atoms, that do not participate in reactions. Furthermore, the Hartree-Fock energy is invariant under rotations under its occupied orbitals. To see this, consider that the wavefunction is a determinant, and that $\det\{UAU^\dagger\} = \det\{U\}\det\{A\}\det\{U^\dagger\} \simeq \det\{A\}$, thus it is invariant under unitary transformations up to a global phase. These unitary transformations can be used to obtain different orbitals that can be used for different purposes, at the cost of losing the energy ordering of the CMOs. For example, one can minimize a measure of spatial extent to obtain *localized orbitals*. These localized orbitals can represent familiar chemical concepts such as bonding/anti-bonding orbitals and lone pairs (see Section 1.3.2 and Chapter 2 for more information about localization techniques).

Density Functional Theory

DFT takes a different approach than Hartree-Fock, where instead of the wavefunction, the *electron density* is the central variable. The electron density can be defined from a state in first quantization $\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_\eta)$ (that could be a superposition of Slater determinants), as:

$$\rho(\mathbf{r}) = \eta \int \cdots \int |\Psi(\mathbf{r}, m_\mathrm{s}, \mathbf{x}_2, \ldots, \mathbf{x}_\eta)|^2 dm_\mathrm{s} d\mathbf{x}_2 \ldots d\mathbf{x}_\eta \tag{1.20}$$

such that $\int \rho(\mathbf{r})d\mathbf{r} = \eta$. In the electronic structure Hamiltonian of Eq. (1.3), the kinetic energy and the electron-electron repulsion terms are fixed, while the potential U_Ne uniquely defines the system under consideration. Thus, for any arbitrary potential[3] $V(\mathbf{r})$, there is a unique mapping to a ground state wavefunction that is the solution of the Schrödinger equation. From this ground state wavefunction, the electron density $\rho(\mathbf{r})$ can be computed. As a result, there is a unique mapping

$$V(\mathbf{r}) \rightarrow \rho(\mathbf{r}). \tag{1.21}$$

The cornerstones of DFT are two theorems by Hohenberg and Kohn [15]. The first theorem states that the above mapping can be inverted:

$$\rho(\mathbf{r}) \rightarrow V(\mathbf{r}) \ + \ \text{constant}. \tag{1.22}$$

---

[3] The potential can not be completely arbitrary. It has to be local, and additionally it needs to be such that the corresponding Hamiltonian has a (bound) η-electron state.

This has serious implications: a given density uniquely determines the Hamiltonian (up to a constant), and in turn the wavefunction and all properties of the many-body problem. The wavefunction can thus be written as a *functional* of the electron density: $\Psi[\rho(\mathbf{r})]$. As the expectation value of the potential is usually an explicit linear function of the density: $\langle V\rangle = \int V(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}$, the total electronic energy can be written as a functional of the density as follows:

$$E[\rho] = F[\rho] + \int V(\mathbf{r})\rho(\mathbf{r})d\mathbf{r}, \tag{1.23}$$

where we defined the *universal functional* that is independent of the potential:

$$F[\rho] = \langle\Psi[\rho]|T_{\mathrm{e}} + U_{\mathrm{ee}}|\Psi[\rho]\rangle \tag{1.24}$$

The second Hohenberg-Kohn theorem provides us with a variational theorem: minimization over all possible[4] ρ will give the ground state energy: $E_0 = \min_\rho E[\rho]$.

A key challenge remains with this approach: how do we determine an explicit formula for the functional $F[\rho]$? Although Hohenberg and Kohn proved its existence, it must ultimately be approximated through some surrogate model[5]. This approximation is central to DFT, and there is a zoo of functionals out there that can work better for some systems than others [17]. Another obstacle that needs to be overcome is the representation of the density: how do we know it represents a valid η-electron state? Kohn and Sham (KS) [18] provided a framework to perform the density optimization in practice. They decompose the universal functional as:

$$F[\rho] = T_{\mathrm{e}}[\rho] + \frac{1}{2}\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' + E_{\mathrm{xc}}[\rho] \tag{1.25}$$

where the second term is the Hartree term, and $E_{\mathrm{xc}}[\rho]$ is the Exchange-correlation functional that needs to be approximated. This decomposition facilitates different level of approximations for the various contributions to the energy; there is usually a trade-off between the level of approximation in exchange and correlation, while the Hartree part is usually treated exactly. Hartree-Fock can also be described in the framework of DFT; the HF procedure is equivalent to a functional that treats the Hartree and Exchange exactly, while correlation is the part that is missing. To perform the optimization, KS showed that any ground state density can be obtained from a single determinant wavefunction. With an appropriate approximation for the exchange-correlation functional, this reduces the problem to a set of SCF equations for the orbitals, exactly like in Hartree-Fock. Thus, all the single-configurational machinery and chemical interpretation can be used in

---

[4] The catch is that the minimization has to be done over densities that result in a local potential (v-representable) that in turn define an η-electron quantum state (N-representable)

[5] It has even been shown that the evaluation of an exact universal functional would need to be computationally hard, for a classical and for a quantum computer alike. [16].

DFT, additionally contributing to its usefulness.

1.3 Active Space Methods

1.3.1 Construction of the Hamiltonian

To capture the full electronic structure problem of a molecular system within the pre-defined basis-set, one must consider a complete set of electronic configurations. The Full Configuration Interaction (FCI) method solves the electronic structure problem by expanding the wave function in a linear combination of all possible Slater determinants:

$$|\Psi\rangle = |\Phi_0\rangle + \sum_{ia} c_i^a \left|\Phi_i^a\right\rangle + \sum_{ijab} c_{ij}^{ab} \left|\Phi_{ij}^{ab}\right\rangle + \dots \tag{1.26}$$

where we defined $|\Phi_0\rangle$ as the Hartree-Fock determinant, $\left|\Phi_i^a\right\rangle$ as a single excitation from the occupied orbital i to the unoccupied orbital a, and $\left|\Phi_{ij}^{ab}\right\rangle$ as a double excitation from the occupied orbitals i and j to the unoccupied orbitals a and b. To find this ground state, one has to diagonalize the Hamiltonian matrix in the basis $\{|\Phi_0\rangle, \left|\Phi_i^a\right\rangle, \left|\Phi_{ij}^{ab}\right\rangle, \dots\}$ for the lowest eigenvalues and vectors. Clever ways of solving this problem without constructing the complete Hamiltonian matrix have been conceived, the main workhorse being Davidson diagonalization that solves the problem for the lowest eigenvalues [19]. While the FCI method provides the exact solution to the electronic structure problem, the state in Eq. (1.26) has $\binom{2N}{\eta}$ configurations, such that even only storing the CI-coefficients in Eq. (1.26) requires exponential classical memory. This makes the FCI method only feasibly for very small systems, typically of the order of tens of spatial orbitals[6].

To still be able to treat larger systems, one can restrict the number of configurations in Eq. (1.26) in multiple ways. One is to truncate the CI expansion by considering only single and double excitations, leading to the Configuration Interaction Singles Doubles (CISD) method [21]. Alternatively, the typical approach in multi-reference quantum chemistry is the use of an active space. In active space methods, the total orbital space is divided up in three subspaces: the core space, the active space and the virtual space, indicated in this thesis with following $i, j, k, l \dots$ t, u, v, w and a, b, c, d respectively – see Figure 1.1. Consequently, in a CI-expansion of an active space wavefunction the core orbitals are always doubly occupied, and the virtual orbitals are always unoccupied. Projecting the total Hamiltonian in Eq. (1.14), defined on all orbitals, onto this restricted space, results

---

[6] The largest exact diagonalization results known to the author were for an electronic structure Hamiltonian that had 24 spatial orbitals and 24 electrons, corresponding to a CI expansion of around one trillion Slater determinants [20].

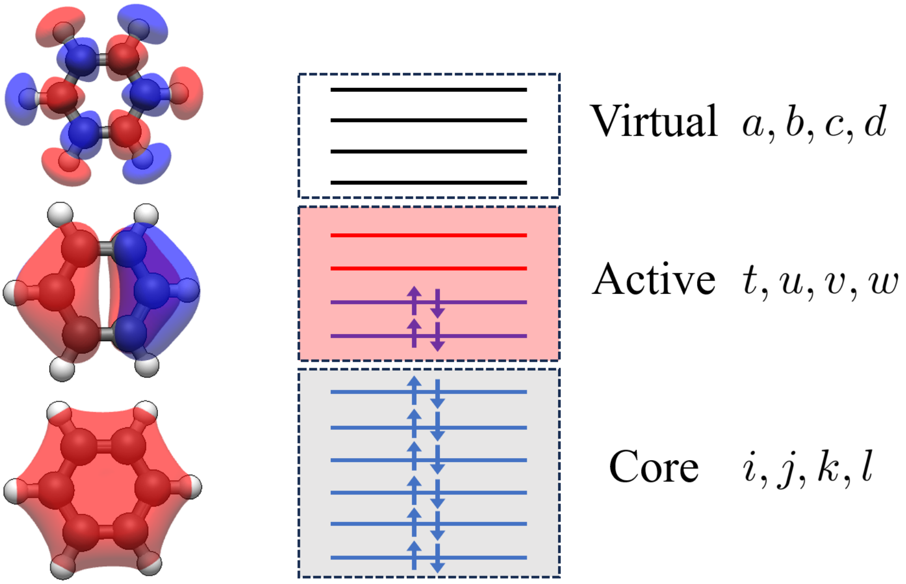


Figure 1.1: Schematic picture of an active space in a molecule. The core orbitals are doubly occupied and the virtual orbitals are unoccupied. On the left, typical examples of all three classes or orbitals are portrayed for the benzene molecule (C_6H_6). A low-lying σ-bonding orbital is always doubly occupied, thus it can go in the core space. In contrast, a high-lying σ^* orbital can be put in the virtual space. The HOMO, which is the π bonding orbital should go in the active space (together with its anti-bonding counterpart).

in:

$$H_{\text{act}} = \sum_{tu} \left[h_{tu} + \sum_{i} (2g_{tuii} - g_{tiiu}) \right] E_{tu} + \frac{1}{2} \sum_{tuvw} g_{tuvw} e_{tuvw} + E_{\text{core}} \tag{1.27}$$

where the adjusted one-electron integrals are due to the effective potentials of the doubly occupied core orbitals, and $E_{\text{core}} = 2\sum_i h_{ii} + \sum_{ij} 2g_{iijj} - g_{ijji}$ is their mean-field energy.

In a Complete Active Space Configuration Interaction (CASCI) calculation, the active space Hamiltonian in Eq. (1.27) is solved with FCI. Once the problem is solved, one can get useful quantities out of a CASCI calculation at no additional cost, by using the wavefunction to calculate expectation values of operators, for example to get molecular properties. The CASCI method provides a more accurate description of the electronic structure of a molecule compared to the Hartree-Fock method, as it includes a larger number of configurations while still being computationally feasible for bigger systems. However, the cap on the active space size, due to the exponential scaling of FCI, still remains.

As elaborated on in Section 1.4, quantum computers are a promising tool to overcome the limitations of active space size, by providing a more space-efficient way to simulate the electronic structure of molecules with larger active spaces. There are also classical methods that substitute the exact diagonalization by an approximate method with a more favourable cost. Some examples are adaptive sampling CI methods [22], coupled-cluster methods [23], tensor network algorithms like Density Matrix Renormalization Group (DMRG) [24], or Quantum Monte Carlo (QMC) [25].

1.3.2 Choosing consistent orbitals

The limited size of the active space is not the only limitation of CASCI. Another limitation comes from the orbitals used to construct the active space, as an accurate description of the PES is very sensitive to this set of reference MOs. In this section we will discuss different methods to obtain a robust orbital basis, e.g. by picking them from a localized basis based on chemical intuition or by minimizing the energy of the active space wavefunction w.r.t. orbital rotations.

Selection of active space

The energy of the Hartree-Fock orbitals can help to give a handle on what orbitals should be in the active space. As either excitations of electrons from low-lying core orbitals to virtuals, or from core orbitals to high lying virtuals pay a high price in energy, these determinants can be expected to have small coefficients in the FCI expansion. Therefore, we can select the active space based on orbitals around the Highest Occupied MO (HOMO) and the Lowest Unoccupied MO (LUMO). This will generally be the orbitals that contribute most to the correlations in the system. However, this approach has some drawbacks: while the Hartree-Fock orbitals might give the lowest-energy solution for a single determinant (see Section 1.2.3), this does not guarantee the lowest energy solution for the active space Hamiltonian in Eq. (1.27).

In addition, an active space selection is often guided by chemical intuition, where the active space is chosen to include the orbitals that are most relevant for the chemical properties of the molecule. For example, in a molecule with a metal center, the active space might include the metal d-orbitals and the ligand orbitals that interact with the metal center. The delocalized nature of the CMOs coming from Hartree-Fock makes it challenging to select them in the above fashion. Instead, localization procedures can be used to identify the bonding orbitals that are known to be important in the reaction. Chapter 2 and 3 contain a more in-depth discussion about localization schemes.

Another problem of Hartree-Fock orbitals arises when computing active spaces along a reaction pathway. Namely, the Hartree-Fock minimum in orbital space might be discontinuous along the reaction coordinate, especially in bond breaking reactions [21, 26]. Different orbital basis might overcome this inconsistency, by for example projecting the orbitals from a previous calculation onto the next geometry to enforce continuity (see chapter 5 for an implementation).

Orbital transformations

To address the limitations of Hartree-Fock orbitals, the Complete Active Space Self-Consistent Field (CASSCF) method was conceived [27]. This method introduces an additional orbital optimization to minimize the energy of the active space wavefunction. In practice, a unitary orbital rotation is parameterized by an exponential operator $e^{-\hat{\kappa}}$ acting on $|\Psi\rangle$ where

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} E^{-}_{pq} \tag{1.28}$$

where κ_{pq} (in the following indicated in matrix form as $\boldsymbol{\kappa}$) is an antisymmetric matrix uniquely determined by $N(N-1)/2$ real parameters, and $E^{-}_{pq} = E_{pq} - E_{qp}$. Conveniently, the expectation value of the new state will be given by

$$E(\boldsymbol{\kappa}) = \langle\Psi|e^{\hat{\kappa}} H e^{-\hat{\kappa}}|\Psi\rangle = \langle\Psi|H(\boldsymbol{\kappa})|\Psi\rangle \tag{1.29}$$

where the transformed Hamiltonian $H(\boldsymbol{\kappa}) = e^{\hat{\kappa}} H e^{-\hat{\kappa}}$ can be easily calculated by transforming the molecular integrals. To see this, we make use of the Baker-Campbell-Hausdorff (BCH) formula, which states that for two operators A and B:

$$e^{A} B e^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \dots \tag{1.30}$$

when one works out the commutators in Eq. (1.30), one finds that the creation and annihilation operators transform as a Bogoliubov transformation: $\tilde{a}^{\dagger}_p = \sum_q a^{\dagger}_q \left[e^{-\boldsymbol{\kappa}}\right]_{qp}$ and $\tilde{a}_p = \sum_q a_q \left[e^{-\boldsymbol{\kappa}}\right]_{qp}$ respectively[7]. Thus, for the energy, performing an orbital transformation on the state is equivalent to transforming one- and two-electron integrals as $h_{pq} \to \tilde{h}_{pq}(\boldsymbol{\kappa})$, $g_{pqrs} \to \tilde{g}_{pqrs}(\boldsymbol{\kappa})$, where **every** index is transformed with the unitary matrix $\left[e^{-\boldsymbol{\kappa}}\right]_{pq}$. This is equivalent to rebuilding the active space Hamiltonian with the new orbitals $\tilde{\phi}_p$:

$$\tilde{\phi}_p(\boldsymbol{\kappa}) = \sum_q \phi_q \left[e^{-\boldsymbol{\kappa}}\right]_{qp}. \tag{1.31}$$

The energy of the active space Hamiltonian in Eq. (1.27) can then be minimized by optimizing the parameters of the unitary transformation $\boldsymbol{\kappa}$. The exponential

---

[7] A closely related result is the Thouless theorem [28], that says that we can generate any single-determinant wavefunction from any other by an orbital rotation $e^{-\hat{\kappa}}$ for $\hat{\kappa}$ defined as Eq. (1.28).

form of the unitary transformation ensures a straightforward way to calculate the analytical gradient and Hessian of the energy with respect to $\boldsymbol{\kappa}$, which can be used in fast optimization algorithms such as the Newton-Raphson method. To calculate the analytical gradient of the energy $E(\boldsymbol{\kappa})$, we compare the BCH expansion of Eq. (1.30) with the Taylor expansion of the $E(\boldsymbol{\kappa})$ in $\boldsymbol{\kappa}$ and equate orders. We find consequently that:

$$\frac{\partial E}{\partial \kappa_{pq}} = \langle \Psi | \left[E^{-}_{pq}, H \right] | \Psi \rangle \tag{1.32}$$

$$\frac{\partial^2 E}{\partial \kappa_{pq} \partial \kappa_{rs}} = \frac{1}{2} \left(1 + P_{pq,rs} \right) \langle \Psi | \left[E^{-}_{pq}, \left[E^{-}_{rs}, H \right] \right] | \Psi \rangle \tag{1.33}$$

where $P_{pq,rs}$ is the permutation operator that exchanges the indices pq and rs. It turns out that these expectation values of commutators can be conveniently expressed in linear combinations of integrals and 1- and 2-particle Reduced Density Matrices (1- and 2-RDMs), see Appendix 5.C in chapter 5 for more details. This is practical for the optimization of the orbitals, as the 1- and 2-RDMs can be calculated efficiently in most multi-configurational solvers and stored more efficiently than the complete CI-vector.

(SA-)CASSCF

Combining the orbital optimization with the CASCI method results in the Complete Active Space Self-Consistent Field (CASSCF) method [29]. In the usual two-step optimization, the orbitals that span the active space are optimized to minimize the energy of the active space wavefunction using the gradient and Hessian in Eq. (1.32)-(1.33). Next, the active space Hamiltonian is rebuilt with the new optimized orbitals and its low-lying states are computed with CASCI. This goes on until convergence of the energy. In this manner, the method becomes more flexible and can capture a wider range of electronic correlations than CASCI. At the same time, the method is less sensitive to the input orbitals, assuming it can find a good minimum in the orbital space, making it more of a *black box*.

When considering excited states, the orbital-optimization procedure might be even more important, as the active space that is optimal for the ground state might not be optimal for the excited states. To address this, the State-Averaged Complete Active Space Self-Consistent Field (SA-CASSCF) method was introduced, which optimizes the orbitals to minimize the energy average of the states [21, 31]. The state-averaged optimization procedure can provide a more accurate description of the electronic structure, especially when there are low-lying states that are close in energy. This is especially important if the states exhibit a topologically protected crossing, called a *conical intersection*. For more information about conical intersections, see chapter 4 and 5. As readily seen in Figure 1.2, an unequal treatment of the states, in this case with an active space based on Hartree-Fock orbitals, does not correctly describe the topology: there

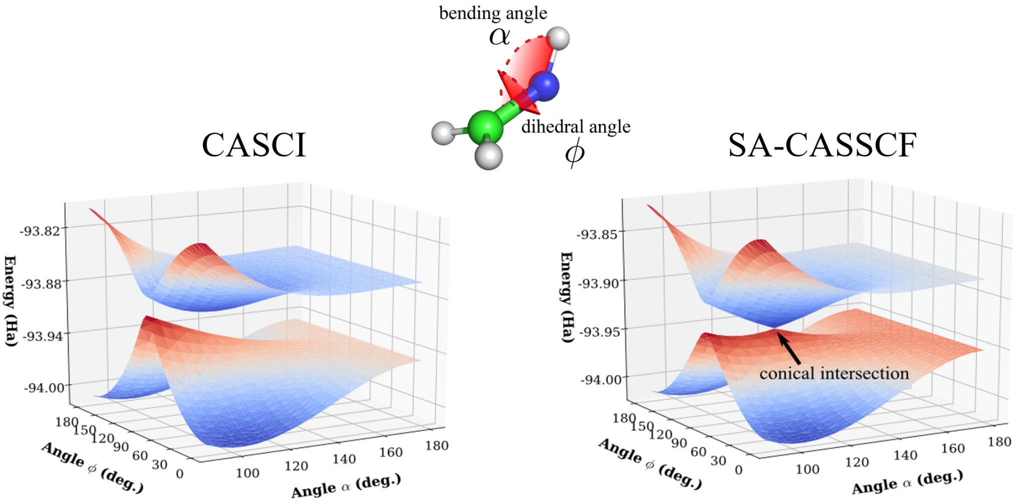


Figure 1.2: Potential energy surfaces of the formaldimine molecule, shown above, for two different active space methods. On the left is CASCI, where the active space Hamiltonian is constructed through Hartree-Fock orbitals. On the right is SA-CASSCF, where the orbitals are optimized for the average of the ground and excited state energies. CASCI does not capture the correct topology of the surface, instead we need the orbital optimization to model the conical intersection. Adapted from [30].

is no conical intersection between the two low-lying states. By optimizing the orbitals democratically, the correct topology of the PES is captured, and the conical intersection is found. The SA-CASSCF method has been widely used to study photochemical systems [32] and other molecular systems with multiple low-lying states [33].

1.3.3 Static and dynamic correlations

The variational theorem says that the ground state has a lower energy than any other normalized state. This means that the Hartree-Fock state is always higher in energy than the ground state. The missing energy is called the *correlation energy*:

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}. \tag{1.34}$$

Thus, by its quantum chemical definition, it precisely describes the deficiency of the best single Slater determinant compared to the exact wavefunction (exact within a given basis set). In post-HF calculations like active space methods, the target is to retrieve as much as possible of the correlation energy as possible. Chemists generally differentiate between two different sources of correlation energy:

static correlation and *dynamic correlation*. While they are not mathematically rigorously defined, they give a handle on the strong and weak points of post-HF methods.

Static correlation is coming from the inability of a mean-field method to describe a superposition of multiple determinants of similar weight. The entanglement between electrons can give rise to long range correlations; thus static correlation is associated with long range interactions. The presence of static correlation in the electronic structure of the molecule is important in the case of near-degeneracies, such as in bond-breaking reactions, transition metal complexes, and other systems with low-lying excited states. To describe static correlation in these systems, multi-configurational methods are a crucial tool. Active space methods are the most common method to deal with static correlation, as the exact diagonalization procedure is able to support any amount of states with similar weight within the limited size active space.

The other source of error in a single determinant is dynamical correlation, that is associated with the close-range behavior of the electrons. A mean-field description namely fails to describe correlated motion of the electrons. The exact ground state of two electrons $\Psi(\mathbf{r}_1, \mathbf{r}_2)$ has an asymptotic behaviour at close distances $\mathbf{r}_1 \approx \mathbf{r}_2$, because of the blowing up of the denominator in the Coulomb repulsion. This is also refered to as the electron-electron cusp. While it is impossible to describe this behaviour exactly in second quantization with a finite basis set, a single determinant is especially insufficient. Instead, we can approximate it by having a lot of determinants with small weights spanned on all spin-orbitals. This behaviour can be retrieved efficiently by using perturbative methods like coupled-cluster, or Møller-Plesset perturbation theory [21, 34]. An active space method is typically lacking in its description of dynamical correlations. Methods like Complete Active Space 2nd order Perturbation Theory (CASPT2) and its extensions try to model excitations from the core and to the virtuals perturbatively to retrieve the dynamical correlation after an active space calculation [35–37].

The contribution of the correlation energy can be extremely important in some systems, especially in open-shell transition metal clusters that have multiple low-lying excited states. Understanding molecular systems with strong electron correlations could have a high real-world impact. For example, *Nitrogenase*, the key enzyme responsible for converting atmospheric nitrogen (N_2) into ammonia (NH_3), involves complex metal clusters, including the iron-molybdenum cofactor (FeMoco). Classical algorithms struggle with ground state estimation of FeMoco because of the high amount of both static and dynamic correlations in this system [38, 39]. Modelling the correlations correctly could help explain how electrons and protons are transferred within the FeMoco to break the strong triple bond of nitrogen molecules. This detailed understanding is essential for improving synthetic catalysts and for biotechnological applications in sustainable agriculture.

1.4 Quantum Algorithms for Chemistry

Single Slater determinant methods like Hartree-Fock are space-efficient on classical computers, as they can be represented using $2N$ classical bits. However, proper eigenstates of this Hamiltonian are of the form of Eq. (1.26), and need to be represented by their full CI-vector that contains $\binom{2N}{\eta}$ floating point numbers. A quantum computer replaces classical bits by *qubits*. Qubits live in a finite-dimensional Hilbert space, and can be in any normalized complex linear combination of 0 and 1: $\alpha\,|0\rangle + \beta\,|1\rangle$, where $|\alpha|^2 + |\beta|^2 = 1$. The basis states can be represented by the computational basis $|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ and $|1\rangle = \begin{pmatrix}0\\1\end{pmatrix}$. We can build a multi-qubit state by taking a tensor product basis of these computational basis states:

$$|\psi\rangle = c_0\,|0\ldots00\rangle + c_1\,|0\ldots01\rangle + \ldots + c_{2^N}\,|1\ldots11\rangle \tag{1.35}$$

which can be written as a 2^N-dimensional vector.

From Eq. (1.35) it is apparent that we can represent a state with 2^{2N} amplitudes, which is bigger than $\binom{2N}{\eta}$ that we need for an CI state, *exactly* with just $2N$ qubits, thus achieving an exponential reduction in space. However, this does not mean the quantum computer will also be exponentially faster in time, implying a quantum advantage. In fact, it has been shown that there exist Hamiltonians of the form Eq. (1.14) for which it is exponentially hard to prepare the ground state even on a quantum computer [40], assuming widely held conjectures in complexity theory[8]. Yet, this is not a reason to abandon all hope; the statement is only true for the worst-case Hamiltonians. As we have seen, classical heuristics are often sufficiently accurate for chemistry systems, and similarly we expect quantum algorithms to help us find accurate enough ground states in some cases. Indeed, the complexity result is not incredibly surprising: we know that there exists systems in nature that do not reach their ground state, like spin-glasses and other frustrated systems. This would be targets where it might be difficult to find the ground state for any type of computer. For natural Hamiltonians in chemistry, there is much larger chance that the ground state can be prepared efficiently using heuristic methods [42].

A quantum algorithm is implemented as a sequence of unitary operations on the qubit registers, that perform specific operations like rotations, bit flips or controlled-NOTs (cNOTs). We describe a quantum algorithm as by a classical description of this sequence of operations. This description is conveniently expressed as a *quantum circuit*. In a quantum circuit, qubits are represented as

---

[8] This result has been formalized in quantum complexity theory. Finding the ground state of Eq. (1.14) is *QMA hard*, which means that it is harder than any problem in the QMA class (Quantum Merlin Arthur). QMA is the complexity class that contains all problems that can be *verified* efficiently on a quantum computer, i.e. the analog of the classical NP [41]. It is strongly believed that verifying problems is much easier than solving it, thus we expect that a QMA-hard problem will be exponentially difficult in general.

lines and quantum gates are represented as boxes that act on one or more qubits at a time. Time flows from left to right. It provides a visual representation of the quantum algorithm and allows for the analysis of the quantum state of the system at different points in the computation (see Figure 1.3). To retrieve information out of the quantum computer, a measurement is done at the end of the computation. The Born rule tells us that if we measure the state Eq. (1.35) in the computational basis, we get as outcome a bit string $|i\rangle$ with probability $|c_i|^2$. We can also measure in another basis by doing a transformation to the another basis. In this manner we can measure any *Pauli string*, i.e. tensor product of Pauli matrices, that are defined as follows:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.36}$$

1.4.1 Paradigms of quantum computing

There are several types of radically different technologies that can be used to store quantum information and implement quantum algorithms. Some examples of physical systems proposed and developed to implement these technologies are superconducting qubits, trapped ions, and topological qubits. Superconducting qubits are solid-state devices that use superconducting electronic circuits to encode quantum information, while trapped ions use laser-cooled ions to encode qubits. Topological qubits are a type of qubit that relies on topological properties of the system to protect the quantum information from errors. Each type of quantum device has its own advantages and disadvantages, such as coherence times, error rates, and scalability. Superconducting qubits are currently the most widely used type of quantum device and are being developed by companies such as IBM [43] and Google [4].

Today's quantum computations are characterized by noise, and exponential decoherence of the quantum state with respect to the length (or *depth*) of the quantum circuit [44, 45]. To combat the noise in devices, quantum computation can be made *fault-tolerant* by storing quantum information redundantly in multiple physical qubits, that together form a *logical qubit*. In this Fault Tolerant Quantum Computation (FTQC) era, the computer will periodically measure some of the qubits to detect and consequently correct errors; we call this *quantum error correction* (QEC). For this process to correct errors faster than they happen, the error probability of a physical qubit per unit of time must be below some constant *threshold* p_{th}, depending on the QEC code. Once it is below this threshold, the error can be made arbitrarily small by concatenating multiple logical qubits into one. The realization of such error corrected devices is the only way forward to implement pivotal algorithms like Shor's algorithm [46, 47] and quantum phase estimation [48, 49] (see Section 1.4.3).

As of today, there has been considerable efforts towards FTQC, with multiple first demonstrations of logical qubits that operate below the quantum threshold [50–

52]. However, estimating the timeline for the realization of a fully error-corrected device is extremely challenging, and might be on the order of decades. This is why a big part of the quantum computing community focuses on how we can make use of quantum devices in the near term. In this current *Noisy Intermediate Scale Quantum* (NISQ) era [53], the physical qubits are used directly. This implores us to keep computations on the device extremely short to avoid disruptions from accumulated noise. To achieve this, every NISQ algorithm generally follows these principles: prepare a state using a short-depth circuit, sample an observable, and repeat. To mitigate the noise, NISQ needs *error mitigation* protocols, that further increase the amount of samples needed to filter out circuit noise.

It is important for any quantum algorithm designer to keep in mind the intended architecture of the quantum device, especially to differentiate between the NISQ or FTQC paradigm. In FTQC, because of the way the error correction works, only certain operations can be implemented fault-tolerantly. This discrete set of operations comprises a *gate set*, and it is essential for FTQC algorithms to consider an efficient decomposition in terms of these gates. In NISQ, the set of allowed operation depends on the details of the device, rather than the details of a QEC code. Therefore, the way to calculate cost in the two cases is different: in NISQ the two-qubit gates are expensive, while in FTQC, small Pauli rotations are expensive, while e.g. cNOTS are cheap. In contrast to FTQC, NISQ imposes a tight budget on the depth of the circuit before decoherence of the state. Additionally, measuring an observable will imply a substantial sampling cost, that has to be factored in when estimating the total complexity of the algorithm.

While the limited capabilities of NISQ are evident, it has been demonstrated that some computations nevertheless can be beyond capabilities of classical computers [4–6]. A specific class of algorithms that try to achieve quantum utility with short depth computations are *hybrid quantum-classical algorithms*. They trade off many short circuit evaluations on a QPU while offloading part of the computation to a CPU. A subset of these hybrid algorithms are Variational Quantum Algorithms (VQAs), where a set of parameters are variationally optimized to find the minimum of some cost function. VQAs encompass a family of algorithms that share the common feature of a *Parameterized Quantum Circuit* (PQC), a short depth circuit containing parameterized quantum gates. We express the vector of variational parameters as the set of classical numbers $\boldsymbol{\theta}$. The PQC then implements the parameterized operation $U(\boldsymbol{\theta})$.

1.4.2 The Variational Quantum Eigensolver

The most common type of VQA is called the Variational Quantum Eigensolver (VQE) introduced by Peruzzo et. al. [54], that tries to estimate the ground state energy of some Hamiltonian H by preparing a state $|\psi(\boldsymbol{\theta})\rangle$ with a PQC and minimizing its parameters. It exists of three steps:

- The variational state $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi_0\rangle$ is prepared on the quantum device.

- The expectation value of the Hamiltonian is measured on the state $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$.
- The result is fed back to the CPU, that uses a classical optimizer that updates the parameters $\boldsymbol{\theta}$ to minimize the energy $E(\boldsymbol{\theta})$.

This cycle is repeated until convergence of the energy. The variational theorem gives a promise that the end result is bounded from below by the ground state energy of H, i.e. $\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \geq E_0$. A schematic of the algorithm is shown in Figure 1.3.

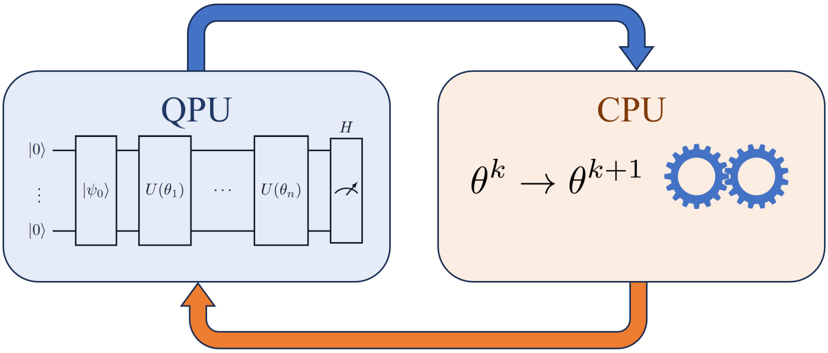


Figure 1.3: Schematic of the VQE algorithm. In the blue box, the ansatz wave function is being prepared on the QPU by first preparing the initial state $|\psi_0\rangle$, and subsequently acting with the unitary $U(\boldsymbol{\theta})$. In this example, unitary depends on n different parameters. After the initialization and unitary evolution, the Hamiltonian is measured by sampling Pauli strings on the state. This has to be repeated many times. In the k-th iteration, this allows us to estimate an energy dependent on $\boldsymbol{\theta}^k$. The classical optimizer proposes a new set of parameters $\boldsymbol{\theta}^{k+1}$, which are then passed on to the quantum computer to evaluate the new energy. This cycle is repeated until some convergence condition is met.

Measurements

In this section, we go into detail how to measure the fermionic Hamiltonian of Eq. (1.14) on a quantum computer. As we can natively only measure Pauli operators on qubits, we will have to map the Hamiltonian of Eq. (1.14), which is a linear combination of fermionic creation and annihilation operators, to a linear combination of Paulis. We call this a *Fermion-to-qubit transformation*:

$$H = \sum_{pq} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} e_{pqrs} \xrightarrow{\text{F-Q}} \sum_i h_i P_i, \tag{1.37}$$

where $P_i \in \{I, X, Y, Z\}^{\otimes N}$ are tensor products of Pauli operators called Pauli strings. The coefficients h_i are determined by the transformation and the integrals of the Hamiltonian.

To construct this mapping, we study the representation of a second quantized state (Eq. (1.11)) on the quantum computer. The most straightforward way is by identifying the $2N$ spin-orbitals directly as $2N$ qubits, thus the second quantized state directly as a qubit state $|f_1, \ldots, f_{2N}\rangle$. To identify the mapping, let us recall how fermionic raising operators act on states in second quantization, using the commutation relationships (1.10):

$$\begin{aligned} a_k^\dagger |f_1, \ldots, f_{2N}\rangle &= a_k^\dagger {a_1^\dagger}^{f_1} {a_2^\dagger}^{f_2} \ldots {a_{2N}^\dagger}^{f_{2N}} |\Omega\rangle \\ &= (-1)^{\sum_{j=1}^{k} f_j} {a_1^\dagger}^{f_1} \ldots a_k^\dagger {a_k^\dagger}^{f_k} \ldots {a_{2N}^\dagger}^{f_{2N}} |\Omega\rangle \\ &= \delta_{f_k,0} (-1)^{\sum_{j=0}^{k} f_j} |f_1, \ldots, 1, \ldots, f_{2N}\rangle \end{aligned} \tag{1.38}$$

Thus, it is readily seen that we need to perform two operations: we convert the k-th qubit from $|0\rangle$ to $|1\rangle$ and take into account the phase factor $(-1)^{\sum_{j=0}^{k-1} f_j}$. The first is easily expressed in Pauli operators: $|1\rangle\langle 0| = \frac{1}{2}(X - iY)$. Additionally, we can account for the phase acting with Pauli-Z operators on the first $k-1$ qubits, arriving at the following transformation:

$$\begin{aligned} a_k^\dagger &= Z_1 \otimes \ldots \otimes Z_{k-1} \otimes \frac{1}{2}(X_k - iY_k) \otimes \mathbb{1}_{k+1} \otimes \ldots \otimes \mathbb{1}_{2N} \\ a_k &= Z_1 \otimes \ldots \otimes Z_{k-1} \otimes \frac{1}{2}(X_k + iY_k) \otimes \mathbb{1}_{k+1} \otimes \ldots \otimes \mathbb{1}_{2N} \end{aligned} \tag{1.39}$$

where we took the Hermitian conjugate to retrieve the last line. This transformation was conceived in 1928 by Jordan and Wigner [55] and is called the Jordan-Wigner transformation.

While it is perhaps the most intuitive way to map fermionic operators to qubit operators, it may not be the optimal one. The *k-locality* of the Pauli strings is defined by the number of qubits they act on non-trivially. This k-locality is important for many algorithms, notably in prominent Hamiltonian time-evolution algorithms, where a high k amounts to a higher cost. As readily seen from Eq. (1.39), the string of Pauli-Z operators make that the k-locality after a Jordan-Wigner transformation is linear in the number of orbitals. Other common transformations, like the Bravyi-Kitaev mapping, can reduce this to $\mathcal{O}(\log(N))$ [56–58]. Nevertheless, the Jordan-Wigner transformation is still most often used, both because of its directly interpretable states, and geometric locality of Hamiltonians. Geometric locality is the physical property of interactions inside a Hamiltonian: a term $h_{pq} a_p^\dagger a_q$ will be very small if the distance (in some metric) between orbital p and q is very large. Thus, terms with a large Pauli-Z string are less important, and can sometimes be neglected completely.

To calculate the energy, we can make use of the linearity of the expectation value of the Hamiltonian:

$$E(\boldsymbol{\theta}) = \langle\Psi(\boldsymbol{\theta})|\sum_i h_i P_i|\Psi(\boldsymbol{\theta})\rangle = \sum_i h_i \langle\Psi(\boldsymbol{\theta})|P_i|\Psi(\boldsymbol{\theta})\rangle\,. \tag{1.40}$$

To measure P_i on the quantum computer, we can measure our qubits in the according X- Y- or Z-basis. This is a stochastic event; if we measure all qubits we will get a single correlated bit string corresponding to the projective measurement. We will have to repeat the state preparation and measurement many times to get a statistical average to obtain an estimation of the energy $\tilde{E}$. Call $m_j^{(i)}$ the j-th measurement outcome, which is a product of the relevant measurement outcomes in the correct basis. Then $\tilde{P}_i = \sum_j \frac{m_j^{(i)}}{M_i}$ is an (unbiased) estimator of P_i, where M_i is the shot budget for term P_i. To determine how many total measurements $M = \sum_i M_i$ we need to take to estimate each the energy to precision ϵ we can use propagation of variance:

$$\mathrm{Var}(\tilde{E}) = \sum_i |h_i|^2 \mathrm{Var}(\tilde{P}_i) \leq \sum_i \frac{2|h_i|^2}{M_i} \tag{1.41}$$

Thus the expected energy error is bounded as $\epsilon \leq \sqrt{\sum_i \frac{2|h_i|^2}{M_i}} \leq \mathcal{O}(1/\sqrt{M})$. The scaling of the precision with the inverse square root of the total amount of measurements is a fundamental restriction called the *sampling limit*. Working within the sampling limit, there is still a lot of room for improvement in the amount of measurements by adjusting shot allocations between the terms in the Hamiltonian[9]. For example, we might want to attribute more shots to Pauli strings with a large coefficient [60]. Additionally, commuting Pauli terms can be measured at the same time [61], further reducing the cost of sampling the Hamiltonian.

Variational Ansätze

The structure of the PQC is called a variational ansatz. The variational ansatz determines the ability of the quantum circuit to represent eigenstates of the Hamiltonian in question, and is thus a crucial part of the VQE. The ansatz should be expressive, and at the same time have a short runtime. Variational ansatz families can be broadly categorized into two classes: the hardware efficient ansatz and the physically inspired ansatz. The latter, exemplified by the Unitary Coupled Cluster (UCC) ansatz [54, 62], is shown to possess more advantageous properties due to its higher expressiveness and consideration of essential system

---

[9] The Hamiltonian in Eq. (1.14) contains $\mathcal{O}(N^4)$ terms. However, this can be reduced by compressing the integrals using their symmetries or by truncation techniques like density fitting [11, 59].

properties, such as symmetries. However, physically inspired ansätze often lack a flexible method for systematically improving accuracy by adjusting the number of layers. Additionally, they can require high coherence times that are currently beyond the reach of near-term quantum devices.

To mitigate these issues, most VQE experiments on quantum devices have utilized hardware efficient ansätze. These ansätze feature a tunable number of gate layers that are natively implemented on the quantum devices, maximizing efficiency. Despite this, they present their own challenges, as they can require a large amount of layers to reach the ground state. Moreover, they often fail to preserve the symmetries of the Hamiltonian in question, like the number of particles N, total spin S^2 or its projection along the z-axis S_z. Consequently, during optimization, the solution can converge to a different symmetry sector than the one of interest.

Designing an ansatz that can efficiently represent the ground state of the system is still an active area of research [7]. The literature has explored middle ground solutions between hardware-efficient and physically inspired ansätze, that combine aspects of both classes. These can be based of the problems Hamiltonian, like the Variational Hamiltonian Ansatz (VHA) [63], or the symmetries of the Hamiltonian, like the Number Preserving Fabric (NPF) ansatz [64]. The latter is used in multiple chapters in this thesis, as it is expressive, preserves essential symmetries of the system (N, S^2, and S_z), and allows for a tunable number of layers.

Optimization

The optimization of the parameters in the PQC is a crucial step in the VQE algorithm. For the optimization we make use of classical optimization algorithms, that update the parameters iteratively to minimize the cost function. Some of these optimizers are *gradient-free*, and others might need the gradient of the cost function, that can be measured on the quantum device in exchange for extra sampling cost [65, 66]. There are multiple challenges in the optimization procedure of the VQE. The optimization process can be computationally expensive, especially for large systems with many parameters. The cost-function landscape might be highly non-convex, so it might be easy to get stuck in some local minimum. Furthermore, it has been shown that especially for unstructured ansätze, the optimizer can get stuck on *barren plateaus* where the norm of the gradient vanishes [67]. Finally, an optimizer that can deal with some degree of noise is imperative in VQE optimization. A comparison of different optimizers for chemistry problems can be found in Ref. [68].

Limitations of VQE

While the VQE exploits the CPU to minimize the quantum resources needed, there are caveats to this approach. There are different sources of error in the final estimation of the ground state energy:

- **Algorithmic errors**: the VQE is a heuristic algorithm and there is no guarantee that the algorithm will converge to the global minimum of the energy. This stems from both the expressibility of the PQC ansatz and the optimization procedure that can get stuck in local minima.

- **Circuit noise**: While the VQE leverages low circuit depths, the quantum gates in the PQC are still imperfect and will introduce errors in the quantum state because of the limited coherence time in the device. This noise can severely affect the accuracy of the computed energy and limit the size of the circuits.

- **Sampling noise**: the expectation value of the Hamiltonian is estimated by sampling the quantum state multiple times. To suppress this statistical error in the measurement to an acceptable level can take an excessive amount of shots depending on the Hamiltonian.

The culmination of these different error sources makes it uncertain whether the VQE can achieve a practical quantum advantage. It is highly debated that the naive application to ground state energies in chemistry is the ultimate target for VQE; we need a too high accuracy to effectively beat competing classical methods. This means that it is crucial to work on both the chemistry side, by identifying problems that are harder for classical methods, and on the quantum algorithms side, by making them more efficient to work within the constraints of near-term devices.

1.4.3 Beyond near-term quantum algorithms

In contrast to NISQ, in the FTQC era, ground state energies are believed to be the one of the most important targets for quantum computational chemistry. Most popular approaches to fault-tolerant energy estimation of many-body Hamiltonians center around the Quantum Phase Estimation (QPE) algorithm [69]. In the QPE algorithm consists of three steps: the preparation of an initial state $|\psi_0\rangle = \sum_k c_k |E_k\rangle$ (written here as a linear combination the qubit Hamiltonian eigenstates, i.e. $H|E_k\rangle = E_k|E_k\rangle$), evolution of the Hamiltonian $U = e^{-iHt}$ controlled on one or multiple ancilla qubits and finally measurement of the ancillas (possibly in Fourier basis). The result is stochastic: we measure $e^{i\tilde{\phi}}$, where $\tilde{\phi}$ is an estimate of $E_k t$, with probability $p_k = |c_k|^2$ to a certain guaranteed precision ϵ. This precision depends on the details of the QPE implementation, but it can be shown that we can go beyond the sampling limit: pushing down the scaling as $\epsilon \sim \mathcal{O}(1/D)$, where D is the total cost of the QPE runs. This is an absolute speed limit, referred to as the Heisenberg limit [70]. In summary, we can efficiently sample from the eigenbasis of the Hamiltonian with a fault-tolerant quantum computer.

The requirements on the depth of QPE circuits are beyond near-term devices, so a fault-tolerant device is necessary. The two bottlenecks of the canonical QPE algorithms are the initial state preparation and the Hamiltonian time evolution. While there has been much work to make the latter problem more efficient, the former has been largely ignored until recently [42, 71]. As the probability of the QPE to return the ground state energy E_0 as probability $p_0 = |c_0|^2$, we need to repeat the experiment $1/p_0$ times on average. We cannot just initialize a random state, as it will have tiny overlap with the true ground state due to the exponentially growing size of the Hilbert space. This problem is also referred to as the *orthogonality catastrophe* [70]. Therefore, we should have a sufficiently accurate heuristic for the preparation of the initial state on the quantum device. While the Hartree-Fock state is the simplest option, that is also trivial to prepare, it has been shown that for some strongly correlated systems, it has an exponentially small overlap with the ground state [71, 72]. Therefore, current research is going beyond the preparation of the HF state, for example considering a sum of Slater determinants coming from for example a truncated CI approach, or directly from a tensor network calculation [71, 73].

There has been promising advancements in the field that try to estimate and reduce the cost of QPE for real-world chemistry problems, like estimating ground states of nitrogenase [8], simulating ruthenium catalysts for carbon capture [11], or simulating lithium-ion batteries [74]. The costs of the QPE algorithm itself sits mainly in the time evolution. Efforts to bring this costs down include decomposing the Hamiltonian using efficient representations of the integrals [11, 75, 76], or methods in first quantization [77, 78].

The applications of these advancements in efficiency of time-evolution go beyond QPE: real-time time dynamics of electrons, nuclei, or both at the same time can also be done. This can be especially relevant in photochemical reactions, where the ground and excited state can get close in energy or even have points of degeneracy (conical intersections). Around these points, the nuclear kinetic energy is of the same order of magnitude as the electronic gap, resulting in complex correlations between electrons and nuclei and a breakdown of the Born-Oppenheimer approximation. The exploration of chemical systems beyond the Born-Oppenheimer approximation remains highly constrained, primarily due to the intricate electrons-nuclei interactions, as well as the challenge of representing space without the use of well-defined atomic orbitals. However, it is theoretically possible to simulate both electrons and nuclei simultaneously using first-quantized quantum simulation techniques [77, 78].

1.5 Outline of the thesis

This thesis seeks to advance quantum computational chemistry by reducing the quantum resources required for accurate simulations through a multi-faceted approach. The scarcity of quantum resources in near-term quantum devices necessitates to optimize every part of the pipeline in a quantum computational chemistry calculation. To make progress towards this goal, we develop classical pre-processing techniques to reduce the quantum resources required, and we look at targets beyond ground state energies that are challenging for classical methods. In chapter 2, we introduce a method to reduce the 1-norm of the electronic structure Hamiltonian by transforming the molecular orbitals. In chapter 3, we present a multi-fragment wavefunction embedding method to simulate chemical systems with fewer qubits. Subsequently, in chapter 4 we discuss the state-averaged orbital-optimized variational quantum eigensolver to simulate photochemical systems with multiple electronic states. In the final two chapters of this thesis, we investigate different targets than ground state energies that are more robust to noise. In chapter 5 we introduce a hybrid quantum algorithm to detect conical intersections in molecular models, a key challenge in photochemistry. Finally, in chapter 6, we develop machine learning density functionals from noisy quantum data. A statement regarding the publications that the individual chapters are based on can be found at the back of this thesis.

Chapter 2: Orbital transformations to reduce the 1-norm of the electronic structure Hamiltonian for quantum computing applications

In chapter 2, we introduce a method to reduce an induced norm λ of the electronic structure Hamiltonian by transforming the molecular orbitals. This norm surfaces as a crucial measure of quantum resources in quantum algorithms both in the NISQ and for FTQC.

In NISQ, optimizing the number of samples needed to measure an expectation value of an operator to a certain precision is crucial. Without advanced strategies like grouping of Hamiltonian terms, it has been shown that the optimal sampling strategy has a number of measurements that is proportional to λ^2 [60]. In FTQC, the norm λ is a key measure of the quantum resources – think runtime and space requirements – needed in advanced Hamiltonian simulation algorithms like QDRIFT [79], Linear Combination of Unitaries and Qubitization [75, 79–82].

As the λ-norm is an induced 1-norm of Hamiltonian coefficients, it is not invariant under unitary transformations of the orbitals. In this chapter we show that the commonly used Canonical Molecular Orbitals, coming from a Hartree-Fock calculation, result in Hamiltonians with exceptionally high λ-norm. In contrast, localized orbitals can be exploited to give much lower λ, as well as providing chemically intuitive bonding orbitals. Besides localization schemes, we introduce an orbital-optimization scheme that is specifically tailored to minimizing λ. We benchmark its effectiveness on a large set of systems, ranging from Hydrogen chains to industrially interesting systems like FeMoco [8] and Ruthenium metal complexes [11].

Chapter 3: FragPT2: Multi-Fragment Wavefunction Embedding with Perturbative Interactions

In chapter 3, we design a novel embedding framework to tackle the challenges of multiple interacting active fragments within large molecular systems.

Using an accurate but computationally expensive method is often not feasible on the whole molecular system at once. Embedding techniques allow for an efficient description of localized correlations, while accounting for their environment on a lower level of theory [83–85]. In this manner, one can for example offload the computationally intensive ground state calculation to two or more separate quantum computers, or execute them serially on one device, simulating one fragment at a time. This results in a significant reduction in space and time requirements. However, embedding methods often face computational challenges, requiring numerous self-consistent cycle iterations and typically falling short in capturing inter-fragment correlations.

The FragPT2 framework is aimed to address both these challenges. It is based on the idea of a one-step embedding: localization of molecular orbitals on pre-defined fragments that consist of a subset of atoms. In this manner we construct separate active spaces for each fragment that are self-consistently embedded in the mean field from other fragments. FragPT2 reintroduces inter-fragment correlations via multi-reference perturbation theory. This approach not only improves the efficiency, but also provides a robust tool for analyzing various inter-fragment interactions, such as dispersion, charge transfer, and spin exchange.

Our method is benchmarked against complex systems, including N_2 dimers, aromatic dimers, and butadiene, demonstrating its capability to successfully handle cases where fragments are defined by cutting through covalent bonds. This paves the way towards the treatment of larger systems with wavefunction based techniques and quantum algorithms.

Chapter 4: Analytical non-adiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver

Simulating photochemical systems requires an accurate description of multiple electronic states. For a qualitatively correct description of photochemical processes, it is essential to treat the ground and excited states on equal terms, i.e., democratically (see Sec. 1.3.2). Additionally, dynamic simulation of photochemical processes necessitates energy gradients with respect to nuclear geometries, as well as non-adiabatic couplings that govern population transfer from the excited state to the ground state. Traditional methods like TD-DFT tend to be biased toward the ground state, leading to inaccuracies in the calculation of energies, gradients, and non-adiabatic couplings. Therefore, this presents an ideal challenge for quantum simulation algorithms.

In Chapter 4, we build on recent work by our group, which introduced the state-averaged orbital-optimized variational quantum eigensolver (SAOOVQE) [30].

SAOOVQE is a hybrid quantum-classical algorithm that integrates the state-averaged complete active space self-consistent field (SA-CASSCF) method with the state-averaged variational quantum eigensolver (SA-VQE). It employs a state-averaged ansatz to represent multiple electronic states, while optimizing both the circuit parameters and orbitals in a democratic manner.

In this work, we present several technical and analytical extensions to SAOOVQE. First, we introduce an efficient state-resolution procedure for finding the adiabatic eigenstates of the system, rather than merely the subspace they span. Second, we derive analytical energy gradients with respect to nuclear coordinates, which are critical for geometry optimization and dynamics simulations. Finally, we derive analytical expressions for non-adiabatic couplings between electronic states, which are crucial for simulating photochemical processes. We benchmark the SAOOVQE method on a model system, using formaldimine (see Fig. 1.2) as a representative molecule for studying photoisomerization in larger biomolecules. Additionally, we perform a geometry optimization to locate the conical intersection between the ground and first-excited electronic states of the molecule.

Chapter 5: A hybrid quantum algorithm to detect conical intersections

In Chapter 5, we explore a key challenge in photochemistry: detecting conical intersections in molecular systems. To address this, we introduce a quantum simulation algorithm specifically designed for the Noisy Intermediate-Scale Quantum (NISQ) era.

Conical intersections are critical points in molecular geometries where electronic potential energy surfaces intersect in a topologically protected way. Near these intersections, the Born-Oppenheimer approximation breaks down, leading to significant consequences for chemical processes such as photoisomerization and non-radiative relaxation. A well-known example is the role of conical intersections in vision, where they enable the isomerization of retinal upon photon absorption, triggering a chemical cascade that results in light perception.

A defining feature of conical intersections is the non-zero Berry phase, a topological invariant that arises on a closed loop in atomic coordinate space. This phase takes on a value of π if the path encircles the intersection manifold, and 0 otherwise. The proposed algorithm tracks the approximate ground state along a specified path using a parametrized quantum circuit, which is iteratively updated with a fixed number of Newton-Raphson steps. A Hadamard test is then applied to extract a single bit of information, revealing whether the Berry phase is π or 0.

Because the outcome of the algorithm is discrete, it remains effective even with bounded cumulative errors; this allows us to prove an analytical bound on the total sampling cost and the procedure's success. We demonstrate the algorithm's application through numerical simulations on the same toy model of the formaldimine molecule ($H_2C{=}NH$).

Chapter 6: Machine Learning Density Functionals from Noisy Quantum Data

In Chapter 6, we introduce a machine learning method to construct density functionals using noisy quantum data. Density Functional Theory (DFT), a cornerstone of quantum chemistry, offers low computational cost for calculating ground state energies based on electron density, following the Hohenberg-Kohn theorems. The first theorem states that the ground state energy is a unique functional of the electron density, while the second theorem provides a way to optimize the orbitals to represent the ground state density using the Kohn-Sham equations. However, the accuracy of DFT calculations hinges on the choice of the exchange-correlation functional, which is often approximated using (semi-)empirical methods.

In this chapter, we propose a machine learning approach to learn the *exact functional*, i.e. the map from electron densities to ground state energies. As Ref. [86] showed, this exact functional can be efficiently learned by a convolutional neural network by training on exact diagonalization (ED) data for small-size Fermi-Hubbard models. However, due to exact methods scaling exponentially with system size, generating the training data for these models becomes infeasible for larger systems.

To overcome this, we train our model on noisy quantum data from near-term devices. While NISQ devices enable beyond-classical simulations of larger systems, algorithms like VQE suffer from sampling noise, circuit noise, and optimization errors. We show that our model can effectively mitigate unbiased sampling noise in the Fermi-Hubbard model, outperforming any individual noisy training point. When trained on data with expressibility and optimization errors, the model offers little improvement over training points, as it reflects the biases present in the data.

Lastly, we demonstrate that the trained models can be applied to new systems with varying potentials using a Kohn-Sham-like approach, leveraging the model's automatic differentiation for reasonably accurate solutions across most instances.

CHAPTER 2

Orbital transformations to reduce the 1-norm of the electronic structure Hamiltonian

2.1 Introduction

Quantum chemistry has been identified as the killer application of quantum computers, which promise to solve problems of high industrial impact that are not tractable for their classical counterparts [8, 9, 11, 87]. Unfortunately, quantum devices are noisy and only shallow circuits can be implemented with a relatively accurate fidelity. Hence, even though we know how to solve the electronic structure problem with, for instance, the quantum phase estimation (QPE) [48, 69, 88–90] algorithm, this remains out of reach in practice and one has to come up with new algorithms that can actually be run within this noisy intermediate-scale quantum (NISQ) era [53]. Reducing circuit depth has been achieved by interfacing quantum and classical devices, thus leading to hybrid quantum-classical algorithms such as the variational quantum eigensolver (VQE) [54]. Since then, variational algorithms have been successfully applied to the estimation of ground-state (see Refs. 91–94 and references therein) and excited-state energies [30, 95–104], as well as molecular properties [105–108].

However, it remains unclear if these algorithms can provide a clear quantum advantage in the long run, especially due to the difficulty in circuit optimization [109] and to the large overhead in the number of measurements required to achieve sufficient accuracy [60, 63], though significant progress has been made recently [60, 61, 110–120]. This led to the development of new strategies for reducing the resources required to implement quantum algorithms (such as gate complexity), for both the fault-tolerant and NISQ era. For simulating the time-evolution operator $U(t) = e^{iHt}$ where H is the electronic Hamiltonian, Campbell has shown that using randomized compiler is better suited than the Trotter-Suzuki decomposition, the so called quantum stochastic drift protocol (qDRIFT) [79]. One can also directly simulate the Hamiltonian by using linear combination of unitaries, generalized by the block encoding or “qubitization” formalism [75, 80–82]. Most of these algorithms have a gate complexity that scales with respect to the parameter $\lambda_Q = \sum_i |h_i|$, where $H = \sum_i h_i P_i$ is the qubit Hamiltonian and P_i are Pauli strings [55]. Lowering the value of λ_Q – usually referred to the 1-norm of the Hamiltonian – has a significant impact on quantum algorithms, and techniques such as double-factorization [11], tensor hypercontraction [76] and n-representability constraints [60] have proven successful in this respect.

In this work, we investigate how the value of the 1-norm and its scaling with respect to the number of orbitals can be reduced by single-particle basis rotations. As a starting point we consider the use of localized orbitals that are commonly available in most quantum chemistry codes. Generating such orbitals presents a simple and effortless way of reducing gate complexity by classical pre-optimization of the Hamiltonian representation.

We show that the use of localized orbitals has a significant impact on the 1-norm of all systems studied, from simple one-dimensional hydrogen chains (also studied in Refs. 60, 76) to much more complex organic and inorganic molecules. Additionally, we connect the expressions of the 1-norm before and after the fermion-to-qubit mapping. This expression is used to define a new cost function for single-particle basis rotations, and is shown to reduce the value of λ_Q even more than standard localization schemes.

The chapter is organized as follows. After introducing the electronic structure problem in Sec. 2.2.1, the impact of the 1-norm on quantum algorithms is detailed in Sec. 2.2.2. Then, we review the localization schemes applied in this work in Sec. 2.2.3 followed by our so-called 1-norm orbital-optimization method in Sec. 2.2.4. Finally, preceded by the computational details in Sec. 2.3, the scaling of the 1-norm when increasing the number of atoms is investigated in Sec. 2.4.1 for hydrogen and alkane chains, followed by a benchmark on several systems in Sec. 2.4.2 and a study of the effect of increasing the active space size in Sec. 2.4.3. Conclusions and perspectives are given in Sec. 2.5.

2.2 Theory

2.2.1 Electronic structure Hamiltonian in second quantization

In quantum chemistry, the second quantization formalism is usually employed to describe the electronic properties of molecules. Under the Born-Oppenheimer approximation, N_e electrons can rearrange around N_a nuclei by occupying a restricted set of N spatial molecular orbitals (MO). The orthonormal MO basis $\{\phi_p(\mathbf{r})\}$ is built from linear combination of atomic orbitals (LCAO) and is usually obtained from an inexpensive mean-field calculation, *e.g.* with the Hartree–Fock (HF) method [21]. In this basis, the molecular Hamiltonian can be expressed in a spin-free form, see (1.14). The one-electron integrals are given in Eq. (1.15). These encode, for each individual electron, the associated kinetic energy and Coulombic interaction with the N_a nuclei of the molecule. The two-electron integrals are given in Eq.(1.16), encoding the Coulombic repulsion between each pair of electrons. Solving the electronic structure problem consists in solving the time-independent Schrödinger equation

$$H\left|\Psi_\ell\right\rangle = E_\ell\left|\Psi_\ell\right\rangle, \tag{2.1}$$

with $|\Psi_\ell\rangle$ an electronic eigenstate with corresponding energies E_ℓ. This is only possible in the case of small molecules and small basis sets, due to the exponential scaling of the computational cost with respect to the size of the system. A commonly employed approach is the so-called active space approximation which divides the MO space into three parts: the core (frozen occupied orbitals), the active, and the virtual spaces (deleted unoccupied orbitals), such that only the electrons inside the active space are treated explicitly. This approximation will be used throughout this work and is equivalent to finding the eigenstates of an effective Hamiltonian H^{act} also called the active space Hamiltonian.

The electronic Hamiltonian can be mapped onto an appropriate representation for quantum computers by doing a fermion-to-qubit transformation (such as Jordan–Wigner [55]), resulting in a linear combination of Pauli strings $P_j \in \{I, X, Y, Z\}^{\otimes N}$,

$$H_Q = \sum_j^{\mathcal{S}} h_j P_j. \tag{2.2}$$

Here, $\mathcal{S}$ denotes the sparsity of the Hamiltonian and generally scales as $\mathcal{O}(N^4)$, but sometimes $\mathcal{O}(N^2)$ for a sufficiently large system and a localized basis [121].

2.2.2 The 1-Norm in quantum computing

The term ‘1-norm’ in quantum computing refers to a norm induced on a Hilbert-space operator by its decomposition as a sum of simpler terms. To calculate a 1-norm, we write an operator H (e.g. a Hamiltonian) in the form

$$H = \sum_j b_j B_j, \tag{2.3}$$

where the B_j are operators on $\mathbb{C}^{2^{2N}}$, and the b_j are complex numbers. Typically the operators B_j are chosen to be either unitary ($B_j^\dagger B_j = 1$) or unital ($B_j^\dagger B_j < 1$).

Method	Toffoli/T complexity
Database method [122]	$\mathcal{O}(N^4\lambda/\epsilon)$
qDRIFT [79]	$\mathcal{O}(\lambda^2/\epsilon^2)$
Qubitization (sparse method) [75]	$\mathcal{O}((N+\sqrt{\mathcal{S}})\lambda/\epsilon)$
Qubitization (single factorization) [75]	$\mathcal{O}(N^{3/2}\lambda_{\mathrm{SF}}/\epsilon)$
Qubitization (double factorization) [11]	$\mathcal{O}(N\lambda_{\mathrm{DF}}\sqrt{\Xi}/\epsilon)$
Tensor hypercontraction [76]	$\tilde{\mathcal{O}}(N\lambda_{\zeta}/\epsilon)$
VQE	# of measurements
State tomography [54, 63, 91]	$M \approx \lambda^2/\epsilon^2$
Basis rotation grouping [117]	$M \approx \lambda_{\mathrm{SF}}^2/\epsilon^2$

Table 2.1: Lowest asymptotic scaling of different quantum algorithms involving the λ parameters. λ denotes the 1-norm of the qubit Hamiltonian in Eq. (2.12), λ_{SF} (denoted λ_W in Ref. 75) the one of the singly-factorized fermionic Hamiltonian, λ_{DF} the one of the doubly-factorised Hamiltonian and λ_{ζ} the one of the non-orthogonal tensor hypercontracted Hamiltonian representation. $\mathcal{S}$ is the sparsity of the electronic Hamiltonian and Ξ the average rank of the second tensor factorization [11, 123]. See text for further details.

We define the 1-norm of H (induced by this decomposition) to be the 1-norm on the vector $\vec{b}$ (with components b_j):

$$\lambda_B = \lambda_B(H) = \sum_j |b_j|. \tag{2.4}$$

(It is common to suppress the dependence on the Hamiltonian H when λ_B is unambiguous.) For example, under the above decomposition into Pauli operators [Eq. (2.2)], we have

$$\lambda_Q \equiv ||\vec{h}||_1 = \sum_j |h_j| \tag{2.5}$$

In general, λ_B is a norm on the space spanned by the set $\{B_j\}$ as long as this set is linearly independent, as the mapping $H \leftrightarrow \vec{b}$ is then bijective and linear. The Pauli operators P_j used above make a natural choice for B_j as they span the set of $2^{2N} \times 2^{2N}$ matrices, and so induce a norm for any operators on the Hilbert space. Pauli operators are also unitary, and may be easily implemented in a quantum circuit [124], making them a common choice for Hamiltonian decompositions. However, unlike other operator size measures, 1-norms may be heavily optimized by a change of operator basis (whereas *e.g.* the 2-norm or infinity-norm are both invariant under unitary rotation). As operator 1-norms play a role in determining the cost of many quantum computing algorithms, this makes them a key target of study in quantum algorithm research.

A versatile method for implementing an arbitrary (non-Hermitian) operator on a quantum device is the linear combination of unitaries (LCU) method [80], which involves a decomposition over a set $\{B_j\}$ of strictly unitary operators. As H is typically not unitary, the original LCU technique requires postselection, but this forms the basis for unitary methods such as qubitization [82]. These in turn underpin many proposals for quantum computing on a quantum computer [11, 76, 123], and the dependence on λ_B is pulled directly through. We give a brief description here of the key identity in LCU methods to show where the dependence on λ_B appears. LCU methods all require a control register with states $|j\rangle$ that encode the index to the operator B_j. (One need only encode those operators B_j that are relevant to the target H.) Then, these methods encode the coefficients of H on the register by preparing the state $\frac{1}{\sqrt{\lambda_B}}\sum_j \sqrt{b_j}|j\rangle$. Then, consider implementing the joint unitary $\sum_j |j\rangle\langle j| \otimes B_j$ (that is, applying the unitary operator B_j to a system conditional on the control register being in the state j). Post-selecting the control register returning to the initial state performs the LCU operation to the register

$$\begin{aligned}\frac{1}{\sqrt{\lambda_B}}\sum_j \sqrt{b_j}\langle j| \left[\sum_j |j\rangle\langle j|B_j\right] \frac{1}{\sqrt{\lambda_B}}\sum_j \sqrt{b_j}|j\rangle \\ = \frac{1}{\lambda_B}\sum_j b_j B_j = \frac{1}{\lambda_B}H. \end{aligned} \tag{2.6}$$

We see here that λ_B emerges naturally as the normalization constant of the LCU unitary. This dependence is passed on to any LCU-based method, suggesting that minimizing λ_B is key to optimizing any such techniques.

The 1-norm can also play a role in methods where Hamiltonian simulation is implemented stochastically. For instance, in the qDRIFT [79] method, hermitian terms B_j in the Hamiltonian are chosen at random, weighted by $|b_j|/\lambda_B$, and exponentiated on a system as $e^{i\tau B_j}$ (λ_B appears here immediately as the normalization of the probability distribution). To first order, this implements the channel on a state $\hat{\rho}$ [79]

$$\mathcal{E}(\hat{\rho}) = \hat{\rho} + i\sum_j \frac{b_j\tau}{\lambda_B}(B_j\hat{\rho} - \hat{\rho}B_j) + O(\tau^2), \tag{2.7}$$

which can be observed to be the unitary evolution of $\hat{\rho}$ under e^{iHdt} for $dt = \tau/\lambda_B$. To approximate time evolution for time t to error ϵ then requires repeating this process $O(2\lambda_B^2 t^2/\epsilon)$ times. This has been extended upon recently [125], yielding a more complicated λ_B dependence.

A final situation where the 1-norm of an operator plays a role is in tomography. One may in principle measure the expectation value of an arbitrary Hermitian operator H on a state ρ by repeatedly preparing $\hat{\rho}$, rotating into the eigenbasis of

H and reading out the device. The variance in averaging M shots of preparation and measurement is directly given by the variance of H on $\hat{\rho}$:

$$\mathrm{Var}\left[\langle H\rangle\right] = \frac{1}{M}\left[\langle H^2\rangle - \langle H\rangle^2\right]. \tag{2.8}$$

However, in practice (especially in the NISQ era) arbitrary rotations into an unknown basis may be costly. As expectation values are linear, given a decomposition of H over $\{B_j\}$, one may instead perform M_j different preparations of $\hat{\rho}$ and measurements in the basis of each B_j and sum the result,

$$\langle H\rangle = \sum_j b_j \langle B_j\rangle, \tag{2.9}$$

with variance

$$\mathrm{Var}\left[\langle H\rangle\right] = \sum_j \frac{|b_j|^2}{M_j}\left[\langle B_j^2\rangle - \langle B_j\rangle^2\right] =: \sum_j \frac{|b_j|^2}{M_j} V_j. \tag{2.10}$$

Typically the V_j are not known in advance, but if B_j are chosen to be unital (by a suitable rescaling of b_j) then one can bound $V_j < 1$. One can confirm (e.g. via the use of Lagrange multipliers [60]) that the least number of measurements required to bound $\mathrm{Var}[\langle H\rangle]$ below some error ϵ^2 can be found by choosing $M_j = M|b_j|/\lambda_B$, and that this yields a total number of measurements

$$M = \sum_j M_j = \frac{\lambda_B^2}{\epsilon^2}, \tag{2.11}$$

first mentioned by Ref. [63]. State tomography is essential in hybrid quantum-classical algorithms such as a VQE. In the VQE, the Hamiltonian is usually expressed as a linear combination of Pauli operators, meaning $\lambda_B = \lambda_Q$. As the number of measurements is one of the biggest overheads in the practical implementation of VQE, lowering the 1-norm would be highly beneficial. One way to do so is to add constraints to the Hamiltonian based on the fermionic n-representability conditions, which has led to a reduction of up to one order of magnitude for hydrogen chains and the H_4 ring molecule [60]. Another hybrid algorithm that would benefit for a decrease in the 1-norm is the constrained-VQE algorithm, as this would decrease the higher-bound used in the penalty term, thus improving convergence [126].

As the above considerations imply, many of the most competitive algorithms have a strong dependence on 1-norms for various choices of the basis decomposition $\{B_j\}$, as shown in Tab. 2.1. The 1-norm induced by a Pauli decomposition, λ_Q (Eq. 2.5), scales with the number of orbitals somewhere in between $\mathcal{O}(N)$ and $\mathcal{O}(N^3)$. This depends on the system, and whether N increases because the number of atoms increases for a fixed number of basis functions per atom, or because the number of basis functions per atom increases while fixing the number of atoms. As far as we are aware of, this scaling has only been numerically benchmarked for the H_4 ring Hamiltonian [60, 76] and one-dimensional hydrogen chains [76], which remain quite far from realistic chemistry problems encountered in enzymatic [8] and catalytic reactions [11]. In this work, we perform a similar benchmark on a larger set of organic and inorganic molecules, from hydrogen and carbon chains to the FeMoco and ruthenium complexes.

The 1-norm used in many of the aforementioned algorithms in Tab. 2.1 takes the following form [75, 76]

$$\lambda = \lambda_T + \lambda_V \tag{2.12}$$

where

$$\lambda_T = \sum_{pq}^{N} \left| h_{pq} + \frac{1}{2} \sum_{r}^{N} (2g_{pqrr} - g_{prrq}) \right| \tag{2.13}$$

and

$$\lambda_V = \frac{1}{2} \sum_{pqrs}^{N} |g_{pqrs}| \tag{2.14}$$

with $\lambda_V \gg \lambda_T$. Note that the difference between Eqs. (2.13) and (2.14) and Eqs. (A10) of Ref. 76 comes from a different convention when writing the electronic Hamiltonian in Eq. (1.14), and the fact that N refers to spatial orbitals in this work instead of spin-orbitals. This norm of the Hamiltonian expressed as a linear combination of unitaries is used in the database method of Babbush *et al.* [122], the qubitized simulation by the sparse method of Berry *et al.* [75] (further improved by Lee and coworkers [76]) and the qDRIFT protocol of Campbell [79].

Further attempts have been made to reduce the number of terms in the Hamiltonian. For instance, one may perform a low rank decomposition of the Coulomb operator V such that the 1-norm of the – now singly-factorised – Hamiltonian reads [75, 76]

$$\lambda_{\mathrm{SF}} = \frac{1}{4} \sum_{\ell=1}^{L} \left(\sum_{p,q}^{N} \left| W_{pq}^{(\ell)} \right| \right)^2 , \tag{2.15}$$

where $W_{pq}^{(\ell)}$ are obtained from a Cholesky decomposition of the g_{pqrs} tensor. Note that λ_{SF} is a higher-bound of λ_V, and so this decomposition tends to increase the 1-norm. To improve over this low-rank factorization, one can write the Hamiltonian in a doubly-factorized form [11] by rotating the single-particle basis to the eigenbasis of the Cholesky vectors, such that the corresponding 1-norm λ_{DF} is much lower than after a single factorization. One can also use the tensor hypercontraction representation of the Hamiltonian [76]. However, applying the qubitization algorithm on this Hamiltonian directly is not efficient, as its associated 1-norm λ_{THC} is even larger than λ_V. To bypass this issue, Lee *et al.* provided a diagonal form of the Coulomb operator by projection into an expanded non-orthogonal single-particle basis, thus leading to a non-orthogonal THC representation of the Hamiltonian with an associated 1-norm λ_ζ that scales better than any prior algorithm [76].

As readily seen in this section, the λ norm depends on how you represent your Hamiltonian. However, there is a unique way to write the Hamiltonian as a sum of unique Pauli strings, or equivalently as a sum of unique products of Majorana operators (see Appendix 2.B). Doing so allows to express the qubit 1-norm λ_Q in Eq. (2.5) as a function of the electronic integrals,

$$\lambda_Q \quad = \quad \lambda_C + \lambda_T + \lambda'_V, \tag{2.16}$$

where

$$\lambda_C \quad = \quad \left| \sum_p^N h_{pp} + \frac{1}{2} \sum_{pr}^N g_{pprr} - \frac{1}{4} \sum_{pr}^N g_{prrp} \right|, \tag{2.17}$$

$$\lambda_T \quad = \quad \sum_{pq}^N \left| h_{pq} + \sum_r^N g_{pqrr} - \frac{1}{2} \sum_r^N g_{prrq} \right|, \tag{2.18}$$

$$\lambda'_V \quad = \quad \frac{1}{2} \sum_{p>r,s>q}^N |g_{pqrs} - g_{psrq}| + \frac{1}{4} \sum_{pqrs}^N |g_{pqrs}|. \tag{2.19}$$

The first term λ_C corresponds to the absolute value of the coefficient of the identity term that emerges when rearranging the Hamiltonian in terms of unique majorana operators, see Eq. (2.44). It is invariant under orbital rotations and can be added to the energy as a classical constant together with the nuclear repulsion energy and, if one employs the active space approximation (see Section 1.3.1), the mean-field energy of the frozen core. Thus, this term (apart from being small compared to λ_T and λ'_V) will not be important for quantum algorithms, and we will leave it out of our results, redefining the 1-norm as: $\lambda_Q = \lambda_T + \lambda'_V$. λ_T represents the absolute values of the coefficients of the quadratic term in Majorana operators, and λ'_V of the quartic term. Notice that λ'_V is slightly different from λ_V of Ref. 76 also given in Eq. (2.14) (which added a slight correction to the one of Ref. 75). In this work, we add another correction that comes from the swapping of two Majorana operators. We confirm that this is the correct form by directly comparing λ_Q with the norm of the qubit Hamiltonian after doing a qubit-to-fermion transformation in Eq. (2.5). Note that $\lambda'_V \leq \lambda_V$ (see appendix 2.B for more information). By expressing the 1-norm of the qubit Hamiltonian in Eq. (2.2) with respect to the electronic integrals, we can compute its value before doing any fermion-to-qubit transformation, which can be costly for large systems.

2.2.3 Localized orbitals

As discussed previously, different approaches have already been considered to minimize the 1-norm of a given Hamiltonian H. In this work, we choose to tackle the problem from a chemistry point of view, by focusing on the use of orbital transformations (*i.e.* single-particle states rotations). More precisely, we investigate the use of orbital localization techniques as a classical pre-optimization method to express the electronic structure Hamiltonian H in a new basis (see Appendix 2.A for details about electronic integrals transformations) that presents natural advantages for quantum computing. Note that exploiting spatial locality to reduce the 1-norm has already been mentioned in Ref. 76, or to reduce the number of significant integrals in Ref. 121.

In computational chemistry, localization schemes represent state-of-the-art orbital-rotation techniques employed in various situations. For example, localized orbitals (LOs) are regularly used to alleviate the computational cost of numerical simulations in post-HF methods such as second order Møller Plesset [127–130], coupled cluster [131–134], and multireference methods [135–138]. They can also be used to partition a system in spatially localized subsystems that are treated at different levels of theory [85, 139–150]. In the context of quantum algorithms for the NISQ era, one may for instance consider performing a calculation with a classical mean-field method to produce localized orbitals and using the orbitals localized in the spatial region of interest as a basis for a calculation with a quantum algorithm. In the current work, we demonstrate that LOs can also be of a significant help beyond isolating the spatial region that is of chemical interest, by reducing the qubit 1-norm λ_Q of the electronic Hamiltonian H after a fermion-to-qubit transformation.

In the following, we introduce the orbital localization schemes considered in this work where the notations $\tilde{\psi}_p(\mathbf{r})$ and $\chi_\mu(\mathbf{r})$ are used to denote orthogonal LOs and non-orthogonal atomic orbitals (AO), respectively.

Lowdin orthogonal atomic orbital

The first approach we investigate to generate LOs is the orthogonalization of atomic orbitals method (OAO). In practice, several techniques exist to produce orthogonal AOs [151, 152]. Here, we focus on Löwdin's method [153, 154], known to generate orthogonal LOs with a shape that is the closest to the original AOs (in the least square sense). In practice, orthogonal Löwdin orbitals $\tilde{\psi}_p(\mathbf{r})$ are built via a linear combination of the N original AOs,

$$\tilde{\psi}_p(\mathbf{r}) = \sum_{\mu=1}^{N} \chi_\mu(\mathbf{r}) \tilde{C}_{\mu p}, \tag{2.20}$$

where the orbital coefficient matrix $\tilde{\mathbf{C}}$ takes a very simple form like

$$\tilde{\mathbf{C}} = \mathbf{S}^{-1/2}, \tag{2.21}$$

and $S_{\mu\nu} = \int \mathrm{d}\mathbf{r} \chi_\mu^\star(\mathbf{r})\chi_\nu(\mathbf{r})$ is the overlap matrix encoding the overlap between different non-orthogonal AOs.

From a practical point of view, this method represents one of the simplest and numerically cheapest localization methods, where the computational cost is dominated by the exponentiation of the overlap matrix and typically scales as $\mathcal{O}(N^3)$. However, this approach is defined with respect to the full AOs' space and cannot be straightforwardly applied to the practical case of active space calculations where only a restricted set of molecular orbitals – formed by linear combination of AOs – is used.

Molecular orbital localization schemes

Fortunately, other localization schemes can be used in any circumstances (*i.e.* with or without active space approximation). Among the possible approaches, we focus on three of them: the Pipek-Mezey [155] (PM), Foster-Boys [156] (FB) and Edmiston-Ruedenberg [157] (ER) methods. From a practical point of view, these methods fundamentally differ from OAO as they generate LOs out from molecular orbitals and not AOs. In practice, all these methods start from a set of orthogonal canonical MOs (CMOs) $\{\phi_s\}_{s=1}^N$ obtained with an initial mean-field calculation (*e.g.* Hartree–Fock or density functional theory). The CMOs form linear combinations of AOs as

$$\phi_s(\mathbf{r}) = \sum_\mu^N \chi_\mu(\mathbf{r}) C_{\mu s}, \tag{2.22}$$

with $\mathbf{C}$ the CMO-coefficient matrix. A set of LOs $\{\tilde{\psi}_p\}_{p=1}^N$ is then generated by applying a unitary transformation matrix $\mathbf{U}$ (with $\mathbf{U}^\dagger\mathbf{U} = \mathbf{U}\mathbf{U}^\dagger = \mathbf{1}$) to express each LO as a linear combination of the original CMOs,

$$\tilde{\psi}_p(\mathbf{r}) = \sum_s^N \phi_s(\mathbf{r}) U_{sp}, \tag{2.23}$$

or, in a more compact matrix form,

$$\tilde{\mathbf{C}} = \mathbf{C}\mathbf{U}, \tag{2.24}$$

where $\tilde{\mathbf{C}}$ is the LO-coefficient matrix. In practice, the shape of the LOs is numerically determined by modifying the unitary $\mathbf{U}$ to optimize a cost function $\mathcal{L}$ based on a relevant localization criterion (depending on the scheme considered). This localization can be prone to many local minima and the orbital-optimization process is usually realized with different numerical techniques (*e.g.* Jacobi rotations [158, 159], gradient descent [160], Newton and trust-region methods [161–163], etc.). In chemistry, the question of how to realize an efficient orbital-optimization process in practice still constitutes an active topic of research which goes way beyond the scope of our manuscript (we refer the interested reader to Ref. 164 and references within). Let us now focus on the different criteria used in the FB, PM and ER methods together with their respective numerical costs.

In the Foster-Boys scheme [156], the localization criterion is the square of the distance separating two electrons $\mathbf{r}_{12}^2 = |\mathbf{r}_2 - \mathbf{r}_1|^2$, and the set of LOs is obtained by minimizing the following cost function

$$\mathcal{L}_{\text{FB}} = \sum_p^N \iint |\tilde{\psi}_p(\mathbf{r}_1)|^2 \mathbf{r}_{12}^2 |\tilde{\psi}_p(\mathbf{r}_2)|^2 \text{d}\mathbf{r}_1 \text{d}\mathbf{r}_2, \tag{2.25}$$

which represents the average value of the distance criterion over the set of orbitals to be optimized. In practice, a single estimation of the cost function $\mathcal{L}_{\text{FB}}$ requires a decomposition in the AO basis that generates five nested sums, thus leading to a scaling of $\mathcal{O}(N^5)$. However, specific manipulations (see Refs. 156 and 164) can reduce the total computational cost of the Foster-Boys method to $\mathcal{O}(N^3)$ multiplied by the number of times $\mathcal{L}_{\text{FB}}$ is called (which intrinsically depends on the optimization algorithm considered in practice). Note that extensions of the Foster-Boys scheme exist and are based on the orbital variance [165] or the fourth moment method [166], resulting in more localized orbitals especially for basis sets augmented by diffuse functions.

In the case of the Edmiston-Ruedenberg method, the inverse distance $1/\mathbf{r}_{12}$ (proportional to the two-body electronic repulsion operator, see Eq. (1.16)) is used as a criterion, and the LOs are obtained by maximizing

$$\mathcal{L}_{\text{ER}} = \sum_p^N \iint |\tilde{\psi}_p(\mathbf{r}_1)|^2 \frac{1}{\mathbf{r}_{12}} |\tilde{\psi}_p(\mathbf{r}_2)|^2 \text{d}\mathbf{r}_1 \text{d}\mathbf{r}_2. \tag{2.26}$$

The computational cost of a single estimation of $\mathcal{L}_{\text{ER}}$ scales as $\mathcal{O}(N^5)$ (multiplied by the number of times $\mathcal{L}_{\text{ER}}$ is called), though it can be reduced up to $\mathcal{O}(N^3)$ by the use of density fitting or the Cholesky decomposition.

Finally, the Pipek-Mezey scheme adopts a very different point of view. PM orbitals are generated by maximising the Mulliken charge of each orbital [155]

$$\mathcal{L}_{\text{PM}} = \sum_{A=1}^{N_A} \sum_p^N (Q_A^p)^2, \tag{2.27}$$

where

$$Q_A^p = \sum_{\mu \in A} \sum_{\nu}^{N} \tilde{C}_{\mu p} S_{\mu\nu} \tilde{C}_{\nu p} \tag{2.28}$$

is the contribution of orbital p to the Mulliken charge of atom A. In this case, the numerical cost of one call of $\mathcal{L}_{\text{PM}}$ essentially depends on the triple sums over the molecular and atomic orbitals in Eqs. (2.27) and (2.28). As a results, the computational cost of the global PM method scales as $\mathcal{O}(N^3)$ (multiplied by the number of times $\mathcal{L}_{\text{PM}}$ is called). Note that the traditional PM method is ill-defined since Mulliken charges are basis set sensitive and do not have a basis set limit. Various partial charges estimates can be used instead of Mulliken charges [167], such as intrinsic orbitals that are basis set insensitive and lead to a cheaper and better behaved localization procedure than PM localization [168, 169].

In summary, the construction of OAOs is the cheapest approach but can't be used when considering active space. Somewhat more expensive are PM and FB which share an equivalent scaling, and finally ER which is the most computationally demanding method.

2.2.4 1-Norm orbital-optimization

One of the main contribution of this chapter is the use of an orbital-optimization (OO) process specifically dedicated to the minimization of the qubit 1-norm λ_Q. To proceed, we introduce a unitary operator

$$\mathbf{U}^{\text{OO}} = e^{-\mathbf{K}} \tag{2.29}$$

with an anti-hermitian generator

$$\mathbf{K}^T = -\mathbf{K} = \begin{pmatrix} 0 & K_{12} & K_{13} & \dots & K_{1N} \\ -K_{12} & 0 & K_{13} & \dots & K_{2N} \\ -K_{13} & -K_{13} & 0 & \dots & K_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -K_{1N} & -K_{2N} & -K_{3N} & \dots & 0 \end{pmatrix}. \tag{2.30}$$

This operator is then used to transform a reference MO basis into a new basis denoted by $\{\phi'_q\}$ with $\mathbf{C}' = \mathbf{C}\mathbf{U}^{\text{OO}}$, for which the optimal orbitals are obtained by minimizing the following cost function,

$$\mathcal{L}_{OO} = \lambda_Q(\{\phi'_q\}), \tag{2.31}$$

by varying the $N(N-1)/2$ off-diagonal independent elements of the matrix $\mathbf{K}$. This optimization can be carried out for a reference MO-basis that consists of canonical Hartree-Fock orbitals, but could equally well be carried out on a subspace of the full MO-basis, e.g. only the set of fractionally occupied orbitals in an active space type calculation, or only a subset of localized orbitals resulting from a preliminary localization using one of the methods discussed above.

In practice, to realize the OO of the 1-norm, we choose to use 'brute-force' optimization algorithms which autonomously estimate the local derivatives of the cost function (*i.e.* gradients and hessians). This choice is motivated by the expression of λ_Q in Eq. (2.16) that contains many absolute values and for which analytical derivatives are clearly non-trivial to estimate. Numerically, the main bottleneck of the OO method is linked to the repetitive transformation of the two-electron integrals realized at each step of the process. As a result, the core algorithm scales as $\mathcal{O}(N^5)$ (similarly to the ER localization scheme). However, with the computational effort deployed to estimate the numerical gradient and hessian of the complicated cost function (given in Eq. (2.31)), this scaling should increase more especially when treating large systems.

2.3 Computational details

The geometries of the small systems considered in this work were optimized using the ADF program of the Amsterdam Modeling Suite (AMS) [170] with a quick universal force-field (UFF) optimization, sufficient for our purposes. These geometries are provided in the Supplementary Material of Ref. [171]. The first geometry in Ref. 8 with a charge of +3 is used for FeMoco and the geometries in Ref. 11 (all with a charge of +1) were used for the Ruthenium metal complexes. The electronic integrals of the Hamiltonian were computed using the restricted Hartree–Fock method from the PySCF package [172]. For large molecules, the active space approximation was invoked according to Section 1.3.1. All the localization schemes used to transform the Hamiltonian were already implemented in the PySCF package. For our 1-norm orbital-optimization scheme, the SLSQP optimizer was used from the SciPy package [173] and the python version of the L-BFGS-B optimizer of the optimParallel package [174]. The choice of these optimizers has been motivated by their capacity to automatically approximate gradients (and Hessians) for minimization. This represents an interesting tool when no evident analytical gradient/Hessian can be determined as for the 1-norm. A fast algorithm to estimate this λ_Q [Eq. (2.16)] was implemented in the OpenFermion package [175], allowing users to calculate 1-norms of large molecular Hamiltonians without employing expensive fermion-to-qubit mappings.

2.4 Results

In this section, we study the scaling of λ_Q with respect to the number of orbitals. First, we fix the basis set and increase the number of atoms, in the spirit of Refs. 60 and 76. Concerning the orbital localization schemes, we allow rotations between the active occupied and virtual spaces as this can lead to a better localized orbital basis.

Then, we benchmark the value of λ_Q for several different chemical systems and active space sizes, using the different localization schemes and our 1-norm orbital-optimized scheme. When considering an active space, we always choose our active space based on the CMOs by considering the N_{act} orbitals around the Fermi level (*i.e.* around the highest occupied and lowest unoccupied molecular orbitals). We localize only *inside* the active space (i.e. the localizing unitary in Eq. (2.23) only has indices corresponding to active orbitals) such that the subspace spanned by the active space remains invariant under the unitary rotations. Although this means that the expectation values of observables remain the same inside the active space when an exact solver is used, one can converge to different expectation values when approximate solvers are considered such as the truncated unitary coupled cluster ansatz (that becomes exact if not truncated [176]). However, rotating from the CMO to the LMO basis will not necessarily deteriorate the results of the (approximate) simulation, as LMOs have significant importance in local correlation treatments in post-HF methods like second order Møller Plesset [127–130], coupled cluster [131–134], embedding approaches [85, 139–150], and multireference methods [135–138] on classical computers. Hence, using localized orbitals could lead to similar advantages in quantum computing simulations, and could also inspire new ansatz proposals based on embedding strategies. This is left for future work.

Finally, to study the scaling of λ_Q for large molecules, we increase the size of the active space N_{act} while fixing the basis set and the number of atoms.

2.4.1 Hydrogen and alkane chains: scaling of the 1-norm by increasing the number of atoms

In this section, we investigate the performance of the localization schemes together with our brute-force OO method, studying the scaling of λ_Q with respect to the number of orbitals by increasing the number of atoms in small systems. Inspired by Refs. 121 and 60, we consider linear chains of hydrogens with a spacing of $r = 1.4\,\text{Å}$ and linear alkane chains (for which the geometry has been optimized). Results are shown in Fig. 2.1 and Tab. 2.2.

The orbital-optimizer method is not included for the linear hydrogen chains, because the gradient-estimating optimizer has trouble finding the approximate gradient in the first step of the optimization. The reason for this is that the localization schemes already come so close to the optimal solution that the optimizer estimates the gradient to be zero or positive in each direction. The STO-3G basis results in the localized orbitals resembling just 1s atomic orbitals on every nucleus which are very localized and therefore have a minimal 1-norm. As linear hydrogen chains are very artificial systems and give limited information on the scaling of λ_Q for actual interesting chemical systems, we thought the results of the conventional localization schemes to be sufficient here.

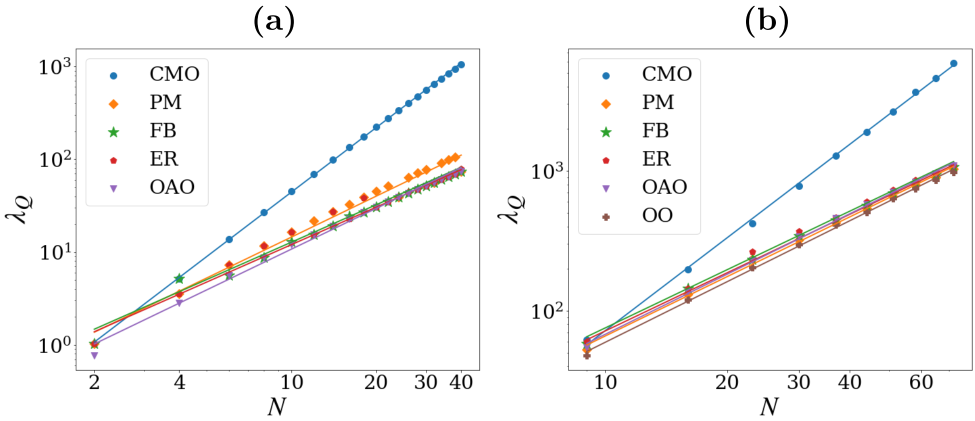


Figure 2.1: The 1-norm λ_Q (in Hartree) with respect to **(a)** the number of atoms in a linear hydrogen chain (same as the number of orbitals) and **(b)** the number of orbitals in a linear alkane chains (ethane, propane, butane, etc. up to decane), for different orbitals in the STO-3G basis. The constant term in the Hamiltonian is ignored. Assuming a polynomial scaling of $\lambda_Q = \mathcal{O}(N^\alpha)$, we fit $\log \lambda_Q = \alpha \log N + \beta$ and show the plot with log-log axes. The fitting and regression coefficients are given in Tab. 2.2.

As readily seen, all localization schemes perform comparably well in reducing the 1-norm compared to canonical molecular orbitals. Indeed, the scaling of the 1-norm decreases significantly from $\mathcal{O}(N^{2.31})$ to $\mathcal{O}(N^{1.34})$ for the hydrogen chains, and from $\mathcal{O}(N^{2.21})$ to $\mathcal{O}(N^{1.38})$ for the alkane chains. In terms of concrete values, for the largest N studied in Fig. 2.1, we can see a reduction of λ_Q by more than a factor 13 for the hydrogen and 5 for the alkane chains, when passing from the CMO basis to the LO ones. In the alkane chains, we do not see any significant distinction between the different schemes. While α for the orbital-optimization scheme is slightly higher than the best localization scheme, it is not so informative here as the λ_Q associated to the OO has the lowest absolute value for any orbital space. In the hydrogen chains, it seems that the Pipek-Mezey scheme is slightly less efficient. Surprisingly, the atomic orbital based Löwdin orthogonalization method do not provide any further improvement over other localized orbitals, which might be due to the simplicity of the systems for which localization schemes can generate extremely localized orbitals.

2.4.2 Benchmarking λ_Q for a variety of molecules and active spaces

In order to benchmark the impact of the localization schemes on the 1-norm, we calculated λ_Q for a variety of molecules and active spaces. Whenever we consider the of orbitals, we include the value of λ_Q obtained from the Hamiltonian

Hydrogen chain	α	β	R^2
CMO	2.31	-1.54	0.9990
PM	1.47	-0.70	0.8832
FB	1.34	-0.61	0.9967
ER	1.34	-0.54	0.9643
OAO	1.46	-0.99	0.9846

Alkane chain	α	β	R^2
CMO	2.21	-0.82	0.9978
PM	1.41	0.93	0.9965
FB	1.38	1.08	0.9976
ER	1.39	1.13	0.9926
OAO	1.43	0.92	0.9954
OO	1.44	0.76	0.9964

Table 2.2: Fitting and regression coefficients from Fig. 2.1.

expressed in the OAO basis. When an active space is considered, we use the frozen-core approximation and we localize the orbitals inside the active space. The resulting λ_Q values are tabulated in Tab 2.3, where the lowest ones obtained from standard localization schemes for each system are in bold. Apart from the different localization schemes, results of our brute force optimizer are also shown. Note that we start the optimization with already localized orbitals, such that it always gives a lower 1-norm than the best localization scheme. On most molecules, we had the best experience with the L-BFGS-B optimizer. Unfortunately, this optimizer was not able to estimate the gradient of λ_Q w.r.t. $\mathbf{U}^{\text{OO}}$ on the smallest molecules (H_2 and LiH), which may be due to infinitesimal changes in λ_Q when changing $\mathbf{U}^{\text{OO}}$. This why we employed the SLSQP optimizer here.

In order to identify the origin of the significant reduction of the 1-norm and its scaling, we focus on the transformation of the two-electron integrals defined in Eqs. (1.16) with the following permutational symmetries (for real-valued integrals)

$$\begin{aligned} & (pq|rs) = (pq|sr) = (qp|rs) = (qp|sr) \\ = \quad & (rs|pq) = (rs|qp) = (sr|pq) = (sr|qp), \end{aligned} \tag{2.32}$$

and we divide the indices of this tensor in seven separate classes: (1) pppp, (2) pqqq, (3) pqpq, (4) ppqq, (5) pqrq, (6) pprs, (7) pqrs, where the indices p,q,r and s are all different. The above notations are used in Fig. 2.2 and correspond to the sum of the absolute values of the electronic integrals associated to these indices [and the ones related by symmetries in Eq. (2.32)], e.g.

Molecule	CAS	CMO	PM	FB	ER	OAO	OO	% Red.
H_2	(2,10)	101	135	116	**93**	103	$90^{(a)}$	10.9%
LiH	(4,19)	185	177	190	178	**153**	$134^{(a)}$	27.6%
H_2O	(10,24)	717	678	710	662	**616**	$576^{(b)}$	19.7%
HLiO	(12,33)	993	787	831	**768**	788	$668^{(b)}$	32.7%
H_2CO	(16,38)	1792	1419	1417	1343	**1327**	$1101^{(b)}$	38.6%
$HNCH_2$	(16,43)	3096	1813	1670	**1588**	1645	$1240^{(b)}$	60.0%
C_3H_6	(24,45)	2316	1223	1145	**1137**	N/A	$995^{(b)}$	57.0%
C_4H_6	(30,45)	1760	**974**	1054	1048	N/A	$812^{(b)}$	53.9%
C_5H_8	(38,45)	1821	818	801	**799**	N/A	$698^{(b)}$	61.7%
HNC_3H_6	(32,50)	3796	1493	1244	**1232**	N/A	$1085^{(b)}$	71.4%
HNC_5H_{10}	(48,50)	2632	1098	**989**	1002	N/A	$842^{(b)}$	68.0%
HNC_7H_{14}	(50,50)	2610	751	**705**	707	N/A	$616^{(b)}$	76.4%

Table 2.3: Values of λ_Q for relatively small test molecules in the cc-pVDZ basis-set. The active space size is indicated as (n_e, N), where n_e is the number of electrons in N the number of spatial orbitals. The lowest 1-norm obtained from standard localization schemes for each system are in bold. Superscripts (a) and (b) refer to the use of the SLSQP and the L-BFGS-B optimizer used for in our 1-norm orbital-optimization scheme (denoted as "OO" in the table), respectively. The rightmost column shows the percentage of reduction obtained from the 1-norm-optimized orbitals compared to the CMOs.

$$
\begin{aligned}
\text{pqrq} \quad &\equiv \sum_{p\neq q\neq r} |g_{pqrq}| + |g_{pqqr}| + |g_{qprq}| + |g_{qpqr}| \\
&= \sum_{p\neq q\neq r} 4|g_{pqrq}| \\
\text{pqrs} \quad &\equiv \sum_{p\neq q\neq r\neq s} |g_{pqrs}|
\end{aligned}
\tag{2.33}
$$

for the cases (5) and (7), respectively. These contributions can give us some insight as to how the localized orbitals affect the norm of the diagonal and off-diagonal parts of the tensor, and are represented in Fig. 2.2 for the second stable intermediate Ruthenium complex in different orbital bases. As readily seen in this diagram, localization schemes do not lead to a uniform reduction of the norm of every term, but tends to maximize some of them (the pppp term), minimize others (the pqrs term) and leaves others relatively unchanged (the ppqq term). However, the pqrs contribution clearly dominates in the CMO basis, being more than one-order of magnitude larger than any of the other terms. Especially the orbital-optimizer can give an even better reduction of this term. This explains why the localization schemes can be used to reduce the 1-norm, as reducing the

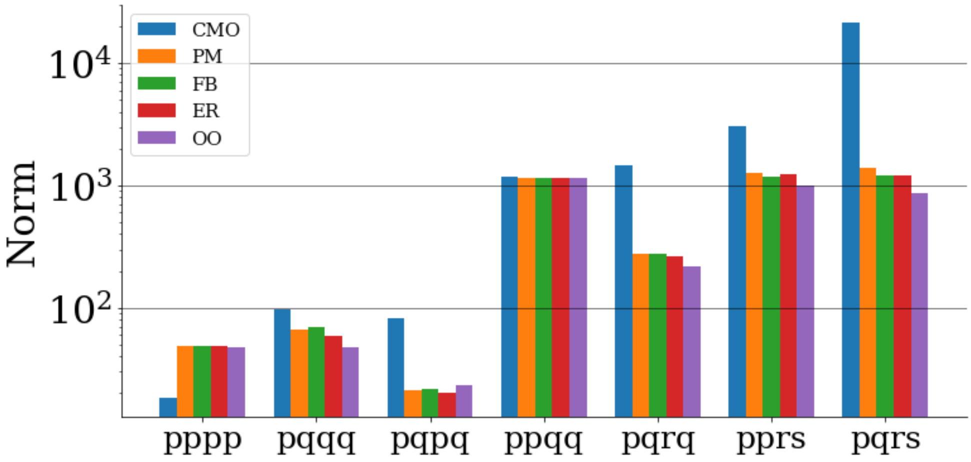


Figure 2.2: (Color) Sum of absolute values of different parts of the two-electron integral tensor on a logarithmic scale (see text for further details on the notations). It shows the second stable Ruthenium intermediate in a minimal ANO-RCC basis for an active space of 100 electrons in 100 spatial orbitals.

pqrs norm by one order of magnitude (as seen in Fig. 2.2) will play a much more important role than increasing the pppp one by the same order of magnitude. As the magnitude of these integrals depends on the overlap densities $\psi_p(\mathbf{r})\psi_q(\mathbf{r})$ and $\psi_r(\mathbf{r})\psi_s(\mathbf{r})$, it is clear that localization will reduce the number of numerically significant contributors. For an extended system, integrals in which p and q as well as r and s are localized on the same atoms are expected to dominate. This explains the observation that the pqrs norm becomes comparable to the ppqq and pprs terms when employing localized orbitals.

2.4.3 Effect of increasing the size of the active space on λ_Q

Let us now investigate how λ_Q scales with respect to the number of active orbitals for large and complex systems. We selected test-molecules based on two important recent papers in the field of quantum computing for quantum chemistry: FeMoco – an iron-molybdenum complex that constitutes the active site of a MoFe protein which plays an important role in nitrogen fixation [8] – and all the stable intermediates and transition states of the catalytic cycle presented in Ref. 11. This catalytic cycle describes the binding and transformations of carbon dioxide CO_2, a molecule with infrared absorption properties that makes it a potent and the most important greenhouse gas.

For the calculations to remain doable in a reasonable amount of time, the ANO-RCC minimal basis (with scalar relativistic corrections) is used. This should be sufficient for the current purpose of studying 1-norm reductions within an

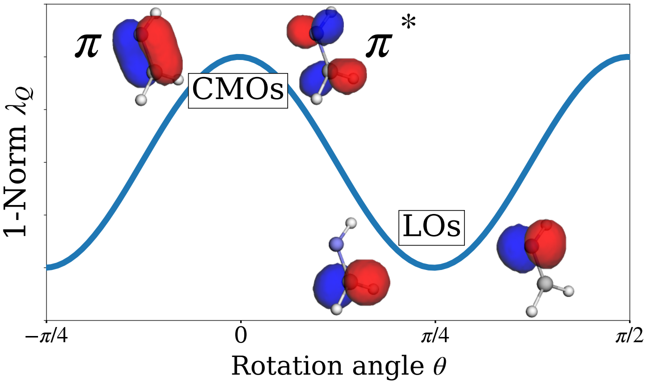


Figure 2.3: An illustration of how to decrease the 1-Norm on the formaldimine molecule: moving from CMOs to LOs with a Jacobi rotation between occupied π-bonding and virtual π^*-antibonding orbitals.

active space, but we note that larger basis sets will be required to reach chemical accuracy [177]. The scaling of λ_Q with respect to the number of active orbitals N_{act} for different orbital bases are reported in details in Fig. 2.4. Comparing the scaling reported in this section and the ones obtained in Sec. 2.4.1, we see that the scaling is larger when increasing the number of active orbitals than when increasing the number of atoms, as expected. However, the impact of localized orbitals on the 1-norm is similar in both cases, as it reduces the scaling by almost one order of magnitude. We also do not see any significant distinction between the 1-norm associated to the different conventional localization schemes. The brute-force orbital-optimizer however, gives a consistent 10-25% improvement over the conventional localization schemes. In the largest active spaces considered ($N_{\text{act}} \geq 85$) the orbital-opimization scheme can converge very slowly due to the need of 4-index transformations at every step, combined with the non-continuous and complicated landscape of λ_Q. Because of this, the user is usually forced to cut off the optimizer at some point, or choose a high threshold of convergence. In spite of this, we got similar good results in the largest active spaces we considered, indicating that the optimizer is not limited to a specific active space size.

The scaling results for all large molecules are summarized in Table 2.4. The scaling lies in the order of $\mathcal{O}(N^{2.6})$ to $\mathcal{O}(N^{2.9})$ when using CMOs, while it lies around $\mathcal{O}(N^2)$ or slightly higher when using LOs. One gets a bit of a misleading picture looking at just the scaling of the orbital-optimizer, as it has the lowest absolute value of λ_Q at every point. The cause is that the cost function of the orbital-optimizer has more local minima and is harder to converge compared to conventional localization schemes.

For the localization schemes it is usual that the bigger the active space is, the more λ_Q can be reduced compared to CMOs. This can be rationalized as follows.

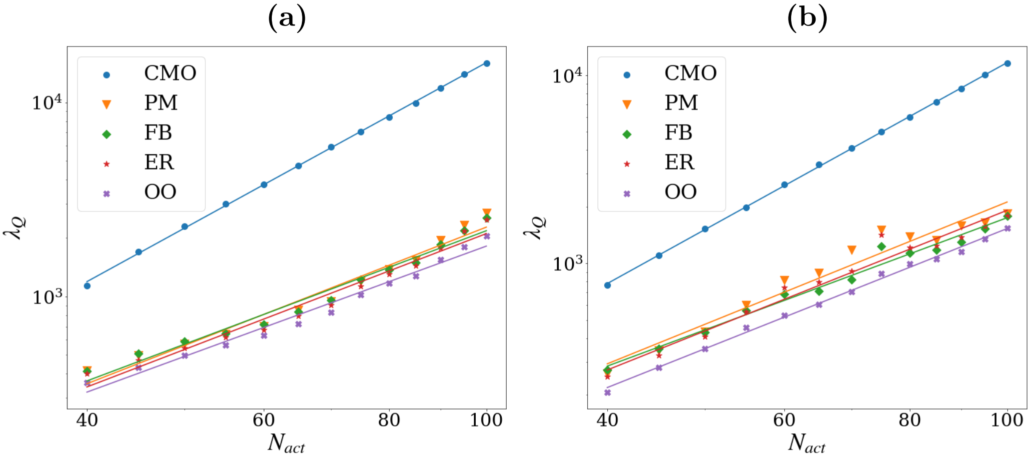


Figure 2.4: Scaling of λ_Q (in Hartree) with respect to the size of the active space for different orbitals for **(a)** FeMoco and **(b)** the Ruthenium transition state II-III in a minimal ANO-RCC basis. The constant term in the Hamiltonian is ignored. Assuming polynomial scaling of $\lambda_Q = \mathcal{O}(N_{\text{act}}^{\alpha})$ we fit $\log \lambda_Q = \alpha \log N_{\text{act}} + \beta$ and show the plot with log-log axes. The fitting parameters and regression coefficients are given in Table 2.4.

When the active space increases, it may involve orbitals localized on atoms which have negligible overlap with the other orbitals. Another possibility is that the additional orbitals may have a positive effect on the localization procedure. As an illustration, see the example shown in Fig. 2.3 where we consider an occupied π-bonding orbital mixing with its associated virtual π^*-antibonding orbital for the formaldimine molecule. By rotating those two orbitals via a Jacobi rotation, we move from a maximal value of λ_Q for $\theta = 0$ (corresponding to the original CMOs) to a minimal value for $\theta = \pi/4$ (corresponding to the new LOs). This way, we see that the larger is the active space the more degrees of freedom we have to perform a better orbital localization. For the biggest active space chosen, an active space of 100 electrons in 100 orbitals, λ_Q is reduced by an order of magnitude. This illustrates the potential of using widely available and computationally cheap classical pre-optimizations, together with our dedicated orbital-optimizer, to reduce the cost of quantum simulations.

2.5 Conclusions

Numerically we have seen that exploiting the locality of the basis gives rise to a lower variance in the Hamiltonian coefficients, reducing λ_Q. This results in a significant reduction of the absolute values of the integrals. The off-diagonal elements of the two-electron tensor play the biggest role here.

To find an even better basis-rotation that minimizes λ_Q, one can use a "brute-

Molecule	α_{HF}	R^2_{HF}	α_{PM}	R^2_{PM}	α_{FB}	R^2_{FB}	α_{ER}	R^2_{ER}	α_{OO}	R^2_{OO}
FeMoco	2.83	0.999	2.03	0.948	1.94	0.958	1.98	0.949	**1.89**	0.969
Ru SI I	2.63	0.996	2.36	0.947	**2.17**	0.987	2.25	0.976	2.21	0.993
Ru SI II	2.83	0.999	2.27	0.953	**2.11**	0.987	2.22	0.977	2.20	0.992
Ru TS II-III	2.95	0.999	2.16	0.892	**1.98**	0.966	2.13	0.923	2.12	0.995
Ru SI V	2.75	0.998	2.42	0.906	**2.13**	0.980	2.33	0.953	2.21	0.993
Ru SI VIII	2.84	0.999	2.22	0.863	**2.12**	0.966	2.18	0.932	2.19	0.990
Ru TS VIII-IX	2.82	0.999	2.23	0.865	**2.09**	0.971	2.19	0.904	2.19	0.988
Ru SI IX	2.8	0.999	2.42	0.924	**2.22**	0.988	2.34	0.966	2.23	0.997
Ru SI XVIII	2.79	0.999	2.31	0.930	**2.12**	0.985	2.26	0.956	2.24	0.993

Table 2.4: Scaling of λ_Q with respect to the active space size N_{act} for different orbital bases, assuming $\log \lambda_Q = \alpha \log N_{\text{act}} + \beta$ with the associated R^2 value. The minimal ANO-RCC basis-set is considered.

force" orbital optimizer as we have done in this work. This method could be realized in practice only because we have a efficient way of computing λ_Q directly in terms of the molecular integrals [Eq. (2.16)], avoiding doing a fermion-to-qubit mapping explicitly. Expressing it as $\lambda_Q = \lambda_T + \lambda'_V$, this improves a bit over the equation of λ_V in Ref. 76, holding into account the representation of the Hamiltonian in terms of unique Pauli strings. The direct orbital optimizer can be used in quite large active spaces, reliably converging for active spaces with up to $N_{\text{act}} \approx 80$ orbitals, sometimes more. For the largest active spaces considered, it converges very slowly due to the need of 4-index transformations at every step which are costly because of the large dimensionality of the system, combined with the non-continuous and complicated landscape of λ_Q. This gives rise to the need to either set a high threshold of convergence or cut off the optimizer at some point where the user is satisfied with the result. While this is problematic, there is no a priori reason not to try this optimizer on the (potentially large) active spaces of the molecule one is considering. This can be a point of further study.

We benchmarked a range of various molecules and active spaces for which we showed to achieve a significant reduction of λ_Q. Apart from this benchmark, we investigated the scaling of λ_Q with the size of the system N. There are multiple paths one can take here to increase N. A popular way to do this in the literature is considering either a hydrogen chain or hydrogen ring, and increase the amount of atoms. We benchmarked the scaling of λ_Q for a hydrogen chain and a linear alkane chain by increasing the number of atoms as well, where our results show a scaling of $\mathcal{O}(N^{2.3})$ and $\mathcal{O}(N^{2.2})$, respectively. Here we saw localized orbitals can give approximately a factor of N lower scaling of $\mathcal{O}(N^{1.3})$ and $\mathcal{O}(N^{1.4})$, respectively. As this gives one limited information how big λ_Q will be in real eventual applications of quantum algorithms, we decided to investigate the scaling on relevant highly correlated molecules such as FeMoco (important in nitrogen fixation) and Ruthenium metal complexes (important in carbon dioxide capture). Even though we used a minimal basis to make the scaling calculations feasible on these molecules, there is no reason to believe λ_Q will scale differently with respect to the active space size on a bigger basis-set. Here we showed that the scaling is a factor of N larger, and with localized orbitals can be brought down to $\lambda_Q \approx \mathcal{O}(N^2)$, depending on the molecule in consideration. Our dedicated orbital-optimizer was able to consistently give an even further improvement of 10-25% on these molecules. As a concrete example, consider the largest active space considered of FeMoco: this factor of N difference would result in a reduction by two orders of magnitude in the amount of measurements needed for tomography [Eq. (2.11)], if one wants chemical accuracy. This indicates that the simple and efficient classical pre-processing by widely available localization techniques, together with our dedicated optimizer, will help to make simulations with large active spaces feasible.

2.A Electronic integrals transformation

Preparing the electronic Hamiltonian in a given orthogonal orbital basis $\{\phi_p(\mathbf{r})\}_{p=1}^N$ is a crucial step for realizing concrete quantum computing applications. In practice, such a process can be realized classically via electronic integral transformations. For this, we assume that the orthogonal orbitals can be expressed as a linear combination of AOs such that

$$\phi_p(\mathbf{r}) = \sum_\mu \chi_\mu(\mathbf{r}) C_{\mu p}, \tag{2.34}$$

where we assume that the functions $\chi_\mu(\mathbf{r})$ as well as the coefficient matrix $\mathbf{C}$ are real-valued, as is usually done in non-relativistic quantum chemistry. Based on our knowledge of this matrix $\mathbf{C}$, one can transform the one- and two-electron integrals from the AO basis to the orthogonal orbital basis. To do so, one starts from the one- and two-electron integrals in the AO basis, respectively $h_{\mu\nu}$ and $g_{\mu\nu\gamma\delta}$, and implements the following two- and four-indexes transformations:

$$h_{pq} = \sum_{\mu\nu} h_{\mu\nu} C_{\mu p} C_{\nu q}, \tag{2.35}$$

and

$$g_{pqrs} = \sum_{\mu\nu\gamma\delta} g_{\mu\nu\gamma\delta} C_{\mu p} C_{\nu q} C_{\gamma r} C_{\delta s}. \tag{2.36}$$

From a computational point of view, the numerical cost of processing electronic integrals is essentially governed by the four-indexes transformation. This transformation is known to scale as $\mathcal{O}(N^5)$ (see Ref. 178), when not employing approximations such as density fitting (see Ref. 179).

2.B Explicit form of λ_Q in terms of molecular integrals

To derive a formula for λ_Q in terms of the molecular integrals, one needs to keep track of what happens to the coefficients in the Hamiltonian (Eq. (1.14)) after a fermion-to-qubit transformation of the fermionic operators. This is non-trivial, since e.g. the Jordan-Wigner transformation transforms a single arbitrary product of fermionic operators $a_p^\dagger a_r^\dagger a_s a_q$ into 16 different Pauli strings, because of the products of $\frac{X-iZ}{2}$ and $\frac{X+iZ}{2}$. The Hamiltonian in the form of (1.14) is actually not best suited for this derivation. Instead, it is helpful to express the Hamiltonian in terms of Majorana operators that have the convenient property of being Hermitian operators that square to identity:

$$\gamma_{p\sigma,0} = a_{p\sigma} + a_{p\sigma}^\dagger, \quad \gamma_{p\sigma,1} = -i\left(a_{p\sigma} - a_{p\sigma}^\dagger\right), \tag{2.37}$$

$$\{\gamma_i, \gamma_j\} = 2\delta_{ij}\mathcal{I}, \quad \gamma_i^\dagger = \gamma_i, \quad \gamma_i^2 = \mathcal{I}. \tag{2.38}$$

To find a representation of the Hamiltonian in terms of Majorana operators, one could directly replace $a_p^\dagger = \frac{\gamma_{p,0} - i\gamma_{p,1}}{2}$ and $a_q = \frac{\gamma_{q,0}+i\gamma_{q,1}}{2}$ in the Hamiltonian, or follow the procedure in Ref. 11. We chose the latter approach, detailed below.

It is well known that the electron repulsion integral tensor, when written as a $N^2 \times N^2$ matrix $g_{(pq),(rs)}$ with the composite indices pq and rs, is positive semi-definite. This makes it possible to define a Cholesky decomposition $\mathbf{g} = \mathbf{L}\mathbf{L}^\dagger$. We write:

$$g_{pqrs} = \sum_\ell \sum_{pqrs} L_{pq}^\ell L_{rs}^\ell. \tag{2.39}$$

with $\mathbf{L}$ a lower triangular matrix. Since g_{pqrs} is symmetric in p, q and in r, s, for a given ℓ, L_{pq}^ℓ is an $N \times N$ symmetric matrix (also called a Cholesky vector in the Cholesky decomposition). This leads to the following expression for the Hamiltonian:

$$\begin{aligned}
H &= \sum_{pq}\sum_\sigma h_{pq} a_{p\sigma}^\dagger a_{q\sigma} + \frac{1}{2}\sum_{pqrs}\sum_{\sigma\tau} g_{pqrs} a_{p\sigma}^\dagger a_{r\tau}^\dagger a_{s\tau} a_{q\sigma} \\
&= \sum_{pq\sigma}\left[h_{pq} - \frac{1}{2}\sum_r g_{prrq}\right] a_{p\sigma}^\dagger a_{q\sigma} \\
&\qquad + \frac{1}{2}\sum_\ell \sum_{pqrs}\sum_{\sigma\tau} L_{pq}^\ell L_{rs}^\ell a_{p\sigma}^\dagger a_{q\sigma} a_{r\tau}^\dagger a_{s\tau} \\
&= \sum_{pq\sigma} h'_{pq} a_{p\sigma}^\dagger a_{q\sigma} + \frac{1}{2}\sum_\ell \left(\sum_{pq\sigma} L_{pq}^\ell a_{p\sigma}^\dagger a_{q\sigma}\right)^2
\end{aligned} \tag{2.40}$$

where $h'_{pq} = h_{pq} - \frac{1}{2}\sum_r g_{prrq}$. Using the relation

$$a^\dagger_{p\sigma}a_{q\sigma} + a^\dagger_{q\sigma}a_{p\sigma} = \begin{cases} \mathcal{I} + i\left(\gamma_{p\sigma,0}\gamma_{p\sigma,1}\right), & p = q \\ \frac{i}{2}\left(\gamma_{p\sigma,0}\gamma_{q\sigma,1} + \gamma_{q\sigma,0}\gamma_{p\sigma,1}\right), & p \neq q \end{cases} \tag{2.41}$$

one can show that

$$\sum_{pq\sigma} M_{pq} a^\dagger_{p\sigma}a_{q\sigma} = \sum_p M_{pp}\mathcal{I} + \frac{i}{2}\sum_{pq\sigma} M_{pq}\gamma_{p\sigma,0}\gamma_{q\sigma,1}, \tag{2.42}$$

where M_{pq} can be replaced by any symmetric matrix (like L^ℓ_{pq} or h'_{pq}). Employing Eq. (2.42) in Eq. (2.40), leads to the expression of the Hamiltonian in terms of Majorana operators:

$$\begin{aligned} H =& \sum_p h'_{pp}\mathcal{I} + \frac{i}{2}\sum_{pq\sigma} h'_{pq}\gamma_{p\sigma,0}\gamma_{q\sigma,1} \\ &+ \frac{1}{2}\sum_\ell\left(\sum_p L^\ell_{pp}\mathcal{I} + \frac{i}{2}\sum_{pq\sigma} L^\ell_{pq}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\right)^2. \end{aligned} \tag{2.43}$$

After working out the square, and substituting back $\sum_\ell \sum_{pqrs} L^\ell_{pq}L^\ell_{rs} = g_{pqrs}$, we obtain

$$\begin{aligned} H =& \left(\sum_p h_{pp} + \frac{1}{2}\sum_{pr} g_{pprr} - \frac{1}{2}\sum_{pr} g_{prrp}\right)\mathcal{I} \\ &+ \frac{i}{2}\sum_{pq\sigma}\left(h_{pq} + \sum_r g_{pqrr} - \frac{1}{2}\sum_r g_{prrq}\right)\gamma_{p\sigma,0}\gamma_{q\sigma,1} \\ &- \frac{1}{8}\sum_{pqrs\sigma\tau} g_{pqrs}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\gamma_{r\tau,0}\gamma_{s\tau,1}. \end{aligned} \tag{2.44}$$

We now want to determine the value of λ_Q after a fermion-to-qubit mapping of this Hamiltonian in which all products of Majorana operators should be distinct. We use a fermion-to-qubit mapping such that single Majorana operators are mapped to single unique Pauli strings, such as the Jordan-Wigner transformation [55]:

$$\gamma_{i,0} \rightarrow X_i Z_{i-1}\ldots Z_0, \text{ and } \gamma_{i,1} \rightarrow Z_i Z_{i-1}\ldots Z_0, \tag{2.45}$$

where we used the composite index $i = p\sigma$. Note this is not only true for the Jordan-Wigner transformation, but also for for example the Bravyi-Kitaev transformation [180]. The only thing left to do is to make sure that the sum over 4 indices used to evaluate λ_Q is indeed truly quartic. Looking at Eq. (2.44), this is not yet the case. If $p\sigma = r\tau$ or $q\sigma = s\tau$ the corresponding Majorana operators will square to identity and will reduce to quadratic and identity terms. Below we employ the permutational symmetry of the real-valued integrals g_{pqrs} to obtain the most compact representation.

We first reorder the Majorana's and distinguish between the same and opposite spin cases:

$$\begin{aligned}
&-\frac{1}{8}\sum_{pqrs\sigma\tau} g_{pqrs}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\gamma_{r\tau,0}\gamma_{s\tau,1}\\
=&\frac{1}{8}\sum_{pqrs}\sum_{\sigma\tau} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\tau,0}\gamma_{q\sigma,1}\gamma_{s\tau,1}\\
=&\frac{1}{8}\sum_{pqrs}\sum_{\sigma} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
+&\frac{1}{8}\sum_{pqrs}\sum_{\sigma\neq\tau} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\tau,0}\gamma_{q\sigma,1}\gamma_{s\tau,1}
\end{aligned}\tag{2.46}$$

For the opposite spin case all operator products are unique and quartic, so no further work is needed. For the same spin case it is useful to identify cases in which two or more indices of the same type of Majorana operators are equal:

$$\begin{aligned}
&\frac{1}{8}\sum_{pqrs}\sum_{\sigma} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
=&\frac{1}{8}\sum_{pq}\sum_{\sigma} g_{pqpq}\gamma_{p\sigma,0}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\gamma_{q\sigma,1}\\
+&\frac{1}{8}\sum_{p\neq r,q}\sum_{\sigma} g_{pqrq}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{q\sigma,1}\\
+&\frac{1}{8}\sum_{q\neq s,p}\sum_{\sigma} g_{pqps}\gamma_{p\sigma,0}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
+&\frac{1}{8}\sum_{p\neq r,q\neq s}\sum_{\sigma} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
=&\frac{1}{4}\sum_{pq} g_{pqpq}\mathcal{I}\\
+&\frac{1}{8}\sum_{p\neq r,q\neq s}\sum_{\sigma} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1},
\end{aligned}\tag{2.47}$$

where we applied the identity relation $\gamma_i^2 = \mathcal{I}$ and for the second term used that integrals are symmetric in exchanging p and r while the product of Majorana operators is antisymmetric under this exchange. This nullifies this and the third term. The first term can be absorbed in the scalar term. For the truly quartic term we may make further use of permutational symmetries to get the most compact form:

$$
\begin{aligned}
&\frac{1}{8}\sum_{p\neq r, q\neq s}\sum_{\sigma} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
=&\frac{1}{8}\sum_{p>r, s>q}\sum_{\sigma}(\\
&g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
+&g_{rqps}\gamma_{r\sigma,0}\gamma_{p\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
+&g_{psrq}\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{s\sigma,1}\gamma_{q\sigma,1}\\
+&g_{rspq}\gamma_{r\sigma,0}\gamma_{p\sigma,0}\gamma_{s\sigma,1}\gamma_{q\sigma,1})\\
=&\frac{1}{4}\sum_{p>r, s>q}\sum_{\sigma}(g_{pqrs}-g_{psrq})\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}
\end{aligned}
\tag{2.48}
$$

Combining all terms we end up with the expression:

$$
\begin{aligned}
H = &\left(\sum_{p}^{N} h_{pp} + \frac{1}{2}\sum_{pr}^{N} g_{pprr} - \frac{1}{4}\sum_{pr}^{N} g_{prrp}\right)\mathcal{I}\\
&+\frac{i}{2}\sum_{pq\sigma}^{N}\left(h_{pq} + \sum_{r}^{N} g_{pqrr} - \frac{1}{2}\sum_{r}^{N} g_{prrq}\right)\gamma_{p\sigma,0}\gamma_{q\sigma,1}\\
&+\frac{1}{4}\sum_{p>r,s>q}^{N}\sum_{\sigma}\left(g_{pqrs}-g_{psrq}\right)\gamma_{p\sigma,0}\gamma_{r\sigma,0}\gamma_{q\sigma,1}\gamma_{s\sigma,1}\\
&+\frac{1}{8}\sum_{pqrs}^{N}\sum_{\sigma\neq\tau} g_{pqrs}\gamma_{p\sigma,0}\gamma_{r\tau,0}\gamma_{q\sigma,1}\gamma_{s\tau,1}.
\end{aligned}
\tag{2.49}
$$

We then take the sum of absolute values of the coefficients and perform the sums over spin in the quartic term explicitly (amounting to a factor of 2), to get the following form of λ_Q:

$$
\lambda_Q \quad = \quad \lambda_C + \lambda_T + \lambda_V', \tag{2.50}
$$

where λ_C corresponds to the constant term in Eq. (2.49), λ_T to the quadratic and λ_V to the quartic term in Majorana operators. They have the form:

$$
\lambda_C \quad = \quad \left|\sum_{p}^{N} h_{pp} + \frac{1}{2}\sum_{pr}^{N} g_{pprr} - \frac{1}{4}\sum_{pr}^{N} g_{prrp}\right|, \tag{2.51}
$$

$$\lambda_T = \sum_{pq}^{N} \left| h_{pq} + \sum_{r}^{N} g_{pqrr} - \frac{1}{2} \sum_{r}^{N} g_{prrq} \right|, \tag{2.52}$$

$$\lambda_V' = \frac{1}{2} \sum_{p>r,s>q}^{N} |g_{pqrs} - g_{psrq}| + \frac{1}{4} \sum_{pqrs}^{N} |g_{pqrs}|. \tag{2.53}$$

This λ_V' is slightly different to the λ_V of Ref. 76 in calculating the λ for the "sparse" algorithm of Berry et al. [75], which should actually be the same as λ_Q. This is due that we took into account the swapping of majorana operators in Eq. (2.48). In Ref. 76, this corresponds to the term $V' = \frac{1}{8} \sum_{\alpha,\beta\in\{\uparrow,\downarrow\}} \sum_{p,q,r,s} V_{pqrs} Q_{pq\alpha} Q_{rs\beta}$. It is hard to see here that the product $Q_{pq\alpha} Q_{rs\alpha}$ is anti-symmetric in swapping p, r and q, s, but it becomes clear when one realizes that $Q_{pq\alpha} = i\gamma_{p\sigma,0}\gamma_{q\sigma,1}$, indicating the usefulness of majorana operators. As the absolute values give that:

$$|g_{pqrs} - g_{psrq}| \leq |g_{pqrs}| + |g_{psrq}| \;\forall\; p, q, r, s \tag{2.54}$$

such that $\lambda_V' < \lambda_V$ (except when g_{pqrs} and g_{psrq} always have opposite sign, in which case they would be equal).

CHAPTER 3

FragPT2: Multi-Fragment Wavefunction Embedding with Perturbative Interactions

3.1 Introduction

Multi-configurational (MC) wavefunction-based methods have long been the workhorse of ab-initio quantum chemistry, particularly for systems with low-lying or degenerate electronic states [181, 182]. Practical MC approaches, such as the complete active space self-consistent field (CASSCF) [27], require defining an active space comprising a subset of the most chemically relevant orbitals. Within this space, electron correlations are calculated exactly by a configuration interaction (CI) wavefunction, a superposition of all electronic configurations formed from a given set of active electrons and orbitals. The number of these configurations scales exponentially with the size of the active space, limiting the application of these methods to small systems. There have been substantial efforts to expand the size of the active space: some try to restrict the number of excitations by partitioning the active space [183–188], others involve adaptive procedure to select the configurations with the largest weights [22, 189]. Radically different approaches to constructing a compressed CI wavefunction include tensor-network algorithms such as the density matrix renormalization group (DMRG) [24], quantum monte carlo (QMC) methods [25], or various kinds of quantum algorithms [9].

A more pragmatic approach for extending multi-configurational computations to larger systems relies on the concepts of *fragmentation* and *embedding* [190–193]. Fragmentation exploits the inherent locality of the problem, describing a system as a composition of simpler subsystems. Each subsystem is then treated with a higher level of theory. The subsystems are then recombined by embedding them in each other's environment at a lower level of theory. The subsystem orbitals can be constructed in various ways, with the most prominent method being Density Matrix Embedding Theory (DMET) [83, 84, 194, 195]. DMET constructs fragment and bath orbitals based on the Schmidt decomposition of a trial low-level (eg. Hartree-Fock) single-determinant wavefunction of the full system. A high-level calculation (e.g. FCI, Coupled-Cluster [85, 196], CASSCF [197], DMRG [85, 198, 199] or auxiliary-field QMC [200]) is then performed on the fragment orbitals. Subsequently, the low-level wavefunction is fine-tuned self-consistently via the introduction of a local correlation potential. Fragmentation and embedding have also been studied in the context of DFT [201, 202]. MC wavefunction-based methods that explicitly construct localized active spaces for each fragment include the Active Space Decomposition method [203], cluster Mean Field (cMF) [204] and Localized Active Space Self-Consistent Field (LASSCF) [149, 150].

While fragmentation methods have shown success in reducing the complexity in treating localized static correlations, they typically don't capture inter-fragment correlations. Especially weak, dynamical, correlations between the different fragments and between fragments and their environment can be crucial for obtaining an accurate description of the full system [205]. In CAS methods, the fragment-environment correlations can be retrieved using Multi-Reference Perturbation Theory (MRPT) [206] methods like Complete Active Space Second-Order Perturbation Theory (CASPT2) [35] and N-Electron Valence Second-Order Perturbation Theory (NEVPT2) [36, 37]. Some methods have been developed to also recover inter-fragment correlations in embedding schemes either variationally [207], perturbatively [204, 208–210], or via a coupled-cluster approach [211]. Although treating strong correlations between fragments remains challenging, there has been some work in this direction [212, 213]. In the field of quantum algorithms, a recent work proposed to treat inter-fragment entanglement with a Unitary Coupled Cluster ansatz using the LASSCF framework [214].

In this work, we introduce and benchmark a novel active space embedding framework, which we call FragPT2. Based on a user-defined choice of two molecular fragments (defined as a partition of the atoms in the molecule), we employ a top-down localization scheme that generates an orthonormal set of localized molecular orbitals, ordered by quasi-energies and assigned to a specific fragment. Using these localized orbitals, we define separate and orthogonal fragment active spaces. Our orbital fragmentation scheme is straightforward, it does not require iterative optimization, and it allows to define fragment orbitals even when the fragments are covalently bonded; on the downside, a good choice of fragments based on chemical intuition is crucial for the success of our method. Within each fragment's active space, we self-consistently find the MC ground state influenced by the mean field of the other fragment (defined as a function of the fragment 1-particle reduced density matrix).

The factorized state obtained with our method has a similar structure to the wavefunction used in LASSCF and cMF, as these methods also construct product state wavefunctions of MC states defined on fragmented active spaces. The cMF method is designed for the 1D and 2D Fermi-Hubbard model. It is based on expressing the ground state wavefunction as a tensor product of many-body states defined on local fragments. The fragment orbitals are then optimized self-consistently to minimize the total energy of the considered product state. Inter-fragment correlations are then recovered in second-order perturbation theory, using excited fragment eigenstates as perturbing functions. On the other hand, LASSCF exploits a modified DMET algorithm to construct fragments. Starting from a product state, a Schmidt decomposition is used to define fragment and bath orbitals for each fragment. Similarly to cMF, the product state and fragment definition are then optimized self-consistently. The resulting method can be made fully variational with respect to both CI and orbital coefficients [150]. In contrast, in our approach, active fragment orbitals are defined in top-down fashion, starting from a set of reference canonical molecular orbitals. Our method is variational with respect to the considered (fragment CI) parameters, and does not require any orbital optimization. As a trade-off for the simplicity of the method, we expect our product wavefunction to have a higher energy than the orbital-optimized LASSCF for the same fragment active space sizes. We instead aim to recover the remaining inter-fragment correlations perturbatively.

To this end, our product state will be used as a starting point for MRPT to recover inter-fragment correlations. The interactions between fragments can be naturally classified on the basis of charge and spin symmetries imposed on the single fragments, offering analytic insight into the nature of these correlations. Differently from cMF, the perturbing functions are chosen on the basis of electronic excitation operators present in the original electronic Hamiltonian, and organized according to a partially contracted basis akin to MRPT methods like PC-NEVPT2 [36, 37]. We apply our method to challenging covalently and non-covalently bonded fragments with moderate to strong correlation, providing qualitative estimates of the contributions from various perturbations to the total correlation energy within the active space.

The rest of this chapter is organized as follows: in Section 3.2 we detail our FragPT2 algorithm for multi-reference fragment embedding. In Section 3.3, we perform numerical tests of the method on a range of challenging chemical systems, ranging from the non-covalently bonded but strongly correlated N_2 dimer to covalently bonded aromatic dimers and the butadiene molecule. In Section 3.4 we present an outlook on future research directions, proposing possible improvements for the method and an application in the field of FragPT2 in the field of variational quantum algorithms. Finally, in Section 3.5 we give concluding remarks.

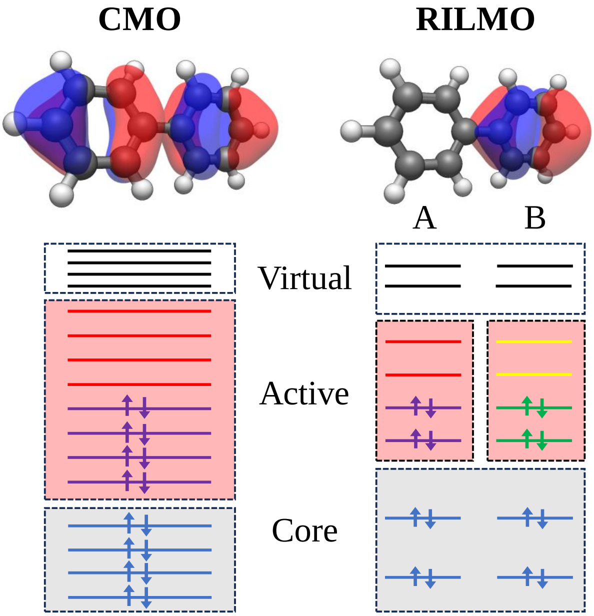


Figure 3.1: Example of fragmentation and definition of the fragment active spaces. **(Left)** Active space selection for the entire biphenyl molecule. The CAS treatment separates the canonical molecular orbitals (CMOs) based on their energy ordering, obtaining a set of doubly-occupied core orbitals, a set of empty virtual orbitals, and a set of active orbitals around Fermi energy used to describe correlations. We illustrate the highest occupied molecular orbital. **(Right)** Fragment active space selection for the left and right fragments of the biphenyl molecule. After the localization procedure, we obtain Recanonicalized Intrinsic Localized Molecular Orbitals (RILMOs), where the orbitals are assigned to either fragment A or B. We can still select core, active and virtual orbitals for each fragment based on an approximate energy ordering, obtained through the recanonicalization procedure. Here we depict the highest occupied RILMO for the right fragment. Using our method, we can half the size of the required active space since the multi-reference solver is applied to just one fragment at a time. The correlations between the localized active spaces can be retrieved afterwards with perturbation theory.

3.2 FragPT2 method

In this section, we introduce a novel method for fragmented multi-reference calculations with perturbative corrections: FragPT2. This method works by dividing the active space of a molecule into localized subspaces that can be treated separately using a MC solver, as illustrated in Figure 3.1. The cost of MC methods scales quickly with the size of the treated active space (e.g. exponentially in the case of FCI); splitting the system into smaller active spaces allows the treatment of larger systems for an affordable computational cost. In this work, we focus on the special case of two active fragments called A and B; however, our method can be promptly generalized to the multiple fragment case as discussed in Section 3.4.3. Our method requires the user to define the molecular fragments as an input. The choice of fragmentation should be based on chemical intuition, aiming at minimizing inter-fragment correlations; a good choice is crucial to the success of the method. Our method allows to recover some inter-fragment correlations, allowing fragmentations that *break a covalent bond* (like the one shown in Figure 3.1 for biphenyl), i.e. where two atoms on either side of a covalent bond are assigned to different fragments. The number of bonds broken in fragmentation should, however, be kept to a minimum.

First, in Sec 3.2.1 we introduce the construction of the localized orbitals and the definition of the fragment active spaces. In Section 3.2.2 we define fragment Hamiltonians by embedding each fragment in the mean field of the other. Applying separate MC solvers to each fragment Hamiltonian, we show how to obtain a fragment product state $|\Psi_0\rangle$ which will be the reference state for subsequent perturbative expansions. Finally, in Section 3.2.3 we decompose the full Hamiltonian into a sum of the solved fragment Hamiltonians and a number of inter-fragment interaction terms. We classify these terms on the basis of fragment symmetries and describe a method to treat them in second-order perturbation theory.

3.2.1 Construction of re-canonicalized intrinsic localized molecular orbitals

In order to define the fragment subspaces, we follow the top-down procedure introduced in references 169 and 215, based on localizing pre-computed molecular orbitals. First, we calculate a set of canonical molecular orbitals (CMOs) for the whole system (other choices for molecular orbitals are discussed in Section 3.5). Distinct Hartree-Fock calculations are also run on each fragment, capped if necessary to saturate bonds severed in the fragmentation. We then choose a valence space, removing a set of hard-core and hard-virtual orbitals far from Fermi energy in both the supermolecular and the fragment calculations. The remaining valence fragment orbitals define the target localized active spaces and are called reference fragment orbitals (RFOs). These RFOs are non-orthogonal and only serve to depolarize the valence CMOs, providing an orthonormal set of intrinsic fragment orbitals (IFOs) of the same dimension as the RFO basis. These IFOs are expressed in the CMO basis and could already be assigned to a particular fragment. They do however mix occupied and virtual spaces and we therefore merely use them to define the localization function in Pipek-Mezey localization [155] of the CMOs. After recanonicalization (block-diagonalizing the Fock matrix within each fragment), we obtain a set of Recanonicalized Intrinsic Localized Molecular Orbitals (RILMOs), partitioned in fragment subspaces, that together span exactly the occupied space of the original CMOs [169] plus the chemically relevant valence virtual space. The active spaces for each fragment are illustrated in Figure 3.1.

In this work we also consider covalently bonded fragments, where there is an ambiguity in assigning one occupied orbital representing the inter-fragment bond to either fragment. The same ambiguity holds for one unoccupied antibonding orbital, which can also be assigned to either fragment. To eliminate this arbitrariness, we introduce a bias so that any such (anti-)bond is always assigned to the first fragment. This enables us to define a natural fragmentation for covalently bonded dimer molecules. As noted already above, in order to generate the required IFO basis for this calculation, we need to deal with "dangling" bonds that are severed in the fragmentation process. For each fragment we simply saturate these by adding a hydrogen atom to the fragment. The thus produced fragment orbitals are well suited as RFOs, but do yield one additional orbital in the span of the RFOs and IFOs. Accepting this feature, the ROSE code reported in reference 169 could be used without modification. In a forthcoming paper, we plan to discuss the localization of higher lying virtuals for which the RILMO generation does need to be modified (see also reference 215 for non-covalently bonded subsystems) by removing the capping basis from the RFO space. For the covalently bonded dimer systems tested in this work, the unmodified RILMO generation could be used with only a bias in the selection procedure to assign both the bond and the antibond to the same fragment.

3.2.2 Fragment embedding

The total Hamiltonian in the combined active space spanned by both fragments is given by

$$H = \sum_{pq \in A \cup B} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs \in A \cup B} g_{pqrs} e_{pqrs}, \tag{3.1}$$

where we use the spin-adapted excitation operators of Eq. (1.12)-(1.13). This Hamiltonian includes all interactions of all active orbitals. Our embedding scheme aims at decomposing this Hamiltonian as $H = H^0 + H'$, where H^0 includes intra-fragment terms and a mean-field inter-fragment term, and can be solved exactly with separate in-fragment MC solvers. The residual inter-fragment interactions H' are treated separately with perturbation theory, as described in Section 3.2.3.

To facilitate the use of separate MC solvers for each fragment, we constrain the wavefunction of the total system to be a product state over the two fragments,

$$\left|\Psi^0\right\rangle = \left|\Psi_A\right\rangle \left|\Psi_B\right\rangle, \tag{3.2}$$

where $|\Psi_X\rangle$ is a many-body wavefunction in the active space of fragment X, similar in spirit to cMF and LASSCF. We further restrict each fragment wavefunction

$|\Psi_X\rangle$ to have fixed, integer charge and spin. Note that the conservation of spin and charge on each fragment is not a symmetry of the subsystem; however, this assumption is crucial to construct separate efficient MC solvers. Inter-fragment charge transfer and spin exchange processes are later treated in perturbation theory.

Under these constraints, we can simplify the expression of H by removing all the terms that do not respect charge and spin conservation on each fragment separately (as their expectation value of $|\Psi^0\rangle$ would anyway be zero). The remaining Hamiltonian can be then decomposed as $H_A + H_B + H_{AB}$, with terms

$$H_X = \sum_{pq\in X} h_{pq}E_{pq} + \frac{1}{2}\sum_{pqrs\in X} g_{pqrs}e_{pqrs} \tag{3.3}$$

(with $X \in \{A, B\}$), that only act non-trivially on a single fragment, and a term

$$H_{AB} = \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}E_{pq}E_{rs} \tag{3.4}$$

(where $g'_{pqrs} = g_{pqrs} - \frac{1}{2}g_{psrq}$), that includes interactions preserving local spin and charge. The term H_{AB} still introduces inter-fragment correlations; one way to make the fragments completely independent would be to also treat this term perturbatively (this is the choice made in SAPT [208]). However, including an effective mean-field interaction (originating from H_{AB}) in the non-perturbative solution improves the quality of our $|\Psi^0\rangle$.

To construct the effective Hamiltonian H_X^{eff} for each fragment we use a mean-field decoupling approach. We write the excitation operator as its mean added to a variation upon the mean: $E_{pq} = \langle E_{pq}\rangle + \delta E_{pq}$. The mean is just the one-particle reduced density matrix (1-RDM) of one of the fragments, $\gamma^X_{pq} = \langle\Psi_X|E_{pq}|\Psi_X\rangle$. By substituting in Eq. (3.4) we obtain

$$\sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\left[E_{pq}\gamma^B_{rs} + \gamma^A_{pq}E_{rs} - \gamma^A_{pq}\gamma^B_{rs} + \delta E_{pq}\delta E_{rs}\right]. \tag{3.5}$$

The term $\delta E_{pq}\delta E_{rs}$ will necessarily have zero expectation value on the product state Eq. (3.2), as $\langle\Psi_X|\delta E_{pq}|\Psi_X\rangle = 0$. Removing this term (which we will later treat perturbatively) we obtain the mean-field interaction

$$H_{\text{mf}} = \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\left[E_{pq}\gamma^B_{rs} + \gamma^A_{pq}E_{rs} - \gamma^A_{pq}\gamma^B_{rs}\right]. \tag{3.6}$$

We can finally define H^0 as

$$H^0 = H_A + H_B + H_{\text{mf}}, \tag{3.7}$$

where all terms are operators with support on only a single fragment, thus the

Algorithm 3.1: Fragment embedding

```
Input: Active space integrals $h_{pq}$, $g_{pqrs}$. Fragment-partitioned sets of
       orbitals: occupied $(O_\mathrm{A}, O_\mathrm{B})$, and virtual $(V_\mathrm{A}, V_\mathrm{B})$ subspaces.
       Convergence threshold $\tau$.
Output: Optimal product state $\left|\Psi^0\right\rangle = |\Psi_A\rangle\,|\Psi_B\rangle$

Start from a Hartree-Fock state: $\gamma^B_{pq} \leftarrow 2\delta_{pq}$ for $p, q \in O_B$;
$\Delta E$, $E^0 \leftarrow \infty$;
while $\Delta E > \tau$ do
    $|\Psi_A\rangle, E^\mathrm{eff}_A \leftarrow$ ground state of $H^\mathrm{eff}_A$ of Eq. (3.9) ;
    $\gamma^A_{pq} \leftarrow \langle\Psi_A|E_{pq}|\Psi_A\rangle$, $p, q \in A$;
    $|\Psi_B\rangle, E^\mathrm{eff}_B \leftarrow$ ground state of $H^\mathrm{eff}_B$ of Eq. (3.10);
    $\gamma^B_{rs} \leftarrow \langle\Psi_B|E_{rs}|\Psi_B\rangle$, $p, q \in B$;
    $E_\mathrm{mf} \leftarrow \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\gamma^A_{pq}\gamma^B_{rs}$;
    $\Delta E \leftarrow E^0 - (E^\mathrm{eff}_A + E^\mathrm{eff}_B - E_\mathrm{mf})$;
    $E^0 \leftarrow (E^\mathrm{eff}_A + E^\mathrm{eff}_B - E_\mathrm{mf})$
return $\left|\Psi^0\right\rangle = |\Psi_A\rangle\,|\Psi_B\rangle$
```

ground state $\left|\Psi^0\right\rangle$ of H^0 is a product state of the form Eq. (3.2). All the terms we removed from H to construct H^0 have zero expectation value on $\left|\Psi^0\right\rangle$, thus it is the *lowest energy product state that respects the on-fragment symmetries.*

To find $\left|\Psi^0\right\rangle$ we minimize $E_0 = \left\langle H^0\right\rangle$ by self-consistently solving separate ground state problems on each fragment. Consider the decomposition

$$E_0 = E_A + E_B + E_\mathrm{mf}, \tag{3.8}$$

where $E_X = \langle\Psi_X|H_X|\Psi_X\rangle$ can be evaluated on a single fragment X and $E_\mathrm{mf} = \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\gamma^A_{pq}\gamma^B_{rs}$ is the mean-field inter-fragment coupling depending on the fragment 1-RDMs. To find $|\Psi_A\rangle$ and $|\Psi_B\rangle$, we iteratively solve for the ground state of the following coupled Hamiltonians:

$$H^\mathrm{eff}_A = H_A + \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs} E_{pq}\gamma^B_{rs} \tag{3.9}$$

$$H^\mathrm{eff}_B = H_B + \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\gamma^A_{pq} E_{rs}, \tag{3.10}$$

thus minimizing all the terms Eq. (3.8). We outline the whole procedure in Algorithm 3.1. Note that this algorithm can be readily generalized to other MC solvers within the fragment that provide access to the state RDMs (e.g. the variational quantum eigensolver, discussed in Sec. 3.4.2).

3.2.3 Multi-reference perturbation theory

While the $\left|\Psi^0\right\rangle$ retrieved from Algorithm 3.1 is a solid starting point, it neglects the correlations between the fragments. If the fragments are sufficiently separated, we expect these correlations to be minimal and recoverable by perturbation theory. We propose using second-order perturbation theory to retrieve the correlation energy of these interactions. The inter-fragment interaction terms can be classified in four categories, based on whether they conserve charge and/or total spin on each fragment: dispersion H'_{disp} (which conserves both charge and spin of the fragments), single-charge transfer $H'_{1\text{CT}}$ and double-charge transfer $H'_{2\text{CT}}$ (that conserve charge nor spin), and triplet-triplet coupling H'_{TT} (that conserves charge but not local spin). Thus, the complete decomposition of the Hamiltonian reads:

$$H = H^0 + H'_{\text{disp}} + H'_{1\text{CT}} + H'_{2\text{CT}} + H'_{\text{TT}}. \tag{3.11}$$

The definition of these terms is given in Table 3.1 and their derivation is reported in Appendix 3.A. We will treat the different perturbations in Eq. (3.11) one at a time. First notice that for every perturbation in Eq. (3.11), the first order energy correction is zero: $E^1 = \left\langle\Psi^0\middle|H'\middle|\Psi^0\right\rangle = 0$. We will focus solely on the second order correction to the energy.

To proceed, we need to choose a basis of *perturbing functions* $\{|\Psi_\mu\rangle\}$ used to define the first-order correction to the wavefunction

$$\left|\Psi^1\right\rangle = \sum_\mu C_\mu \left|\Psi_\mu\right\rangle. \tag{3.12}$$

For the exact second order perturbation energy, we should consider all Slater determinants that can be obtained by applying the terms within H to the set of reference determinants. While this full space of perturbing functions is smaller than the complete eigenbasis of H, it is still unpractically large, and approximations need to be introduced. To choose a compact and expressive basis, we look at the perturbation under consideration. Every perturbative Hamiltonian can be expanded in a linear combination of two-body operators:

$$H' = \sum_{\mu\in A}\sum_{\nu\in B} g_{\mu\nu} O^A_\mu O^B_\nu. \tag{3.13}$$

where O^X_μ is either identity or a product of Fermionic operators on fragment X and $g_{\mu\nu}$ are combinations of one- and two-electron integrals. Their explicit form can be found in appendix 3.B.

Consider the following (non-orthogonal) basis:

$$\left|\Psi_{\mu\nu}\right\rangle = O^A_\mu O^B_\nu \left|\Psi^0\right\rangle. \tag{3.14}$$

This partially-contracted basis is a natural choice for compactly representing the

Perturbation	**Fragment matrix element**
$H'_{\text{disp}} = \sum_{pq\in A}\sum_{rs\in B} g'_{pqrs}\left(E_{pq}-\gamma^{A}_{pq}\right)\left(E_{rs}-\gamma^{B}_{rs}\right)$	$\langle\Psi_X\vert E_{lk}H^{\text{eff}}_X E_{tu}\vert\Psi_X\rangle$
$H'_{\text{1CT}} = \sum_{p\in A}\sum_{q\in B}\left[h_{pq}-\sum_{r\in A} g_{prrq}\right]E_{pq} + \sum_{p\in B}\sum_{q\in A}\left[h_{pq}-\sum_{r\in B} g_{prrq}\right]E_{pq} + \sum_{pqr\in A}\sum_{s\in B} g_{pqrs}E_{pq}\left[E_{rs}+E_{sr}\right] + \sum_{pqr\in B}\sum_{s\in A} g_{pqrs}E_{pq}\left[E_{rs}+E_{sr}\right]$	$\langle\Psi_X\vert a_m E_{lk}H^{\text{eff}}_X E_{tu}a^{\dagger}_v\vert\Psi_X\rangle$ $\langle\Psi_X\vert a^{\dagger}_m E_{lk}H^{\text{eff}}_X E_{tu}a_v\vert\Psi_X\rangle$
$H'_{\text{2CT}} = \frac{1}{2}\sum_{pr\in A}\sum_{qs\in B} g_{pqrs}E_{pq}E_{rs} + \frac{1}{2}\sum_{pr\in B}\sum_{qs\in A} g_{pqrs}E_{pq}E_{rs}$	$\langle\Psi_X\vert a_l a_k H^{\text{eff}}_X a^{\dagger}_t a^{\dagger}_u\vert\Psi_X\rangle$ $\langle\Psi_X\vert a^{\dagger}_l a^{\dagger}_k H^{\text{eff}}_X a_t a_u\vert\Psi_X\rangle$
$H'_{\text{TT}} = -\sum_{pq\in A}\sum_{rs\in B} g_{psrq}t_{pq,rs}$	$\langle\Psi_X\vert T^{(1,0)}_{lk}H^{\text{eff}}_X T^{(1,0)}_{tu}\vert\Psi_X\rangle$ $\langle\Psi_X\vert T^{(1,1)}_{lk}H^{\text{eff}}_X T^{(1,-1)}_{tu}\vert\Psi_X\rangle$ $\langle\Psi_X\vert T^{(1,-1)}_{lk}H^{\text{eff}}_X T^{(1,1)}_{tu}\vert\Psi_X\rangle$

Table 3.1: Summary of the perturbations and the cost of PT2. The left column reports the different inter-fragment perturbations, while the right column reports the form of the matrix elements of H^0 required to compute each perturbation; estimating these matrix elements on the fragment state is the most expensive part of FragPT2. If done naively, by writing out the full fragment Hamiltonians as a contraction between integrals, this could require estimating 4-RDMs for the H'_{disp}, H'_{2CT} and H'_{TT}, and 5-RDMs for H'_{1CT}.

wavefunctions that interact with $\left|\Psi^0\right\rangle$ through the perturbations in H' [36].

Following the choice of perturbing functions, we estimate the matrix elements $\left\langle\Psi_{\mu\nu}\middle|H^0\middle|\Psi_{\kappa\lambda}\right\rangle$ in this basis. The overlap $\langle\Psi_{\mu\nu}|\Psi_{\kappa\lambda}\rangle$ must also be computed in order to be able to contract with the $g_{\mu\nu}$ to yield $\left\langle\Psi^0\middle|H'\middle|\Psi_{\mu\nu}\right\rangle$. To obtain the coefficients $C_{\mu\nu}$ that define the first-order correction to the wavefunction, we solve the following linear equations:

$$\sum_{\kappa\lambda} \left\langle \Psi_{\mu\nu} \middle| H^0 - E^0 \middle| \Psi_{\kappa\lambda} \right\rangle C_{\kappa\lambda} + \left\langle \Psi_{\mu\nu} \middle| H' \middle| \Psi^0 \right\rangle = 0. \tag{3.15}$$

Then the second-order correction to the energy is given by:

$$E^2 = \left\langle \Psi^0 \middle| H' \middle| \Psi^1 \right\rangle = \sum_{\mu\nu} \left\langle \Psi^0 \middle| H' \middle| \Psi_{\mu\nu} \right\rangle C_{\mu\nu}. \tag{3.16}$$

The total second-order PT correction can be expressed as the sum of the different perturbations:

$$E^2 = E^2_{\mathrm{disp}} + E^2_{\mathrm{1CT}} + E^2_{\mathrm{2CT}} + E^2_{\mathrm{TT}}. \tag{3.17}$$

The procedure is summarized in Algorithm 3.2.

Algorithm 3.2: FragPT2

```
Input: Active space integrals and orbital partitioning as in Algorithm 3.1.
       Zeroth-order product state $|\Psi^0\rangle = |\Psi_A\rangle |\Psi_B\rangle$.
Output: Second-order perturbative corrections: $E^2_{\mathrm{disp}}$, $E^2_{\mathrm{1CT}}$, $E^2_{\mathrm{2CT}}$ and
        $E^2_{\mathrm{TT}}$

for $H' \in \{H'_{\mathrm{disp}}, H'_{\mathrm{1CT}}, H'_{\mathrm{2CT}}, H'_{\mathrm{TT}}\}$ do
    Choose perturbing functions as in Eq. (3.14);
    Compute matrix elements $\langle \Psi_{\mu\nu} | H^0 | \Psi_{\kappa\lambda} \rangle$;
    Compute matrix elements $\langle \Psi_{\mu\nu} | H' | \Psi^0 \rangle$;
    $C_{\mu\nu} \leftarrow$ solution of Eq. (3.15);
    $E^2 \leftarrow \sum_{\mu\nu} \langle \Psi^0 | H' | \Psi_{\mu\nu} \rangle C_{\mu\nu}$
return $E^2_{\mathrm{disp}}$, $E^2_{\mathrm{1CT}}$, $E^2_{\mathrm{2CT}}$ and $E^2_{\mathrm{TT}}$
```

Computing the matrix elements $\left\langle \Psi_{\mu\nu} \middle| H^0 \middle| \Psi_{\kappa\lambda} \right\rangle$ is the most expensive part of our algorithm. The tensor product form of the zeroth-order wavefunction significantly reduces the algorithm's cost by allowing the matrices to factorize in the expectation values of operators on the different fragments, that in turn can be expressed as combinations of fragment RDMs. We outline the idea here, and refer the reader to Appendix 3.B for the formal derivation for every perturbation:

$$\begin{aligned}\left\langle\Psi_{\mu\nu}\middle|H^0\middle|\Psi_{\kappa\lambda}\right\rangle = & \\ & \left\langle\Psi_A\middle|O_\mu^{A\dagger}H_A^{\text{eff}}O_\kappa^A\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O_\nu^{B\dagger}O_\lambda^B\middle|\Psi_B\right\rangle \\ & +\left\langle\Psi_A\middle|O_\mu^{A\dagger}O_\kappa^A\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O_\nu^{B\dagger}H_B^{\text{eff}}O_\lambda^B\middle|\Psi_B\right\rangle \\ & +E_{\text{mf}}\left\langle\Psi_A\middle|O_\mu^{A\dagger}O_\kappa^A\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O_\nu^{B\dagger}O_\lambda^B\middle|\Psi_B\right\rangle.\end{aligned} \tag{3.18}$$

If there are $N_A N_B$ two-body terms in Eq. (3.13), the number of matrix elements that one needs to estimate on each fragment is $\frac{1}{2}N_A N_B(N_A N_B + 1)$. However, if the amount of matrix elements becomes too expensive, it is possible to alleviate the cost without sacrificing much accuracy, for example by using a more compact basis of perturbing functions. For a discussion of further reductions of the cost, see Section 3.4.1.

3.3 Numerical demonstration

In this section we demonstrate our method by applying it to a range of molecular systems that are well-suited targets for bipartite fragmentation. We have chosen three sets of systems. The first system consists of two N_2 molecules at a distance of 2.0Å, with a (close to equilibrium) bond length of 1.2Å. In contrast to the other structures, we do not need to cut through a covalent bond and can treat each molecule as a separate fragment. We examine the results of our method while stretching the nitrogen bond in one of the fragments; this is known to rapidly increase static correlation in this system and thus is a good benchmark for the multi-reference method. The second type of systems we consider comprises a set of aromatic dimers, where two aromatic rings of different kinds are connected by a single covalent bond. Cutting through this bond, we investigate the correlation energies of the dimers with respect to the dihedral angle of the ring alignment. These systems exhibit strong correlation whenever the rings are in the same plane and low correlation when the rings are perpendicular to each other: they are thereby suitable to benchmark both regimes. The final system is butadiene, as the simplest example of the class of polyene molecules that are much studied as 1-D model systems [216, 217] as well as for their importance in various applications [218–220]. Here we cut through the single covalent bond between the middle carbons and investigate the correlation energy with respect to the stretching of the double bonds in a single fragment. This system, albeit slightly artificial, is intriguing due to the significant static correlation within the fragments induced by the dissociating bonds, coupled with substantial dynamic inter-fragment correlation.

3.3.1 Numerical simulation details

We construct the localized orbitals using a localization scheme implemented in the ROSE code [221]. The FragPT2 method is implemented completely inside the quantum chemical open-source software package PySCF [222]. Algorithm 3.1 uses the FCI solver of the program to get the optimal product state of the fragments. The matrix elements in Eq. (3.18) by exploiting the software capabilities to manipulate CI-vectors and estimate higher order RDMs. Finally we implemented Algorithm 3.2 that solves Eq. (3.15) and (3.16) for every perturbation in Eq. (3.11). To assess the accuracy of our algorithm, we compare the fragment embedding energy E^0 (from Algorithm 3.1), the FragPT2 energy $E^0 + E^2$ including the perturbative correction (from Algorithm (3.16)), and the exact ground state energy E^{exact} of the Hamiltonian in Eq. (3.1) (calculated with CASCI in an a full-molecule active space of dobule size). The N_2 dimer and aromatic dimer calculations are done in a cc-pVDZ basis set, while butadiene is treated in a 6-31G basis.

3.3.2 N_2 dimer

As an initial test system, we consider a dimer of nitrogen molecules, i.e. $N_2{-}N_2$. To increase the static correlation within the fragment, we dissociate one of the nitrogen molecules. This bond breaking is modeled using three occupied and three virtual localized orbitals in the active space, representing the σ bond and the two π bonds. This results in an active space of six electrons in six orbitals for each fragment. The results in Figure 3.2 clearly demonstrate the failure of the Hartree-Fock method due to the high degree of correlation within the fragment. Our multi-reference solver within the localized active spaces successfully addresses this issue, with E^0 providing a good description of the ground state. There is some minor inter-fragment correlation, and our perturbative correction brings us closer to the exact solution.

Our data further shows that the perturbative correction arises mainly from the single-charge transfer contribution. Notably, the double-charge transfer and triplet-triplet coupling are zero everywhere. Additionally, we find that for stretched bond lengths, the dispersion interaction between the fragments is minimal. The ability to identify the character of the relevant interactions between fragments is a further advantage of our method.

3.3.3 Aromatic dimers

Here we focus on aromatic dimers, i.e. molecules with two aromatic rings that are attached by a single covalent bond. The simplest such system considered is two phenyl rings, known as biphenyl, shown in Figure 3.1. As the biphenyl case is highly symmetric, other similar molecules can be generated by substituting various ligands for one of the hydrogen atoms, or a nitrogen for a carbon in

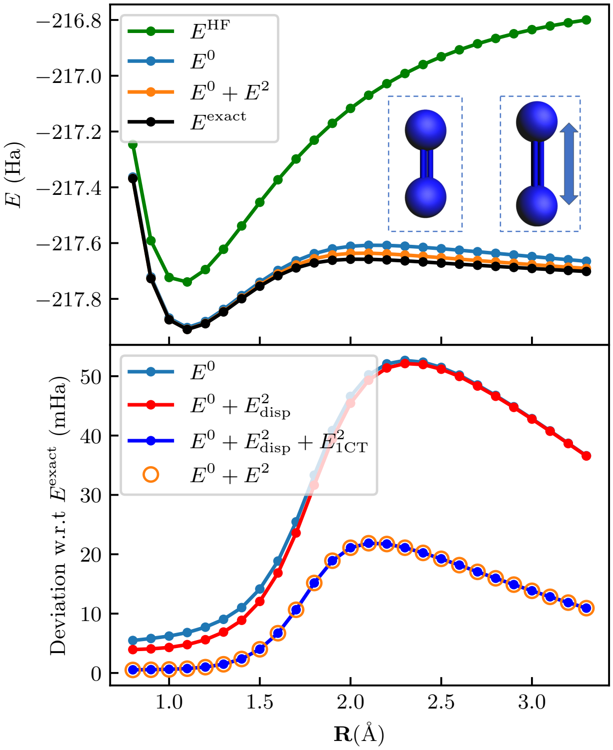


Figure 3.2: Potential energy curve for the N_2 dimer. The upper panel shows a comparison of the curves obtained through Hartree-Fock (E^{HF}), fragment embedding (E^0), FragPT2 ($E^0 + E^2$), and full-molecule CASCI (exact). The fragment active spaces each comprise six electrons in six spatial orbitals, corresponding to the triple bonding and anti-bonding orbitals. In contrast to Hartree-Fock, the fragment embedding energy E^0 captures the correct behaviour of the system, while E^2 gives an additional, small correction in the direction of the exact solution. The bottom panel reports the deviation w.r.t. the exact result over the potential energy curve, where we sequentially add the different perturbative corrections described in Table 3.1. We first add the dispersion correction E^2_{disp} (red line) and then the single-charge transfer contribution E^2_{1CT} (blue line), showing the other contributions are zero by plotting the full FragPT2 energy $E^0 + E^2$ (orange dots).

the phenyl rings. In this manner, we generate a comprehensive benchmark on a variety of systems. Our set of examples is motivated from the different classes of biaryl systems studied by Sanfeliciano *et al.* in the context of drug design [223].

---

[1] The considered molecules are: (a) biphenyl (b) 2-cyanobiphenyl (c) 2-(2-methoxyphenyl)pyridine (d) 2-(4-cyanophenyl)pyridine (e) 4-methoxybiphenyl (f)

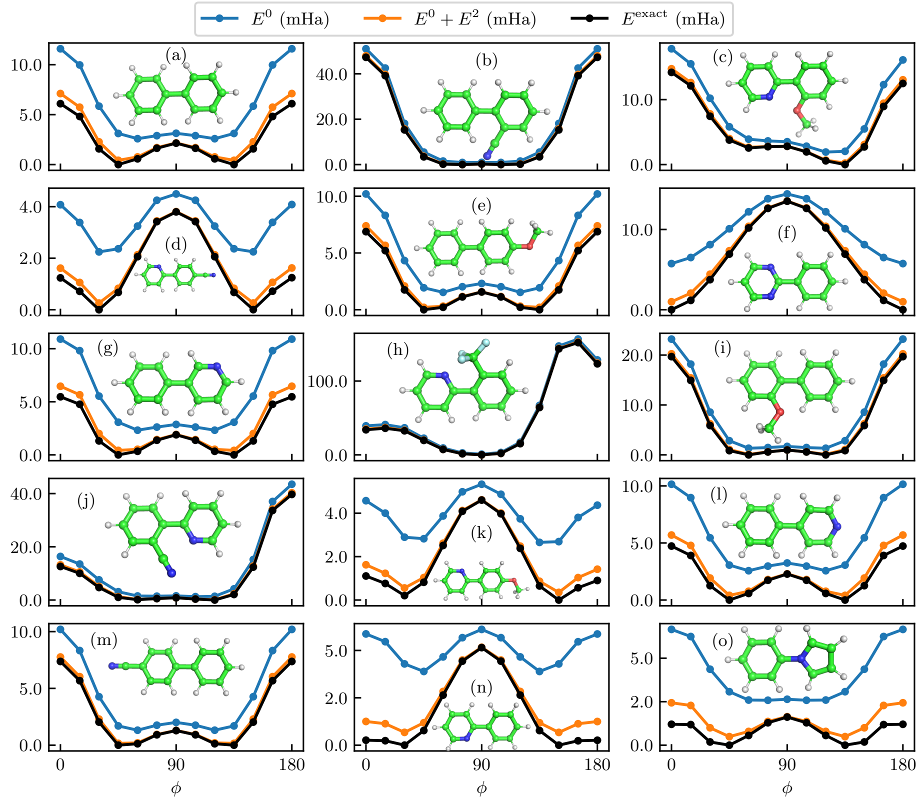


Figure 3.3: Potential energy curves for the set of aromatic dimers[1]. The orbitals are localized on the fragments naturally defined by the two aromatic rings (including the respective ligands). We select an active space of six electrons in six orbitals on each fragment that comprise the conjugated $\pi - \pi^*$ system (with some exceptions, see Section 3.3.3). This is small enough to verify our method against an exact calculation with twelve electrons in twelve orbitals. The blue line presents the fragment embedding energy E^0. The orange line includes the second-order perturbation energy for all considered perturbations, representing the FragPT2 energy $E^0 + E^2$. The black line reports the exact calculation E^{exact}.

To construct the fragment active spaces, we consider the conjugated $\pi - \pi^*$ system on each ring, typically resulting in six electrons distributed across six orbitals for each fragment. There are a few exceptions to this rule. For pyrrole rings, the relevant aromatic orbitals comprise six electrons in five orbitals.

2-phenylpyrimidine (g) 3-phenylpyridine (h) 2-(2-trifluoromethylphenyl)pyridine (i) 2-methoxybiphenyl (j) 2-(2-cyanophenyl)pyridine (k) 2-(4-methoxyphenyl)pyridine (l) 4-phenylpyridine (m)4-cyanobiphenyl (n) 2-phenylpyridine (o) N-phenylpyrrole.

Furthermore, for rings that include a CN or OCH_3 substituent [i.e. (c-f), (i-k) and (m) in Figure 3.3], there is a low-lying π orbital and high-lying π^* orbital that mix with a p orbital of the substituent. These orbitals are excluded from the active space of these fragments, reducing the active space to four electrons in four orbitals. This only provides additional insight into the performance of our method with asymmetric active space sizes in the fragments.

For each dimer, we vary the dihedral angle ϕ of the two planes spanned by the rings, thus rotating over the covalent bond. This gives a potential energy curve with a high variance of correlation energy: if the rings are perpendicular, the aromatic systems are localized and the correlation between the fragments is low. Instead, if the rings are aligned, we expect to see a high amount of correlation between the fragments, and thus a breakdown of the description of E^0. The results of our method compared to the exact energies are given in Figure 3.3.

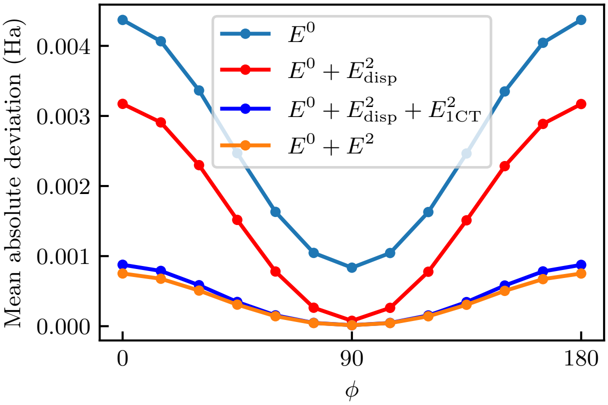


Figure 3.4: Average errors for the aromatic dimer set. Mean Absolute Deviation in total energy with respect to the exact result for the complete set of aromatic dimers shown in Figure 3.3, where we vary the dihedral angle ϕ of the two aromatic rings. We show the result of sequentially adding the different perturbative corrections described in Tab. 3.1. The top curve represents the error of fragment embedding energy E^0. To this, we first add the dispersion correction E^2_{disp}, which is giving a constant shift along the dihedral angle. Then, the (single) charge transfer correction E^2_{1CT} crucially corrects for the behaviour where the rings are aligned. Finally, we add the double-charge transfer term E^2_{2CT} and triplet-triplet term E^2_{TT} together, recovering the final FragPT2 energy $E^0 + E^2$. These last terms contribute an additional small shift to the aligned rings configuration.

Our data shows that, for each of the molecules and values of ϕ, E^0 recovers at least 93% of the correlation energy (with an average of 97%). While this is high in absolute terms, the shape of the potential energy curves for these models can be qualitatively wrong. As expected, a product state is not a good approximation if the rings are aligned, as the aromatic system will be delocalized over the molecule. This causes the interactions between the fragments to play a more significant role. The product state is on the other hand a good approximation when the rings are perpendicular, there pushing E^0 to 99% of the correlation energy. This causes an imbalance between the two configurations and calls for the need to treat the interactions. When we compute the second-order perturbation energy E^2, it is shown in Figure 3.3 that sometimes E^0 finds a different minimum than the exact state. In these cases especially, the perturbative corrections need to be calculated to give a more correct shape of the potential energy curve. In Figure 3.4, one can see that the division of the perturbation energies can be very constructive in determining the important contributions of the system in question. In case of aromatic dimers, two interactions are important: dispersion and single-charge transfer. While the former takes care of a constant shift over the dihedral angles, the latter is much larger when the aromatic rings are aligned, thus crucial in retrieving the right behaviour of PES. The double-charge transfer and triplet-triplet spin exchange terms are not important in these class of molecules.

3.3.4 Butadiene

Butadiene (C_4H_6) is the final test system that we consider. We define the two fragments by cutting through the middle bond of the molecule. We study the energy of the system while stretch the double bond onto dissociation inside one of the fragments, thereby testing our method to increasing amounts of static correlation inside the fragment. Dissociating the bond additionally causes the leftover molecule to be a radical, thus increasing significantly the strength of the interaction between the fragments.

We define the active spaces by taking the $\pi - \pi^*$ and $\sigma - \sigma^*$ system of the double bonds of both fragments. This results in an active space of four electrons in four orbitals for each fragment.

The potential energy curves are shown in the upper panel of Figure 3.5. It can be clearly seen that the multi-reference product state is a correct description at the equilibrium geometry, but its performance is somewhere in between the Hartree-Fock and the exact solution at dissociation. To improve on it, we clearly need the perturbative corrections.

If we analyze the contributions to the perturbative correction plotted in the lower panel of Figure 3.5, we see that $H'_{1\mathrm{CT}}$ interaction is the most important (contributing around 80 $\sim$ 88% to E^2). Notably, in this system the H'_{TT} contribution is large as well (contributing around 8 $\sim$ 12% to E^2). This is in line with chemical intuition, as this system has low-lying triplet states [224]; a singlet-coupled double triplet excitation may therefore contribute significantly to the ground state wave function. Again, the ability to separately analyze the different classes of inter-fragment interactions is useful here, as it allows to consider the correlation in polyenes in terms of products of local excitations.

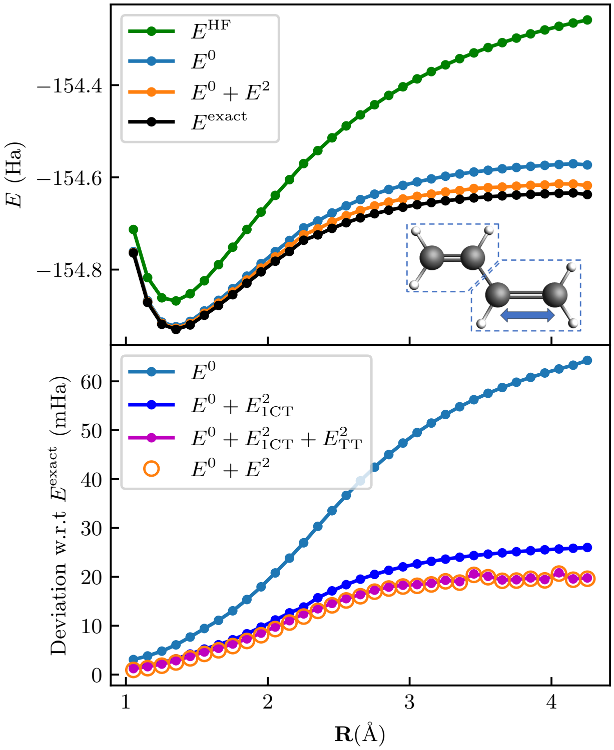


Figure 3.5: Potential energy curves of butadiene. The fragments are chosen by cutting through the middle bond and subsequently stretching the double bond of one of the fragments, as illustrated in the inset. The curves are color coded like in Figure 3.2, and show that both intra-fragment and inter-fragment correlations are important to recover the correct behaviour. In particular, inter-fragment correlations are explained by the vicinity of the two fragments and by the radical that is left over after dissociation of the stretched bond. In the lower panel, we show once more the result of sequentially adding the different perturbative corrections described in Table 3.1. We first add the single-charge transfer contribution $E^2_{1\mathrm{CT}}$ (blue line) and then the triplet-triplet coupling E^2_{TT} (purple line), showing the other contributions are zero by plotting the full FragPT2 energy $E^0 + E^2$ (orange dots).

3.4 Outlook

3.4.1 Computational efficiency

In order to estimate the perturbative corrections in Table 3.1 we have to construct high-order k-RDMs for all $k \leq 5$. These RDMs are tensors with up to 10 indices: constructing and storing them explicitly is computationally expensive. Several methods to evaluate and store high order RDMs in a compressed form have been proposed in the context of e.g. NEVPT2 theory [225, 226]. Resolution of identity (RI) [227], cumulant expansions [228], tensor contraction with integrals [229, 230] are some of the ways to circumvent this bottleneck. Future resesarch should cosider how these methods can be applied in the specific case of FragPT2.

To further improve the efficiency of FragPT2, we can consider modifications to the part our algorithm that calculates the perturbative corrections. For example, to circumvent the need to calculate all the different elements of the RDMs, we can compress the basis of perturbing functions in Eq. (3.14). One option is to set the coefficients C_μ of Eq. (3.12) by the integrals of the perturbation H' under consideration. This gives a single (unnormalized) perturbing function $\left|\Psi^1\right\rangle = H' \left|\Psi^0\right\rangle$, known in literature as a *strongly contracted basis*. This strongly contracted form has been applied with some success in the context of NEVPT2 [36]. The bottleneck of the algorithm then becomes estimating higher order powers of the Hamiltonian and the perturbations on $\left|\Psi^0\right\rangle$, effectively equivalent to the first order of a moment expansion [36, 231]. Another possible approach to reducing the cost of computing perturbations relies on stochastic formulations of MRPT, which have also been studied in the context of strongly-contracted NEVPT2 [232–234]. In these approaches, the necessary quantities were determined in a quantum Monte-Carlo (QMC) framework.

3.4.2 Integration with quantum algorithms

In this manuscript, we focused on solving the single fragments with FCI, but our framework is compatible with any method that can recover RDMs of fragment wavefunctions. Quantum algorithms have emerged as promising tools for tackling classically-hard electronic structure problems, but they come with specific limitations distinct from those of classical algorithms. Fragmentation and embedding techniques are critical for defining tasks suited to quantum algorithms, enabling a focus on strongly-correlated active sites while reducing problem sizes. Recent studies have explored integrating quantum algorithms into embedding schemes, including SAPT (for both near-term variational [235, 236] and fault-tolerant [237] approaches) and LASSCF [214, 238]. In this context, we discuss integrating FragPT2 with the variational quantum eigensolver (VQE) [54, 91].

The VQE prescribes to prepare on a quantum device an ansatz state $|\Psi(\boldsymbol{\theta})\rangle$, as a function of a set of classical parameters $\boldsymbol{\theta}$ which are then optimized to minimize the state energy $E(\boldsymbol{\theta}) = \langle\Psi(\boldsymbol{\theta})| H |\Psi(\boldsymbol{\theta})\rangle$. Having access to a quantum device allows to produce states which can be hard to represent on a classical computer, enabling the implementation of ansätze such as unitary coupled cluster [54, 239] and other heuristic constructions [64, 92, 240]; however, sampling the energy and other properties from the quantum state incurs a large sampling cost, which is worsened by the required optimization overhead.

Integrating FragPT2 with the VQE is straightforward. For each fragment X, a separate parameterized wavefunction $|\Psi_X(\boldsymbol{\theta}_X)\rangle$ is represented, reconstructing an ansatz $|\Psi_A(\boldsymbol{\theta}_A)\rangle|\Psi_A(\boldsymbol{\theta}_B)\rangle$ for the product state Eq. (3.2). As no quantum correlation is needed, multiple wavefunctions can be prepared in parallel in separate quantum devices, or even serially on the same device; this can allow to treat larger chemical systems with limited-size quantum devices. We can find the lowest-energy product state directly by minimizing the expectation value of the Hamiltonian

$$E(\boldsymbol{\theta}_A, \boldsymbol{\theta}_B) = \langle\Psi_A(\boldsymbol{\theta}_A)|\langle\Psi_B(\boldsymbol{\theta}_B)|H|\Psi_A(\boldsymbol{\theta}_A)\rangle|\Psi_B(\boldsymbol{\theta}_B)\rangle . \tag{3.19}$$

this energy can be estimated by measuring the one- and two-body reduced density matrices separately on each fragment. As shown in Section 3.2.2, the minimum energy product state matches the solution of the mean field embedding.

Integrating VQE with fragmentation techniques can help describe binding energies, proposed in literature with a method based on symmetry adapted perturbation theory and termed SAPT(VQE) [235, 236]. SAPT(VQE) addresses the same terms as Algorithm 3.1, but uses a perturbative expansion instead of mean-field coupling for terms dependent on fragment 1- and 2-RDMs. It employs two non-orthogonal orbital sets for the fragments, limiting this method to non-covalently bonded fragments. Inspired by SAPT(VQE), Algorithm 3.1 could be adapted to measure interaction energies. A thorough comparison of the two methods and studying their dependence on molecular orbitals and atomic basis set is a promising area for future work. SAPT has also recently been applied to fault-tolerant quantum algorithms, overcoming some of the limitations of near-term devices [237].

As per Algorithm 3.2, to recover the perturbative corrections accounting for inter-fragment interactions we need to extract higher-order reduced density matrices from each fragment's wavefunction. Perturbation theory for the variational quantum eigensolver has been studied in the context of recovering dynamical correlations [241, 242]. Using measurement optimization techniques from [61, 117], estimating all the elements of the k-RDMs on a fragment active space of N orbitals to a precision ϵ requires $O(\epsilon^{-2}N^k)$ samples. In practice this makes naively estimating the perturbative corrections very costly, especially for the single-charge transfer terms $H'_{1\mathrm{CT}}$ that require 5-RDMs (see Table 3.1). An interesting direction for future research might consider using shadow tomography and its fermionic extension [243, 244] to estimate RDMs to all orders at the same time.

3.4.3 Further extensions

Extension to multiple fragments — This chapter focused on the case of two active fragments. However, it is relatively straightforwardly applied to more. The lowest energy product state can be retrieved by trivially extending the algorithm, looping through the fragments and solving exactly the active fragment feeling the mean-field of the inactive fragments, until reaching convergence. Subsequently, we can treat the inter-fragment interactions that can span four fragments at a time at most (as the Hamiltonian is a two-body operator), which is a coupled charge transfer excitation. While the perturbing functions then have to be extended to these types of excitations, the matrix elements that one has to estimate will factorize in the same way, and the algorithm will not be more costly than for two fragments (i.e. no higher order RDMs will have to be estimated). Working out the exact expressions and implementing a truly many-fragment algorithm is part of future work.

Localized orbitals beyond Hartree-Fock — The Hartree-Fock determinant is known to be an unstable reference in dissociating systems and other highly-correlated molecules [21, 245]. To generate the input orbitals, one might want to change from a cheap mean-field method to a slightly more expensive CASSCF calculation with a small active space. As the localization scheme can handle any input orbitals, our method can be trivially adapted to a better choice of reference orbitals that already takes into account some correlation. Additionally, one can include intra-fragment orbital-optimization during the fragment embedding (Algorithm 3.1). A simple approach would involve using a CASSCF solver on the individual fragments, with orbital rotations constrained to each fragment to keep the fragments separated. This could enhance the method's accuracy and provide a better starting point for perturbation theory. In this spirit, a version LASSCF [149] or vLASSCF [150] could be recovered as an extension of our method where orbital rotations between fragment active spaces are also allowed and optimized self-consistently.

NEVPT2-like perturbations — So far we have treated the interactions only inside the complete active space, i.e. our H from Eq. (3.1) involves indices within either active fragment. To retrieve more of the dynamical correlation energy, the core idea of NEVPT2 is to include excitations involving also the inactive orbitals in a perturbative way. We can build on top of our previous approach by including additional perturbations and perturbing functions. Correspondingly, we can augment our Hamiltonian from Eq. (3.1) as,

$$H = H^0 + H^{'}_{act} + H^{'}_{inact} \tag{3.20}$$

where $H^{'}_{inact}$ consists of the various classes of perturbations involving excitations from the core to the active space, the active space to the virtual space and the core to the active space. For the form of these perturbations, see reference 36. It is straightforward to extend the methods from NEVPT2 to the case of multiple active fragments, and again the matrix elements will factorize on different fragments in the same way, relieving the need to estimate additional matrix elements on the multi-reference fragment solvers.

3.5 Conclusion

In this work, we introduced a novel multi-reference multi-fragment embedding framework called FragPT2. We showed that our method gives accurate results for a reduced cost in active space size, especially when the fragments are well-separated. Our comprehensive numerical benchmarks on a variety of molecules show that: 1. Intra-fragment static correlation can be retrieved by an MC product state ansatz (E^0) 2. Inter-fragment correlation can be treated as a perturbative correction (E^2) 3. A combination of these is needed to recover the correct shape of the potential energy curve. Using a decomposition of the Hamiltonian based on fragment symmetries, we can distinguish the contributions to the inter-fragment correlation in E^2, providing insight into important interactions within the studied systems. Furthermore, our adapted localization scheme allows to define molecular fragments that cut through covalent bonds. In this case, perturbative corrections describing inter-fragment charge transfer (and, to a lesser extent, triplet-triplet spin exchange) are crucial for accurately describing the system. Future research directions include improving the efficiency of high-order RDM estimation, integrating FragPT2 with variational quantum algorithms, and extensions to multiple fragments for broader applicability.

Our multi-reference embedding scheme could find broad applications, for instance in understanding the spatial dependence of the correlation energy in π-stacked systems and other biochemically important systems [246], modelling supramolecular complex formation [247] in metal ion separation, or in analyzing metallophylic interactions [248–250].

3.A Hamiltonian decomposition

In this appendix, we give a detailed derivation of the terms in the decomposition Eq. (3.11) of the full active-space molecular Hamiltonian.

We start from the full Hamiltonian Eq. (3.1), and we rewrite the quartic excitation operators $e_{pqrs} = E_{pq}E_{rs} - \delta_{qr}E_{ps}$ in terms of the quadratic E_{pq}, obtaining

$$H = \sum_{pq} h_{pq}E_{pq} - \frac{1}{2}\sum_{pq}\sum_{r} g_{prrq}E_{pq} + \frac{1}{2}\sum_{pqrs} g_{pqrs}E_{pq}E_{rs}. \tag{3.21}$$

We will separately deal with the terms that conserve charge on each fragment (in Appendix 3.A.1) and those that transfer charge between fragments (in Appendix 3.A.2). It can be easily identified whether a term preserves charge on each fragment by counting the number of electrons moved across orbitals, as all orbitals $\{p, q, r, s\}$ pertain to either fragment A or B.

3.A.1 Charge-conserving terms

In this section we derive the charge-conserving inter-fragment terms H'_{disp} and H'_{TT}, as well as the on-fragment Hamiltonians H_A and H_B.

The one-body term of Eq. (3.21) only conserves charge if p and q are both in the A fragment or both in the B fragment, these terms will respectively be part of H_A and H_B. The two-body term of Eq. (3.21) includes terms where all p, q, r, s are part of the same fragment: these will also be part of H_A and H_B.

The other possible options that preserve charge while including terms on both fragments are:

$$
\begin{aligned}
&\frac{1}{2}\sum_{pq\in A}\sum_{rs\in B} g_{pqrs}E_{pq}E_{rs} + \frac{1}{2}\sum_{pq\in B}\sum_{rs\in A} g_{pqrs}E_{pq}E_{rs} \\
+&\frac{1}{2}\sum_{pq\in A}\sum_{rs\in B} g_{psrq}E_{ps}E_{rq} + \frac{1}{2}\sum_{rs\in B}\sum_{pq\in A} g_{rqps}E_{rq}E_{ps}.
\end{aligned} \tag{3.22}
$$

It is straightforward to show that the first two terms are equivalent by using the symmetry $g_{pqrs} = g_{rspq}$ and the excitation operator commutation relations $[E_{pq}, E_{rs}] = E_{ps}\delta_{qr} - E_{rq}\delta_{ps}$. These terms represent the Coulomb interactions between the fragments. The latter two terms describe the exchange interactions between the fragments.

We rewrite the exchange term using Fermionic commutation rules – using the notation p_X being an orbital index on fragment X we get

$$
\begin{aligned}
E_{p_A s_B}E_{r_B q_A} =& \sum_{\sigma\tau} a^\dagger_{p_A\sigma}a_{s_B\sigma}a^\dagger_{r_B\tau}a_{q_A\tau} \\
=& \sum_{\sigma\tau}\left(\delta_{r_B s_B}\delta_{\sigma\tau}a^\dagger_{p_A\sigma}a_{q_A\tau} - a^\dagger_{p_A\sigma}a_{q_A\tau}a^\dagger_{r_B\tau}a_{s_B\sigma}\right) \\
=& \delta_{r_B s_B}E_{p_A q_A} \\
&- a^\dagger_{p_A\alpha}a_{q_A\alpha}a^\dagger_{r_B\alpha}a_{s_B\alpha} - a^\dagger_{p_A\beta}a_{q_A\beta}a^\dagger_{r_B\beta}a_{s_B\beta} \\
&- a^\dagger_{p_A\alpha}a_{q_A\beta}a^\dagger_{r_B\beta}a_{s_B\alpha} - a^\dagger_{p_A\beta}a_{q_A\alpha}a^\dagger_{r_B\alpha}a_{s_B\beta} \\
=& \delta_{r_B s_B}E_{p_A q_A} - S^{0,0}_{p_A q_A}S^{0,0}_{r_B s_B} \\
&- T^{1,0}_{p_A q_A}T^{1,0}_{r_B s_B} + T^{1,1}_{p_A q_A}T^{1,-1}_{r_B s_B} + T^{1,-1}_{p_A q_A}T^{1,1}_{r_B s_B}
\end{aligned} \tag{3.23}
$$

where we use the definition of the spin operators [21]:

$$S^{(0,0)}_{p_X q_X} = \frac{1}{\sqrt{2}}E_{p_X q_X} = \frac{1}{\sqrt{2}}(a^\dagger_{p_X\alpha}a_{q_X\alpha} + a^\dagger_{p_X\beta}a_{q_X\beta}) \tag{3.24}$$

$$T^{(1,0)}_{p_X q_X} = \frac{1}{\sqrt{2}}(a^\dagger_{p_X\alpha}a_{q_X\alpha} - a^\dagger_{p_X\beta}a_{q_X\beta}) \tag{3.25}$$

$$T^{(1,1)}_{p_X q_X} = -a^\dagger_{p_X\alpha}a_{q_X\beta} \tag{3.26}$$

$$T^{(1,-1)}_{p_X q_X} = a^\dagger_{p_X\beta} a_{q_X\alpha} \tag{3.27}$$

and

$$\begin{aligned} a^\dagger_{p\alpha} a_{q\alpha} &= \frac{1}{\sqrt{2}}(S^{(0,0)}_{pq} + T^{(1,0)}_{pq}) \\ a^\dagger_{p\beta} a_{q\beta} &= \frac{1}{\sqrt{2}}(S^{(0,0)}_{pq} - T^{(1,0)}_{pq}). \end{aligned} \tag{3.28}$$

Notice that the last three terms of Eq. (3.23) conserve the total spin of the system, but flip the local spin of the individual fragments. Separating out these terms from the expansion of Eq. (3.22) we obtain the triplet-triplet interaction Hamiltonian:

$$H'_{\mathrm{TT}} = \sum_{pq\in A}\sum_{rs\in B} g_{psrq} T^{1,0}_{p_A q_A} T^{1,0}_{r_B s_B} + T^{1,1}_{p_A q_A} T^{1,-1}_{r_B s_B} + T^{1,-1}_{p_A q_A} T^{1,1}_{r_B s_B} \tag{3.29}$$

$$= -\sum_{pq\in A}\sum_{rs\in B} g_{psrq} t_{pq,rs} \tag{3.30}$$

where $t_{pq,rs} = T^{1,0}_{pq} T^{1,0}_{rs} - T^{1,1}_{pq} T^{1,-1}_{rs} - T^{1,-1}_{pq} T^{1,1}_{rs}$.

After substracting this term from Eq. (3.22), we arrive at the following expression for the Hamiltonian that includes all fragment charge-conserving and spin-conserving terms: $H_A + H_B + H_{AB}$, which can be easily split in the three terms

$$H_A = \sum_{pq\in A} h_{pq} E_{pq} - \frac{1}{2}\sum_{pqr\in A} g_{prrq} E_{pq} + \sum_{pqrs\in A} g_{pqrs} E_{pq} E_{rs} \tag{3.31}$$

$$H_B = \sum_{pq\in B} h_{pq} E_{pq} - \frac{1}{2}\sum_{pqr\in B} g_{prrq} E_{pq} + \sum_{pqrs\in B} g_{pqrs} E_{pq} E_{rs} \tag{3.32}$$

$$H_{\mathrm{mf}} + H'_{\mathrm{disp}} = \sum_{pq\in A}\sum_{rs\in B} \left[g_{pqrs} - \frac{1}{2} g_{psrq} \right] E_{pq} E_{rs}, \tag{3.33}$$

where the last row can be further split in a mean-field interaction term H_{mf} and a dispersion term H'_{disp}: the mean-field interaction is defined self-consistently on the basis of the solution of the Hamitlonian $H^0 = H_A + H_B + H_{\mathrm{mf}}$, as we showed in Section 3.2.2.

3.A.2 Charge transfer terms

We now work on separating the terms that involve charge transfers between the fragments. As the molecular Hamiltonian contains only one-body and two-body terms, we only need to consider single-charge and double-charge transfers, respectively classified as part of H'_{1CT} and H'_{2CT}.

We first isolate the single-charge transfer terms. These include the single-body terms of Eq. (3.21) where p and q pertain to different fragments:

$$\sum_{p\in A}\sum_{q\in B} h_{pq}E_{pq} + \sum_{p\in B}\sum_{q\in A} h_{pq}E_{pq}; \tag{3.34}$$

along with the two-body terms where three indices pertain to a fragment and one pertains to the other:

$$\begin{aligned}&\frac{1}{2}\left[\sum_{pqr\in A}\sum_{s\in B} + \sum_{pqs\in A}\sum_{r\in B} + \sum_{prs\in A}\sum_{q\in B} + \sum_{qrs\in A}\sum_{p\in B}\right] g_{pqrs}e_{pqrs}\\ +&\frac{1}{2}\left[\sum_{pqr\in B}\sum_{s\in A} + \sum_{pqs\in B}\sum_{r\in A} + \sum_{prs\in B}\sum_{q\in A} + \sum_{qrs\in B}\sum_{p\in A}\right] g_{pqrs}e_{pqrs},\end{aligned} \tag{3.35}$$

where for brevity we write multiple sums (in brackets) sharing the same summand (in parentheses). We can simplify this using the symmetries of the two-body integral $g_{pqrs} = g_{rspq} = g_{qprs} = g_{pqsr}$ and the commutation relations of excitation operators $[E_{pq}, E_{rs}] = E_{ps}\delta_{qr} - E_{rq}\delta_{ps}$, obtaining

$$\begin{aligned}&-\sum_{p\in A}\sum_{q\in B}\left[\sum_{r\in A} g_{prrq}\right]E_{pq} - \sum_{p\in B}\sum_{q\in A}\left(\sum_{r\in B} g_{prrq}\right)E_{pq}\\ &+\sum_{pqr\in A}\sum_{s\in B} g_{pqrs}E_{pq}\left(E_{rs}+E_{sr}\right) + \sum_{pqr\in B}\sum_{s\in A} g_{pqrs}E_{pq}\left(E_{rs}+E_{sr}\right).\end{aligned} \tag{3.36}$$

Combining this with the one-body term Eq. (3.34) we define the single-charge transfer term

$$\begin{aligned}H'_{1\mathrm{CT}} =& \sum_{p\in A}\sum_{q\in B}\left[h_{pq} - \sum_{r\in A} g_{prrq}\right]E_{pq} + \sum_{p\in B}\sum_{q\in A}\left[h_{pq} - \sum_{r\in B} g_{prrq}\right]E_{pq}\\ &+\sum_{pqr\in A}\sum_{s\in B} g_{pqrs}E_{pq}\left[E_{rs}+E_{sr}\right] + \sum_{pqr\in B}\sum_{s\in A} g_{pqrs}E_{pq}\left[E_{rs}+E_{sr}\right].\end{aligned} \tag{3.37}$$

The double-charge transfer is simpler, as there are just two two-body terms that allow for a double-charge transfer:

$$H'_{2\mathrm{CT}} = \frac{1}{2}\sum_{pr\in A}\sum_{qs\in B} g_{pqrs}E_{pq}E_{rs} + \frac{1}{2}\sum_{pr\in B}\sum_{qs\in A} g_{pqrs}E_{pq}E_{rs}. \tag{3.38}$$

One can verify that $H = H_A + H_B + H_{\mathrm{mf}} + H'_{\mathrm{disp}} + H'_{\mathrm{TT}} + H'_{1\mathrm{CT}} + H'_{2\mathrm{CT}}$.

3.B Fragment matrix elements

This section contains derivations of the expressions for the zeroth-order Hamiltonian matrix elements $\langle\Psi_{\mu\nu}|H^0|\Psi_{\kappa\lambda}\rangle$ for every perturbation $H' = \sum_{\mu\nu} g_{\mu\nu} O^A_\mu O^B_\nu$. We will focus here on estimating these matrix elements exactly without any approximation, resulting in the need for higher order RDMs. For a discussion of future work to improve efficiency, see Section 3.4.1. In general, the expressions are given as:

$$\begin{aligned}\left\langle\Psi_{\mu\nu}\middle|H^0\middle|\Psi_{\kappa\lambda}\right\rangle =& \left\langle\Psi_A\middle|O^{A\dagger}_\mu H^{\text{eff}}_A O^A_\kappa\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O^{B\dagger}_\nu O^B_\lambda\middle|\Psi_B\right\rangle\\ &+\left\langle\Psi_A\middle|O^{A\dagger}_\mu O^A_\kappa\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O^{B\dagger}_\nu H^{\text{eff}}_B O^B_\lambda\middle|\Psi_B\right\rangle\\ &+E_{\text{mf}}\left\langle\Psi_A\middle|O^{A\dagger}_\mu O^A_\kappa\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O^{B\dagger}_\nu O^B_\lambda\middle|\Psi_B\right\rangle,\end{aligned} \tag{3.39}$$

$$\left\langle\Psi^0\middle|H'\middle|\Psi_{\kappa\lambda}\right\rangle = \sum_{\mu\nu} g_{\mu\nu}\left\langle\Psi_A\middle|O^{A\dagger}_\mu O^A_\kappa\middle|\Psi_A\right\rangle\left\langle\Psi_B\middle|O^{B\dagger}_\nu O^B_\lambda\middle|\Psi_B\right\rangle, \tag{3.40}$$

where we defined the perturbing functions as $|\Psi_{\mu\nu}\rangle = O^A_\mu O^B_\nu |\Psi_A\rangle|\Psi_B\rangle$. We have:

$$\begin{aligned}&\left\langle\Psi_X\middle|O^X_\mu H^{\text{eff}}_X O^X_\nu\middle|\Psi_X\right\rangle =\\ &\sum_{pq\in X}\left(h_{pq}+\sum_{rs\in Y} g'_{pqrs}\gamma^Y_{rs}-\sum_{r\in X} g_{prrq}\right)\left\langle\Psi_X\middle|O^X_\mu E_{pq} O^X_\nu\middle|\Psi_X\right\rangle\\ &+\sum_{pqrs\in X} g_{pqrs}\left\langle\Psi_X\middle|O^X_\mu E_{pq}E_{rs}O^X_\nu\middle|\Psi_X\right\rangle\end{aligned} \tag{3.41}$$

Thus, the relevant operator matrix elements to estimate are $\left\langle\Psi_X\middle|O^X_\mu E_{pq}E_{rs}O^X_\nu\middle|\Psi_X\right\rangle$, $\left\langle\Psi_X\middle|O^X_\mu E_{pq}O^X_\nu\middle|\Psi_X\right\rangle$ and $\left\langle\Psi_X\middle|O^X_\mu O^X_\nu\middle|\Psi_X\right\rangle$.

3.B.1 Dispersion

For the case of the dispersion perturbation, we use the perturbing functions

$$E_{tu}E_{vw}\left|\Psi^0\right\rangle\left[tu\in A,\, vw\in B\right]. \tag{3.42}$$

Thus, we straightforwardly identify $O^X_\mu \rightarrow E_{pq}$ and the most expensive matrix element to compute is:

$$\langle\Psi_X|E_{vw}E_{pq}E_{rs}E_{tu}|\Psi_X\rangle, \tag{3.43}$$

i.e. a 4-particle reduced density matrix.

3.B.2 Single-charge transfer

The single-charge transfer case is a bit more complicated. We like to preserve the total spin, so we use the spin-free excitation operators to define the perturbing functions as:

$$E_{tu}E_{vw}\left|\Psi^0\right\rangle \quad \begin{bmatrix} tuv \in A, w \in B \\ tuw \in A, v \in B \\ tuv \in B, w \in A \\ tuw \in B, v \in A \end{bmatrix}. \tag{3.44}$$

That means a straightforward decomposition into $O_\mu^A O_\nu^B$ is more intricate because of the sum over spin. Instead, we have $\sum_\sigma E_{p_X q_X} a^\dagger_{v_X\sigma} a_{w_Y\sigma}$ and $\sum_\sigma E_{p_Y q_Y} a^\dagger_{v_X\sigma} a_{w_Y\sigma}$. Let us work out the matrix elements explicitly for the first case in Eq. (3.44). This is easily generalizable to the other cases. The total matrix element becomes:

$$\begin{aligned} &\left\langle \Psi_{k_A l_A m_A n_B} \middle| H^0 \middle| \Psi_{t_A u_A v_A w_B} \right\rangle = \\ &\sum_{\sigma\tau} \left\langle \Psi_A \middle| a_{m\sigma} E_{lk} H_A^{\text{eff}} E_{tu} a^\dagger_{v\tau} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} a_{w\tau} \middle| \Psi_B \right\rangle \\ &+ \sum_{\sigma\tau} \left\langle \Psi_A \middle| a_{m\sigma} E_{lk} E_{tu} a^\dagger_{v\tau} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} H_B^{\text{eff}} a_{w\tau} \middle| \Psi_B \right\rangle \\ &+ E_{\text{mf}} \sum_{\sigma\tau} \left\langle \Psi_A \middle| a_{m\sigma} E_{lk} E_{tu} a^\dagger_{v\tau} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} a_{w\tau} \middle| \Psi_B \right\rangle \end{aligned} \tag{3.45}$$

As E_{pq} and H_X^{eff} preserve spin, and $|\Psi_X\rangle$ are eigenfunctions of S^2, we can replace the double sum over spin by a single one as $\sigma = \tau$. Now substituting $O_\mu^X \to E_{tu} a^\dagger_{v\sigma}$ in Eq. (3.41) the most expensive object to estimate will be:

$$\left\langle \Psi_X \middle| a_{m\sigma} E_{lk} E_{pq} E_{rs} E_{tu} a^\dagger_{v\sigma} \middle| \Psi_X \right\rangle, \tag{3.46}$$

i.e. a 5-particle reduced density matrix.

3.B.3 Double-charge transfer

For the double-charge transfer we have the following set of perturbing functions:

$$E_{tu}E_{vw}\left|\Psi^0\right\rangle \begin{bmatrix} tv \in A, uw \in B \\ uw \in A, tv \in B \end{bmatrix}. \tag{3.47}$$

This makes the total matrix element equal to:

$$\begin{aligned}
&\left\langle \Psi_{k_A l_B m_A n_B} \middle| H^0 \middle| \Psi_{t_A u_B v_A w_B} \right\rangle = \\
&\sum_{\sigma\tau\kappa\lambda} \left\langle \Psi_A \middle| a_{m\sigma} a_{k\tau} H_A^{\text{eff}} a^\dagger_{t\kappa} a^\dagger_{v\lambda} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} a^\dagger_{l\tau} a_{u\kappa} a_{w\lambda} \middle| \Psi_B \right\rangle \\
+ &\sum_{\sigma\tau\kappa\lambda} \left\langle \Psi_A \middle| a_{m\sigma} a_{k\tau} a^\dagger_{t\kappa} a^\dagger_{v\lambda} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} a^\dagger_{l\tau} H_B^{\text{eff}} a_{u\kappa} a_{w\lambda} \middle| \Psi_B \right\rangle \\
+E_{\text{mf}} &\sum_{\sigma\tau\kappa\lambda} \left\langle \Psi_A \middle| a_{m\sigma} a_{k\tau} a^\dagger_{t\kappa} a^\dagger_{v\lambda} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| a^\dagger_{n\sigma} a^\dagger_{l\tau} a_{u\kappa} a_{w\lambda} \middle| \Psi_B \right\rangle .
\end{aligned} \tag{3.48}$$

Here there also simplifications possible regarding the sum over spin. Namely, only the following options are non-zero:
$\sigma, \tau, \kappa, \lambda \in \{\alpha\alpha\alpha\alpha, \beta\beta\beta\beta, \alpha\beta\alpha\beta, \alpha\beta\beta\alpha, \beta\alpha\alpha\beta, \beta\alpha\beta\alpha\}$. Regardless, making the identification $O^X_\mu \to a^\dagger_{t\sigma} a^\dagger_{v\sigma}$ in Eq. (3.41), the most expensive object to estimate is:

$$\left\langle \Psi_X \middle| a_{m\sigma} a_{k\tau} E_{pq} E_{rs} a^\dagger_{t\kappa} a^\dagger_{v\lambda} \middle| \Psi_X \right\rangle , \tag{3.49}$$

and similarly for $O^X_\mu \to a_{t\sigma} a_{v\sigma}$, this is as expensive as estimating a 4-particle reduced density matrix.

3.B.4 Triplet-triplet

The triplet-triplet perturbing functions are given by:

$$t_{tuvw} \left| \Psi^0 \right\rangle \left[tu \in A,\, vw \in B \right] , \tag{3.50}$$

where $t_{tuvw} = T^{1,0}_{tu} T^{1,0}_{vw} - T^{1,1}_{tu} T^{1,-1}_{vw} - T^{1,-1}_{tu} T^{1,1}_{vw}$ (see Appendix 3.A.1 for their definition). Observe that, for any fragment operator O^X that preserves spin, all matrix elements $\left\langle \Psi_X \middle| T^{(1,m')} O^X T^{(1,m')} \middle| \Psi_X \right\rangle$ are *only* non-zero if $m + m' = 0$, where $m, m' \in \{-1, 0, -1\}$. Thus, we can make the following statement:

$$\begin{aligned}
&\langle \Psi^0_{klmn} | O^A O^B | \Psi^0_{tuvw} \rangle \\
&= \langle \Psi^0 | t^\dagger_{k_A l_A, m_B n_B} O^A O^B t_{t_A u_A, v_B w_B} | \Psi^0 \rangle \\
&= \langle \Psi_A | T^{(1,0)}_{lk} O^A T^{(1,0)}_{tu} | \Psi_A \rangle \, \langle \Psi_B | T^{(1,0)}_{nm} O^B T^{(1,0)}_{vw} | \Psi_B \rangle \\
&+ \langle \Psi_A | T^{(1,1)}_{lk} O^A T^{(1,-1)}_{tu} | \Psi_A \rangle \, \langle \Psi_B | T^{(1,-1)}_{nm} O^B T^{(1,1)}_{vw} | \Psi_B \rangle \\
&+ \langle \Psi_A | T^{(1,-1)}_{lk} O^A T^{(1,1)}_{tu} | \Psi_A \rangle \, \langle \Psi_B | T^{(1,1)}_{nm} O^B T^{(1,-1)}_{vw} | \Psi_B \rangle
\end{aligned} \tag{3.51}$$

Finally, the matrix element of H^0 in this basis is thus equal to:

$$\begin{aligned} &\left\langle \Psi_{k_A l_A m_B n_B} \middle| H^0 \middle| \Psi_{t_A u_A v_B w_B} \right\rangle = \\ &\sum_{m+m'=0} \left\langle \Psi_A \middle| T_{lk}^{(1,m)} H_A^{\mathrm{eff}} T_{tu}^{(1,m')} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| T_{nm}^{(1,m')} T_{vw}^{(1,m)} \middle| \Psi_B \right\rangle \\ &+ \sum_{m+m'=0} \left\langle \Psi_A \middle| T_{lk}^{(1,m)} T_{tu}^{(1,m')} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| T_{nm}^{(1,m')} H_A^{\mathrm{eff}} T_{vw}^{(1,m)} \middle| \Psi_B \right\rangle \\ &+ E_{\mathrm{mf}} \sum_{m+m'=0} \left\langle \Psi_A \middle| T_{lk}^{(1,m)} T_{tu}^{(1,m')} \middle| \Psi_A \right\rangle \left\langle \Psi_B \middle| T_{nm}^{(1,m')} T_{vw}^{(1,m)} \middle| \Psi_B \right\rangle . \end{aligned} \tag{3.52}$$

The most expensive object to estimate in the triplet-triplet case is then, identifying $O_\mu^X \to T_{tu}^{(1,m)}$ in Eq. (3.41):

$$\left\langle \Psi_X \middle| T_{lk}^{(1,m)} E_{pq} E_{rs} T_{tu}^{(1,m')} \middle| \Psi_X \right\rangle , \tag{3.53}$$

where $m + m' = 0$. This is equivalent in cost to measuring a 4-particle reduced density matrix.

CHAPTER 4

Analytical non-adiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver

4.1 Introduction

Many fundamental processes in nature, such as photosynthesis and vision, are triggered by light absorption. Thus, a proper description of the associated primary light-induced photochemical events requires a quantum-mechanical approach able to treat accurately both the ground and the excited electronic states. Although density functional theory (DFT) and its time-dependent extension to excited states (TDDFT) have seen huge progress in treating molecular and condensed matter systems near equilibrium [251–253], these approaches are not adequate to accurately describe photochemical reaction paths where the Born-Oppenheimer approximation breaks down for several strongly coupled electronic states that get very close in energy [254]. Especially couplings between the first excited and ground states are problematic because of the single-reference character of many quantum chemical methods (for instance the popular time-dependent density functional theory, TDDFT approach). While single-reference approaches with spin-flip excitations might help to overcome some of these limitations [255], in general more accurate and computationally demanding multi-configurational wavefunction approaches are required for modeling these intrinsically non-adiabatic cases. A good example is the description of the prototypical photoisomerization process in the retinal chromophore of rhodopsin, one of the most studied events in photobiology [256]. Schematically, after the initial photoexcitation, this event proceeds via the relaxation in the first excited state (S_1) towards a conical intersection (CoIn) region. Here, the population is transferred back to the ground state (S_0) where the isomerization is completed. In order to describe dynamically this type of event, one needs the knowledge of the potential energy surfaces (PES) for the electronic states involved in the process, typically S_0 and S_1. Moreover, one should also efficiently compute the gradient of the PES with respect to the nuclear displacement, which in a semiclassical non-adiabatic molecular dynamics scheme provides the forces driving the nuclear subsystem [254, 256, 257]. Finally, it is also crucial to estimate the non-adiabatic coupling terms between the two electronic states, which eventually determine the conical topography of the crossing between the two PES and the dynamical coupling that results in population transfer between the two states [258, 259]. The challenge in computational quantum chemistry is to obtain all these necessary ingredients at an affordable numerical cost and yet with good accuracy.

Methods that are able to provide both non-adiabatic couplings and a correct description of the PES topology and topography (double cone of dimension two with respect to variations of the nuclear coordinates) of conical intersections require, formally, that the problem be solved at the very end with a final Hamiltonian diagonalization. When the crossing occurs between the first excited and ground states, this implies a democratic treatment of both wavefunctions within a common Slater determinant basis set, which in practice calls for a state-averaged (SA) orbital optimization. This can be achieved in-principle by the state-averaged multiconfigurational self-consistent field (SA-MCSCF) method [21]. In practice, the diagonalization step is the principal bottleneck and one has to consider small complete active spaces (CAS), thus leading to the state-averaged complete active space self-consistent field (SA-CASSCF) method [21]. However, this decrease in complexity comes at the expense of a missing dynamical correlation treatment, that is usually recovered by multireference quasidegenerate perturbation techniques, as in the XMS-CASPT2 [260], XMCQDPT2 [261], or QD-NEVPT2 methods [262, 263]; see also Ref. 264 for a comparative discussion of the correct treatment of degeneracies with a selection of excited-state approaches.

With the advent of quantum computing, the dream of a very large CAS becomes possible again, thus turning small SA-CASSCF into large SA-CASSCF which should be good enough to account for a qualitatively correct description of the wave function and also include a substantial part of the (previously missing) so-called dynamical correlation. Note that even with relatively small active spaces, the dynamical correlation can be retrieved a posteriori by other techniques on quantum computers, with no additional qubits or circuit depth, but at the expense of more measurements, as described by Takeshita *et al.* [265]. Recently, the *quantum analogue* of SA-CASSCF has been introduced by Yalouz *et al.* [30] based on a state-averaged orbital-optimized (SA-OO) extension of the variational quantum eigensolver (VQE) algorithm [54, 91], thus referred to as the SA-OO-VQE algorithm. While SA-OO-VQE has been shown to provide an accurate and democratic description of both the ground and first-excited PES [30], its extension to excited-state quantum dynamics requires the knowledge of energy gradients and non-adiabatic couplings. In this work, we show how these properties can be analytically estimated on a quantum computer within the SA-OO-VQE framework, following the coupled-perturbed equations [266–272]. In analogy with Ref. 30, the performance of our algorithm is illustrated on the minimal Schiff base model (*i.e.* the formaldimine molecule), for which results are indistinguishable from its classical analogue, the (coupled-perturbed) SA-CASSCF method.

The chapter is organized as follows: in Section 4.2.1, we briefly introduce quantum chemistry for excited states for pedagogical purposes, from the Born-Oppenheimer approximation to the SA-MCSCF method. Turning to quantum computing in Section 4.2.2, a summary of the SA-OO-VQE is given, and a way to extract the eigenstates (*i.e.* the adiabatic states) is provided, where we also discuss the alternative choice of having diabatic or adiabatic states within the SA-OO-VQE algorithm. Furthermore, the analytical estimation of energy gradients and non-adiabatic couplings is described, and they are compared with classical methods. Using the equations for the analytical gradients, a geometry optimization to the degeneracy point is executed in Sec. 4.4.3 as a simple illustration. A more involved optimization to the minimal energy crossing point (MECI) that requires the knowledge of non-adiabatic couplings is performed in Sec. 4.4.4. Conclusions and perspectives are finally discussed in Sec. 4.5.

4.2 Theory

4.2.1 Quantum Chemistry for excited states

Born–Oppenheimer and the adiabatic approximation

One of the most fundamental approximations used in theoretical chemistry is the adiabatic approximation between electrons and nuclei, which most often takes the form of the Born–Oppenheimer approximation and sometimes of the Born–Huang approximation (the latter being essentially used for highly-accurate treatments of vibrations in small molecules). In both cases, non-adiabatic couplings due to the action of the kinetic energy operator of the nuclei on the parametric dependence of the adiabatic electronic wavefunctions are neglected; however, the Born–Huang approximation considers nuclear-mass-dependent diagonal corrections that are to be added to the potential energy surface obtained as a single adiabatic eigenvalue of the clamped-nucleus Hamiltonian.

Such approximations are justified by the small ratio of electronic over nuclear masses, which results in very different energy and time scales in the vast majority of cases. However, electronic degeneracies may occur at certain nuclear geometries (Jahn–Teller crossings due to symmetry, or, more generally, conical intersections). At such points, the two intersecting potential energy surfaces take locally the shape of a double cone (over a subspace of dimension 2 for a two-state crossing). The two nuclear displacements that lift degeneracy to first order are usually called branching-space vectors; their directions can be identified to the energy gradient difference and first-order non-adiabatic coupling (NAC) vector. Formally, the x-component of the NAC vector between two electronic states $|\Psi_I\rangle$ and $|\Psi_J\rangle$ is defined by

$$D_{IJ} = \left\langle \Psi_I \middle| \frac{\partial}{\partial x} \Psi_J \right\rangle . \tag{4.1}$$

where x represents a given nuclear coordinate and the wavefunctions depend parametrically on it (integration, however, is performed over the electronic Hilbert space only). If the wavefunctions considered in Eq. (4.1) are exact, they yield

$$D_{IJ} = \frac{1}{E_J - E_I} \langle \Psi_I | \frac{\partial H}{\partial x} | \Psi_J \rangle , \tag{4.2}$$

in virtue of the off-diagonal Hellmann–Feynman theorem. The magnitude of the NAC vector is ill-defined at a conical intersection, since it diverges as the inverse of the energy difference [see Eq. (4.2)]. The numerator, however, is well-defined and often called the derivative coupling vector; note that the nomenclature is not fixed in the literature. It can be viewed as a transition gradient. The other vector that forms the branching space together with the derivative coupling is the gradient (half) difference,

$$G_{IJ} = \frac{1}{2} \left(\langle \Psi_J | \frac{\partial H}{\partial x} | \Psi_J \rangle - \langle \Psi_I | \frac{\partial H}{\partial x} | \Psi_I \rangle \right) . \tag{4.3}$$

The vectors G_{IJ} and $(E_J - E_I)D_{IJ}$ – often denoted g and h vectors or x_1 and x_2 vectors in this context – play symmetrical roles: they form the two directions that make the adiabatic energy difference increase to first order from zero at a conical intersection. They actually are undetermined up to within a mutual rotation, which directly reflects the freedom in defining two specific degenerate eigenstates (see, e.g., Ref. 273).

In addition to being essential for the correct capture of the conical topography of crossings, non-adiabatic couplings are required for describing the coupled equations that govern the nuclear components of the molecular wavefunction. As already pointed out, they become large when the energy gap between electronic states decreases, which is why conical intersections are key for describing radiationless processes whereby population is transferred among electronic states. In practice, non-adiabatic quantum dynamics is often better described in terms of quasidiabatic electronic states that result from a unitary transformation of a relevant subset of coupled adiabatic states. They vary smoothly enough with respect to nuclear coordinates to allow for neglect of kinetic couplings but introduce instead nonzero potential couplings.

Further investigations of conical intersections are beyond the scope of the present work and the literature on the subject is vast. We refer for example to Ref. 274 for a comprehensive review of relevant concepts. Finally, let us stress that the ability of a computational method to describe correctly the topography of a conical intersection is intimately related to the formal possibility of using analytic derivative techniques for evaluating non-adiabatic couplings [264]. This somewhat relies on the fact that the final step of the whole computational procedure should be viewed as a Hamiltonian submatrix diagonalization that provides several eigenstates democratically within the same subspace. The state-averaged multi-configurational self-consistent-field (SA-MCSCF) method is an evident option in this context, with analytic derivatives applied similarly to diagonal and off-diagonal terms [268].

State-averaged multi-configurational self-consistent-field method (SA-MCSCF)

The electronic structure Hamiltonian with its one- and two-body integrals are defined in Eq. (1.14)-(1.16). Due to the exponential increase of the configuration space in the number of molecular orbitals, it is of common use to select only a restricted (and, ideally, relevant) part of it in practical calculation, for instance by considering the active space approximation where the orbital space is separated into a set of frozen occupied, active and virtual orbitals. In such a reduced configuration space, the configuration interaction method is not invariant anymore under orbital rotations [21, 31] and the choice of orbitals will influence the quality of the result. Hence, one has to consider the re-optimization of the orbitals, thus leading to the MCSCF model for which wavefunction reads:

$$|\Psi(\boldsymbol{\kappa},\mathbf{c})\rangle = e^{-\hat{\kappa}} \left(\sum_i c_i \, |\Phi_i\rangle \right), \tag{4.4}$$

where $\{|\Phi_i\rangle\}$ are Slater determinants or configuration state functions, and $\hat{U}_{\mathrm{O}}(\boldsymbol{\kappa}) = e^{-\hat{\kappa}}$ is the orbital-rotation operator. The latter is defined as follows in the spin-restricted formalism with real algebra:

$$\hat{\kappa} = \sum_{p>q}^{\mathrm{MOs}} \kappa_{pq} (\hat{E}_{pq} - \hat{E}_{qp}). \tag{4.5}$$

The parameters of the wavefunction in Eq. (4.4) are determined by variationally optimizing the expectation value of the energy:

$$E = \min_{\boldsymbol{\kappa},\mathbf{c}} \frac{\langle \Psi(\boldsymbol{\kappa},\mathbf{c})| \, H \, |\Psi(\boldsymbol{\kappa},\mathbf{c})\rangle}{\langle \Psi(\boldsymbol{\kappa},\mathbf{c}) | \Psi(\boldsymbol{\kappa},\mathbf{c})\rangle}. \tag{4.6}$$

In order to have a democratic description of ground and excited states, one can simultaneously optimize several MCSCF states that are generated from the same orbital basis. As extensively discussed in Ref. 21, it is convenient to introduce an exponential unitary parametrization of the configuration space with nonredundant variables,

$$\hat{U}_{\mathrm{C}}(\mathbf{S}) = e^{-\hat{S}}, \tag{4.7}$$

where

$$\hat{S} = \sum_{J} \sum_{K>J} S_{KJ} \left(\left| \Psi_K^{(0)} \right\rangle \left\langle \Psi_J^{(0)} \right| - \left| \Psi_J^{(0)} \right\rangle \left\langle \Psi_K^{(0)} \right| \right) \tag{4.8}$$

and

$$\left| \Psi_I^{(0)} \right\rangle = \sum_i c_{Ii}^{(0)} \left| \Phi_i \right\rangle \tag{4.9}$$

are initial orthonormal states built from the same set of molecular orbitals. Within the SA-MCSCF model, the wavefunctions are subject to a double-exponential parametrization

$$\left| \Psi_I(\boldsymbol{\kappa}, \mathbf{S}) \right\rangle = e^{-\hat{\kappa}} e^{-\hat{S}} \left| \Psi_I^{(0)} \right\rangle, \tag{4.10}$$

where, according to the generalization of the Rayleigh–Ritz variational principle for an ensemble of ground and excited states [275], the parameters are variationally optimized by minimizing the state-averaged energy

$$E^{\text{SA-MCSCF}} = \min_{\boldsymbol{\kappa}, \mathbf{S}} \sum_I w_I \left\langle \Psi_I(\boldsymbol{\kappa}, \mathbf{S}) \right| H \left| \Psi_I(\boldsymbol{\kappa}, \mathbf{S}) \right\rangle, \tag{4.11}$$

where $\sum_I w_I = 1$ and the states are automatically orthonormalized as they are generated from unitary transformations of the initial orthonormal states $\{ \left| \Psi_I^{(0)} \right\rangle \}$. Note that due to the orbital optimization, the converged individual and state-averaged energies may vary with the weights. In practice, the equal weight SA-MCSCF (where all weights are equal) is usually considered. Finally, the dependence on $\boldsymbol{\kappa}$ in the wavefunctions can actually be transferred to the electronic integrals in the Hamiltonian, *i.e.* $h_{pq} \rightarrow h_{pq}(\boldsymbol{\kappa})$ and $g_{pqrs} \rightarrow g_{pqrs}(\boldsymbol{\kappa})$, such that Eq. (4.11) equivalently reads

$$E^{\text{SA-MCSCF}} = \min_{\boldsymbol{\kappa}, \mathbf{S}} \sum_I w_I \left\langle \Psi_I(\mathbf{S}) \right| H(\boldsymbol{\kappa}) \left| \Psi_I(\mathbf{S}) \right\rangle, \tag{4.12}$$

where $H(\boldsymbol{\kappa}) = \hat{U}_{\mathrm{O}}^{\dagger}(\boldsymbol{\kappa}) H \hat{U}_{\mathrm{O}}(\boldsymbol{\kappa})$ is the MO-basis transformed Hamiltonian.

While the SA-MCSCF method allows for a democratic description of ground and excited states, it is only variational with respect to the state-averaged energy, so that an individual state is not variational. This makes the calculation of analytical energy gradients of each individual state more complicated, as it requires the introduction of specific Lagrangians and the solution of so-called coupled-perturbed equations, as further discussed in Sec. 4.2.2.

4.2.2 Estimation of energies, analytical gradients and non-adiabatic coupling on a quantum computer

State-averaged orbital-optimized variational-quantum-eigensolver (SA-OO-VQE)

The variational quantum eigensolver (VQE) [54, 91] represents one of the most promising methods to estimate the ground-state energy on near-term quantum computers. As suggested by the name of the algorithm, the VQE relies on the Rayleigh–Ritz variational principle and consists in finding the closest approximation to the ground-state wavefunction thanks to a given *ansatz* (defined by a parametrized unitary operation $\hat{U}(\boldsymbol{\theta})$). Applying this unitary operation to a chosen initial state (usually very easy to prepare, such as the Hartree–Fock (HF) Slater determinant $|\mathrm{HF}\rangle$) leads to a parametrized trial wavefunction $|\Psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})\,|\mathrm{HF}\rangle$, from which the associated energy is estimated by repeated measurements of the quantum circuit. Unfortunately, the extension of the VQE algorithm to excited state is not trivial, as a variational estimation of the excited-state energies can only be defined under orthogonal constraints. Such constraints have been considered by adding penalization terms to the Hamiltonian, thus leading to the state-specific variational quantum deflation (VQD) algorithm [101–104] where each state is determined by a separate minimization (or only two minimizations in total if the first one is performed on a state-average ensemble [276]). Other extensions can treat excited states on the same footing, but still favor the ground state [95, 97, 277, 278]. However, the proper description of conical intersections or avoided crossings require a democratic description of both the ground and excited states. Such an equal footing treatment can be achieved by performing a single minimization (or resolution) for all states sharing the same ansatz, as in multistate-contracted VQE (MC-VQE) [100, 279], fully-weighted subspace-search VQE (SS-VQE) [98], variance-VQE [280] and the quantum filter diagonalization method [279, 281]. Inspired by the SS-VQE method of Nakanishi *et al.* [98], we proposed the (equi-weighted) state-averaged orbital-optimized VQE (SA-OO-VQE), that can be seen as a combination of a state-averaged VQE (SA-VQE) and a state-averaged orbital-optimization (SA-OO) procedure. Let us briefly summarize each step of the SA-OO-VQE, focusing on an equi-ensemble of two-states (the extension to more electronic states is straightforward).

1. **Initialization:** Initialize the circuits with two orthonormal states $|\Phi_{\mathrm{A}}\rangle$ and $|\Phi_{\mathrm{B}}\rangle$.

2. **SA-VQE:** Apply a quantum ansatz (*i.e.* a given quantum circuit) to transform both initial states into trial states $|\Psi_{\mathrm{A}}(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\Phi_{\mathrm{A}}\rangle$ and $|\Psi_{\mathrm{B}}(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})|\Phi_{\mathrm{B}}\rangle$, and find the optimal set of ansätze parameters that

minimizes the state-averaged energy

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} E^{\text{SA-OO-VQE}}(\boldsymbol{\kappa}, \boldsymbol{\theta}) \tag{4.13}$$

for a fixed orbital basis $\boldsymbol{\kappa}$, where the state-average energy reads

$$\begin{aligned} &E^{\text{SA-OO-VQE}}(\boldsymbol{\kappa}, \boldsymbol{\theta}) = \\ &w_{\text{A}} \langle \Psi_{\text{A}}(\boldsymbol{\theta}) | H(\boldsymbol{\kappa}) | \Psi_{\text{A}}(\boldsymbol{\theta}) \rangle + w_{\text{B}} \langle \Psi_{\text{B}}(\boldsymbol{\theta}) | H(\boldsymbol{\kappa}) | \Psi_{\text{B}}(\boldsymbol{\theta}) \rangle \end{aligned} \tag{4.14}$$

with w_{A} and w_{B} the weights attributed to each state with the normalization condition $w_{\text{A}} + w_{\text{B}} = 1$. Note that this energy is lower-bounded by the ensemble energy of the exact two lowest eigenstates (denoted by $|\Psi_0\rangle$ and $|\Psi_1\rangle$) of $H(\boldsymbol{\kappa}^*)$ in the active-space approximation, according the variational principle [275].

3. **SA-OO:** Rotate the orbital basis to find the optimal set of parameters that minimize the state-averaged energy

$$\boldsymbol{\kappa}^* = \arg\min_{\boldsymbol{\kappa}} E^{\text{SA-OO-VQE}}(\boldsymbol{\kappa}, \boldsymbol{\theta}) \tag{4.15}$$

(*e.g.* with Newton-Raphson), for a fixed set of parameters $\boldsymbol{\theta}$.

4. **SA-OO-VQE:** Repeat steps 2 and 3 until the state-average energy is minimized with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$, *i.e.* find

$$(\boldsymbol{\kappa}^*, \boldsymbol{\theta}^*) = \arg\min_{\boldsymbol{\kappa}, \boldsymbol{\theta}} E^{\text{SA-OO-VQE}}(\boldsymbol{\kappa}, \boldsymbol{\theta}). \tag{4.16}$$

As discussed in previous works [30, 98], the lower bound in Eq. (4.14) is uniquely defined if $w_A > w_B$, but is invariant under any rotation between $|\Psi_0\rangle$ and $|\Psi_1\rangle$ in the equi-ensemble case ($w_A = w_B$). Hence, considering the case $w_A = w_B$ does not guarantee that the optimized states $|\Psi_A(\boldsymbol{\theta}^*)\rangle$ and $|\Psi_B(\boldsymbol{\theta}^*)\rangle$ are the closest approximation of the eigenstates $|\Psi_0\rangle$ and $|\Psi_1\rangle$. However, this enforces the definition of a well-defined two-state subspace spanned by either Ψ_A and Ψ_B or Ψ_0 and Ψ_1, such that the latter are eigenstates. Forcing this correspondence (that we refer to as the *state-resolution*) is a complicated task that can be handled in different ways. Considering $w_A > w_B$ is a straightforward solution, but this constraint may complicate the SA-VQE optimization considerably [98]. Additional tricks can be used in the equi-ensemble case, by considering additional cost-functions to be maximized [98], a classical diagonalization [100], or another type of cost-functions that use the variance of the states [280]. In Sec. 4.2.2, we discuss another approach, inspired by the one of Nakanishi [98], to solve the state-resolution of a two-state ensemble within the SA-OO-VQE algorithm, when the initial states are the HF Slater determinant and any singlet singly-excited configuration interaction (CIS) state. Note that while we focus on those particular initial states in this manuscript, any other choice could in principle be considered.

State-resolution procedure

In this section, we propose another method to capture the active-space eigenvectors of $H(\boldsymbol{\kappa})$), which requires few additional gates and a negligible increase in the number of measurements. Our approach consists in taking advantage of the rotational invariance of the equi-ensemble state-averaged energy, in order to postpone the state-resolution to the very end of the SA-OO-VQE algorithm. Considering the equi-ensemble (*i.e.* $w_{\mathrm{A}} = w_{\mathrm{B}}$), after convergence of the SA-OO-VQE algorithm, the resulting Hilbert space spanned by $|\Psi_{\mathrm{A}}(\boldsymbol{\theta}^*)\rangle$ and $|\Psi_{\mathrm{B}}(\boldsymbol{\theta}^*)\rangle$ is a good approximation to the subspace spanned by the SA-CASSCF states – classical analogue of the SA-OO-VQE method –, within the same active space. (As discussed in Sec. 4.2.2, the SA-OO-VQE states are not constrained to be the eigenvectors of H, or, equivalently, to form the adiabatic basis that diagonalizes $H(\boldsymbol{\kappa}^*)$). To resolve the ground and first excited SA-OO-VQE *eigenstates* (which should be good approximations to $\left|\Psi_0^{\text{SA-CASSCF}}\right\rangle$ and $\left|\Psi_1^{\text{SA-CASSCF}}\right\rangle$), we propose to implement a rotation between the initial states $|\Phi_{\mathrm{A}}\rangle$ and $|\Phi_{\mathrm{B}}\rangle$, such that the new rotated initial states become

$$\begin{aligned} |\Phi_0(\varphi)\rangle &= \cos\varphi|\Phi_{\mathrm{A}}\rangle + \sin\varphi|\Phi_{\mathrm{B}}\rangle, \\ |\Phi_1(\varphi)\rangle &= -\sin\varphi|\Phi_{\mathrm{A}}\rangle + \cos\varphi|\Phi_{\mathrm{B}}\rangle, \end{aligned} \tag{4.17}$$

where we have the property $|\Phi_1(\varphi)\rangle = |\Phi_0(\varphi + \pi/2)\rangle$. After applying the ansatz with optimized parameters $\boldsymbol{\theta}^*$, these new rotated initial states evolve to

$$\begin{aligned} |\Psi_0(\varphi, \boldsymbol{\theta}^*)\rangle &= \cos\varphi|\Psi_{\mathrm{A}}(\boldsymbol{\theta}^*)\rangle + \sin\varphi|\Psi_{\mathrm{B}}(\boldsymbol{\theta}^*)\rangle, \\ |\Psi_1(\varphi, \boldsymbol{\theta}^*)\rangle &= -\sin\varphi|\Psi_{\mathrm{A}}(\boldsymbol{\theta}^*)\rangle + \cos\varphi|\Psi_{\mathrm{B}}(\boldsymbol{\theta}^*)\rangle, \end{aligned} \tag{4.18}$$

which leads to a rotation between the final SA-OO-VQE states. Note that they remain orthonormal and, by virtue of the equi-ensembles properties, lead to the same state-averaged energy that is invariant with respect to φ. The state-resolution amounts to finding the value $\varphi \rightarrow \varphi^*$ that minimizes the energy of $|\Psi_0(\varphi, \boldsymbol{\theta}^*)\rangle$,

$$\varphi^* = \arg\min_{\varphi} \langle\Psi_0(\varphi, \boldsymbol{\theta}^*)| \, H(\boldsymbol{\kappa}^*) \, |\Psi_0(\varphi, \boldsymbol{\theta}^*)\rangle \tag{4.19}$$

(or, equivalently, maximizes the one of $|\Psi_1(\varphi, \boldsymbol{\theta}^*)\rangle$), thus making both $|\Psi_0(\varphi^*, \boldsymbol{\theta}^*)\rangle$ and $|\Psi_1(\varphi^*, \boldsymbol{\theta}^*)\rangle$ approximated eigenstates of $H(\boldsymbol{\kappa}^*)$. Satisfying Eq. (4.19) can be seen as the fifth step of the SA-OO-VQE algorithm (see Sec. 4.2.2 for the first four steps). In Fig. 4.1, we show the short-depth circuit we specifically developed to perform the rotation between the HF determinant $|\Phi_{\mathrm{A}}\rangle = |\mathrm{HF}\rangle$ and a singlet-excited CIS state $|\Phi_{\mathrm{B}}\rangle = -\hat{E}_{\mathrm{hl}} \, |\mathrm{HF}\rangle \, /\sqrt{2}$, where 'h' and 'l' refer to

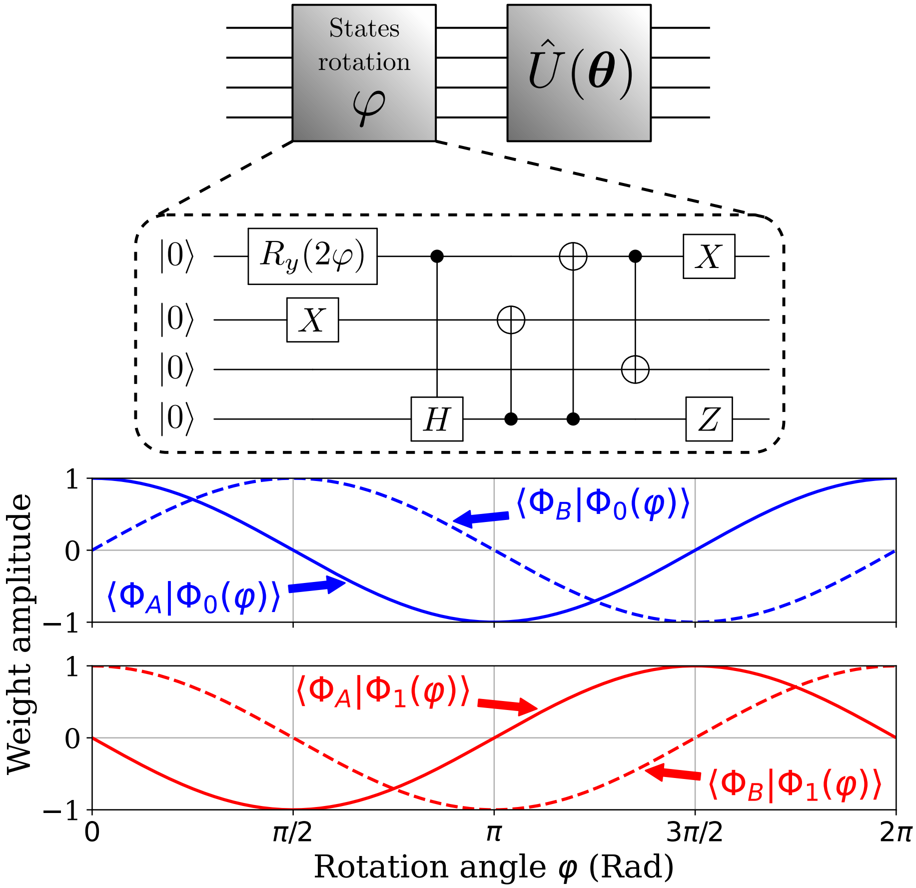


Figure 4.1: Upper panel: Short-depth quantum circuit specifically designed for the state-resolution of the SA-OO-VQE algorithm, to build the state $|\Phi(\varphi)\rangle = \cos\varphi\,|\Phi_\mathrm{A}\rangle + \sin\varphi\,|\Phi_\mathrm{B}\rangle$, i.e. a rotation between the HF state $|\Phi_\mathrm{A}\rangle = |\mathrm{HF}\rangle$ and the HOMO-LUMO singlet CIS state $|\Phi_\mathrm{B}\rangle = \hat{E}_\mathrm{hl}|\mathrm{HF}\rangle/\sqrt{2}$. **Lower panel:** Overlaps $\langle\Phi_A|\Phi_{0/1}(\varphi)\rangle$ and $\langle\Phi_B|\Phi_{0/1}(\varphi)\rangle$ as a function of the rotation parameter φ (with $|\Phi_\mathrm{A}\rangle = |\mathrm{HF}\rangle$ and $|\Phi_\mathrm{B}\rangle = \hat{E}_\mathrm{hl}|\mathrm{HF}\rangle/\sqrt{2}$).

the HOMO and LUMO orbitals, respectively. In practice, the circuit works as follows (for the sake of simplicity, we focus on a 2-spatial-orbital (4-spin-orbital) – *i.e.* 4-qubit – and 2-electron CAS). Starting with the 4 qubits in the $|0\rangle$ state, a $R_y(2\varphi)$ rotation gate and a X gate are applied to the first and second qubit, respectively, thus leading to the quantum superposition

$$R_y^0(2\varphi)X^1\,|0000\rangle = \cos\varphi\,|0100\rangle + \sin\varphi\,|1100\rangle\,. \tag{4.20}$$

Then, a controlled-Hadamard gate transforms the state $|1100\rangle$ into $\frac{1}{\sqrt{2}}(|1100\rangle + |1101\rangle)$, which evolves into $\frac{1}{\sqrt{2}}(|1110\rangle + |0001\rangle)$ after applying three CNOT gates. The first term in the right hand side of Eq. (4.20) remains invariant with respect to the aforementioned operations, such that the state now reads

$$\cos\varphi\,|0100\rangle + \frac{\sin\varphi}{\sqrt{2}}(|1110\rangle + |0001\rangle). \tag{4.21}$$

Finally, we apply X and Z gates to the first and last qubits, respectively, to arrive at the final expression

$$\begin{aligned}|\Phi_0(\varphi)\rangle &= \cos\varphi\,|1100\rangle + \frac{\sin\varphi}{\sqrt{2}}\,(|0110\rangle - |1001\rangle)\\ &= \cos\varphi\,|\Phi_\mathrm{A}\rangle + \sin\varphi\,|\Phi_\mathrm{B}\rangle\,,\end{aligned} \tag{4.22}$$

where $|\Phi_\mathrm{A}\rangle = |\mathrm{HF}\rangle$ and $|\Phi_\mathrm{B}\rangle = -\hat{E}_\mathrm{hl}|\mathrm{HF}\rangle/\sqrt{2} = (|0110\rangle - |1001\rangle)/\sqrt{2}$ is the HOMO–LUMO singlet-excited CIS state. Replacing $\varphi \to \varphi + \pi/2$ in Eq. (4.22), one recovers $|\Phi_1(\varphi)\rangle$ in Eq. (4.17), such that the parameter φ can be tuned to realize any real linear combination between $|\Phi_\mathrm{A}\rangle$ and $|\Phi_\mathrm{B}\rangle$, as illustrated in the lower panels of Fig. 4.1. Note that this circuit is valid for any singlet-excited CIS state $|\Phi_\mathrm{B}\rangle$, by simply applying the quantum gates to the qubits associated to the orbitals involved in the excitation.

Note that the idea introduced here for the state-resolution procedure in SA-OO-VQE follows closely the one proposed by Nakanishi *et al.* (Sec II.A. of Ref. 98). Indeed, the additional circuit in Fig. 4.1 is equivalent to their additional unitary operation $V(\phi)$, for which we provide an explicit form for any two-state ensemble (with a specific focus on initial states that are the HF and any singlet-excited CIS states). Note also that a SA-OO and SA-VQE algorithms are alternatively employed in our method. The resulting SA-OO-VQE subspace is then more meaningful in terms of electronic correlations as it (ideally) provides analog results as in the SA-CASSCF method, contrary to the SS-VQE scheme that is equivalent to the CASCI method. Because we work with an equi-ensemble, the state-resolution can be performed at the very end of the SA-OO-VQE algorithm only. This attractive feature of the equi-ensemble SA-OO-VQE allows in principle to spare a lot of unnecessary quantum resources, as one can still end up with the (approximate) eigenstates without requiring harder optimization procedures or additional quantum measurements at each instance of the SA-VQE algorithm.

Analytical gradients

Molecular properties can be accessed by estimating energy gradients with respect to a given perturbation [164]. Analytical expressions, that are cheaper and more precise than finite difference techniques, have been derived in the context of ground-state VQE in Refs. [106–108, 282]. In this section, we turn towards the question of the analytical evaluation of individual-state nuclear energy gradient with the SA-OO-VQE algorithm (which will be noted $|\Psi_I\rangle$, with $I = 0, 1, \ldots$). As opposed to the state-specific orbital-optimized VQE introduced by Mizukami *et al.* [283] and Takeshita *et al.* [265], each set of variational parameters (in our case $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$) is not optimized to minimize each individual-state energy,

$$\frac{\partial E_I(\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial \kappa_{pq}} \neq 0, \frac{\partial E_I(\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial \theta_n} \neq 0, \tag{4.23}$$

but rather to minimize the state-averaged energy,

$$\frac{\partial E_{\mathrm{SA}}(\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial \kappa_{pq}} = \frac{\partial E_{\mathrm{SA}}(\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial \theta_n} = 0, \tag{4.24}$$

where in both Eq. (4.23) and Eq. (4.24) it is implicit that the gradients are evaluated at the converged parameters. This renders the estimation of the individual-state nuclear energy gradients more complicated, as it has to take into account the non-variational character of the method. Fortunately, one can build analytical Lagrangians that are fully variational with respect to every parameter [284], such that their optimization facilitates the estimation of the targeted quantities (*e.g.* energy derivatives and non-adiabatic couplings in our case).

Following this strategy, we build an individual-state Lagrangian $\mathcal{L}_I$ that depends on all the parameters as follows,

$$\mathcal{L}_I = E_I + \left(\sum_{pq} \overline{\kappa}^I_{pq} \frac{\partial E_{\mathrm{SA}}}{\partial \kappa_{pq}} - 0 \right) + \left(\sum_{n} \overline{\theta}^I_{n} \frac{\partial E_{\mathrm{SA}}}{\partial \theta_{n}} - 0 \right). \tag{4.25}$$

Note that, based on the state-averaged variational conditions given in Eq. (4.24), the correspondence $\mathcal{L}_I = E_I$ holds here. In the definition of the Lagrangian $\mathcal{L}_I$, the parameters $\overline{\kappa}^I_{pq}$ and $\overline{\theta}^I_n$ are Lagrange multipliers designed to make it fully stationary such that

$$\frac{\partial \mathcal{L}_I}{\partial \overline{\kappa}^I_{pq}} = \frac{\partial \mathcal{L}_I}{\partial \overline{\theta}^I_{n}} = \frac{\partial \mathcal{L}_I}{\partial \kappa_{pq}} = \frac{\partial \mathcal{L}_I}{\partial \theta_{n}} = 0. \tag{4.26}$$

To fulfil the stationary conditions in Eq. (4.26), the Lagrange multipliers are determined by solving the so-called coupled-perturbed equations

$$\begin{aligned}\frac{\partial \mathcal{L}_I}{\partial \kappa_{rs}} &= \frac{\partial E_I}{\partial \kappa_{rs}} + \sum_{pq} \overline{\kappa}^I_{pq} \mathcal{H}^{\mathrm{OO}}_{pq,rs} + \sum_n \overline{\theta}^I_n \mathcal{H}^{\mathrm{CO}}_{n,rs} = 0,\\ \frac{\partial \mathcal{L}_I}{\partial \theta_m} &= \frac{\partial E_I}{\partial \theta_m} + \sum_{pq} \overline{\kappa}^I_{pq} H^{\mathrm{OC}}_{pq,m} + \sum_n \overline{\theta}^I_n \mathcal{H}^{\mathrm{CC}}_{n,m} = 0,\end{aligned} \tag{4.27}$$

where we have introduced

$$\mathcal{H}^{\mathrm{OO}}_{pq,rs} = \frac{\partial^2 E_{\mathrm{SA}}}{\partial \kappa_{pq} \partial \kappa_{rs}}, \tag{4.28}$$

$$\mathcal{H}^{\mathrm{CC}}_{n,m} = \frac{\partial^2 E_{\mathrm{SA}}}{\partial \theta_n \partial \theta_m}, \tag{4.29}$$

$$\mathcal{H}^{\mathrm{CO}}_{n,rs} = \frac{\partial^2 E_{\mathrm{SA}}}{\partial \theta_n \partial \kappa_{rs}}, \tag{4.30}$$

which correspond to matrix elements of the (state-averaged) orbital hessian $\mathcal{H}^{\mathrm{OO}}$, circuit hessian $\mathcal{H}^{\mathrm{CC}}$ and circuit-orbital hessian $\mathcal{H}^{\mathrm{CO}}$ (with $\mathcal{H}^{\mathrm{CO}} = (\mathcal{H}^{\mathrm{OC}})^T$). The remaining terms

$$\mathcal{G}^{\mathrm{O},I}_{rs} = \frac{\partial E_I}{\partial \kappa_{rs}} \quad \text{and} \quad \mathcal{G}^{\mathrm{C},I}_{m} = \frac{\partial E_I}{\partial \theta_m}, \tag{4.31}$$

are elements of the circuit gradient vector $\boldsymbol{\mathcal{G}}^{\mathrm{C},I}$ and the orbital gradient vector $\boldsymbol{\mathcal{G}}^{\mathrm{O},I}$ of the state $|\Psi_I(\boldsymbol{\theta})\rangle$. The orbital gradient for individual states $\mathcal{G}^{\mathrm{O},I}_{pq}$ can be relatively easily computed from their one- and two-RDMs and MO coefficients. The state-averaged orbital Hessian $\mathcal{H}^{\mathrm{CC}}_{pq,rs}$ can also be determined from the state-averaged one- and two-RDMs and MO coefficients [267, 271]. For more information, see Appendix 5.C in Chapter 5 for a concise formulation. Therefore, they do not require any additional measurements on the quantum computer. However, the first and second derivatives of the state-averaged energy with respect to the ansätze parameters requires many more measurements (but no additional qubits or deeper circuit depth). According to the parameter-shift rule [285], 2^{2n} and 2^{4n} measurements are required to compute the gradient and Hessian with respect to ansätze parameters, respectively, for an ansatz with up to n-fold fermionic excitation operators. Note that $n = 2$ is usually considered, as the trotterized-UCCSD ansatz can be made arbitrarily exact, as shown by Evangelista *et al* [176], thus corresponding to 16 and 256 expectation values per element of the $\mathcal{H}^{\mathrm{CO}}$ and $\mathcal{H}^{\mathrm{CC}}$ matrices, respectively. Finally, the number of parameter-shifted RDMs to be measured on the quantum computer is directly related to the number of ansätze-parameters N_p. For the circuit-orbital Hessian $\mathcal{H}^{\mathrm{CO}}$, N_p parameter-shifted RDMs need to be measured, while $N_p(N_p+1)/2$ are required to estimate the (symmetric) circuit-circuit Hessian matrix $\mathcal{H}^{\mathrm{CC}}$. The total number of ansätze-parameters also scales with the number of active orbitals as $\mathcal{O}(N_{\mathrm{act}}^4)$ for the generalized UCCD ansatz considered in this work. For the sake of conciseness, we refer the interested reader to Appendix 4.A for more details about the estimation of the above Hessian matrices and gradients vectors.

Assuming we have evaluated the necessary Hessian matrices and gradient vectors out of a quantum circuit following Appendix 4.A, the Lagrange multipliers $\overline{\kappa}^I_{pq}$ and $\overline{\theta}^I_n$ satisfying the conditions in Eq. (4.26) are determined on a classical computer by solving the following matrix equation

$$\begin{pmatrix} \mathcal{H}^{\mathrm{OO}} & \mathcal{H}^{\mathrm{OC}} \\ \mathcal{H}^{\mathrm{CO}} & \mathcal{H}^{\mathrm{CC}} \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\kappa}}^I \\ \overline{\boldsymbol{\theta}}^I \end{pmatrix} = - \begin{pmatrix} \boldsymbol{\mathcal{G}}^{\mathrm{O},I} \\ \boldsymbol{\mathcal{G}}^{\mathrm{C},I} \end{pmatrix}. \tag{4.32}$$

Inserting these Lagrange multipliers back into Eq. (4.25) makes the Lagrangian fully stationary, and the property $\frac{\mathrm{d}E_I}{\mathrm{d}x} = \frac{\mathrm{d}\mathcal{L}_I}{\mathrm{d}x} = \frac{\partial \mathcal{L}_I}{\partial x}$ holds [284]. Hence, the energy derivative $\frac{\mathrm{d}E_I}{\mathrm{d}x}$ can be evaluated as follows:

$$\begin{aligned} \frac{dE_I}{dx} &= \sum_{pq} \frac{\partial h_{pq}}{\partial x} \gamma^{I,\mathrm{eff}}_{pq} + \frac{1}{2} \sum_{pqrs} \frac{\partial g_{pqrs}}{\partial x} \Gamma^{I,\mathrm{eff}}_{pqrs} \\ &\quad + \sum_J \sum_n w_J \bar{\theta}^I_n \mathcal{G}^{\mathrm{C},J}_n (\tfrac{\partial H}{\partial x}), \end{aligned} \tag{4.33}$$

with effective 1- and 2-RDMs defined by

$$\boldsymbol{\gamma}^{I,\mathrm{eff}} = \boldsymbol{\gamma}^I + \tilde{\boldsymbol{\gamma}}^{I,\mathrm{SA}} \tag{4.34}$$

$$\boldsymbol{\Gamma}^{I,\mathrm{eff}} = \boldsymbol{\Gamma}^I + \tilde{\boldsymbol{\Gamma}}^{I,\mathrm{SA}} \tag{4.35}$$

with $\gamma^I_{pq} = \langle \Psi_I | \hat{E}_{pq} | \Psi_I \rangle$ and $\Gamma^I_{pqrs} = \langle \Psi_I | \hat{e}_{pqrs} | \Psi_I \rangle$ regular RDMs of the reference state $|\Psi_I\rangle$, supplemented by corrective state-averaged RDMs $\tilde{\boldsymbol{\gamma}}^{I,\mathrm{SA}}$ and $\tilde{\boldsymbol{\Gamma}}^{I,\mathrm{SA}}$ (encoding orbital contributions) with matrix elements

$$\tilde{\gamma}^{I,\mathrm{SA}}_{pq} = \sum_o \big(\gamma^{\mathrm{SA}}_{oq} \overline{\kappa}^I_{op} + \gamma^{\mathrm{SA}}_{po} \overline{\kappa}^I_{oq} \big) \tag{4.36}$$

$$\begin{aligned} \tilde{\Gamma}^{I,\mathrm{SA}}_{pqrs} = \sum_o \big(& \Gamma^{\mathrm{SA}}_{oqrs} \overline{\kappa}^I_{op} + \Gamma^{\mathrm{SA}}_{pors} \overline{\kappa}^I_{oq} \\ & + \Gamma^{\mathrm{SA}}_{pqos} \overline{\kappa}^I_{or} + \Gamma^{\mathrm{SA}}_{pqro} \overline{\kappa}^I_{os} \big). \end{aligned} \tag{4.37}$$

Note that building these effective matrices does not require any additional measures from the quantum circuit as the RDMs $\boldsymbol{\gamma}^I$ and $\boldsymbol{\Gamma}^I$ are already evaluated during the SA-OO-VQE to estimate the state-averaged energy. The circuit gradient

$\mathcal{G}^{\mathrm{C},J}(\frac{\partial H}{\partial x})$ introduced in Eq. (4.33) is defined such that

$$\mathcal{G}_n^{\mathrm{C},J}(\tfrac{\partial H}{\partial x}) = \frac{\partial}{\partial \theta_n} \langle \Psi_J | \frac{\partial H}{\partial x} | \Psi_J \rangle \tag{4.38}$$

and can be estimated out of a quantum circuit in the same way as for a generic energy gradient using for example the parameter-shift rule (cf Appendix 4.A). The change being here that the central operator is now the nuclear derivative of the Hamiltonian $\partial H / \partial x$ which can be evaluated on a classical computer as shown in Appendix 4.B. Note that, we also refer the interested reader to Appendix 4.B for practical details about nuclear derivatives of electronic integrals (in Eq. (4.33)) which can be evaluated on a classical computer with common quantum chemistry packages.

Interestingly, compared to its classical analogue SA-CASSCF, note that a unique set of ansätze parameters $\boldsymbol{\theta}$ is considered to simultaneously find both ground and first excited states in SA-OO-VQE, instead of the configuration-interaction (CI) coefficients for each state (denoted by $\mathbf{c_0}$ and $\mathbf{c_1}$). This results in a much reduced size of the parameter space,

$$\dim(\boldsymbol{\theta}) \ll \dim(\mathbf{c}_0) + \dim(\mathbf{c}_1). \tag{4.39}$$

This has important consequences, as the original CP-MCSCF equations sometimes cannot be solved due to memory issues in storing all the matrix elements of $\mathcal{H}^{\mathrm{CC}}$, although some alternative implementations have been proposed to overcome this problem (see Ref. 271 and references therein). Hence, the classical complexity in solving the coupled-perturbed equations [Eq. (4.32)] is considerably reduced in SA-OO-VQE compared to SA-CASSCF, at the expense of a lower accuracy (as the SA-VQE solver is not exact in contrast to SA-CASSCF).

Non-adiabatic couplings

Non-adiabatic couplings have been calculated recently by Tamiya *et al.* [286] in the context of SS-VQE without any orbital optimization. In this work, we provide an analytical approach to estimate non-adiabatic couplings within the SA-OO-VQE algorithm, for which the state-averaged orbital-optimization procedure implies a more involved derivation. The definition and Hellmann–Feynman formula for the NAC, D_{IJ}, have been given above – see Eq. (4.1) and Eq. (4.2) – in the ideal case of exact adiabatic eigenstates. It is well-known in the practical context of an MCSCF ansatz that this term actually splits into two contributions: (i) a typically larger CI-contribution, which obeys a Hellmann–Feynman like formula (except that eigenstates are now CI-coefficient vectors and the Hamiltonian operator is replaced by its finite matrix representation in the CSF basis set); (ii) a typically smaller CSF-contribution, which accounts for molecular orbital gradients (via both their expansion coefficients and the overlaps among the primitive atomic basis functions); see, e.g., Ref. 268. While the latter CSF contribution is usually straightforward to estimate, the former CI contribution is a more involved term which should take into account the non-variational character of MCSCF wavefunctions. Fortunately, coupled-perturbed equations have been derived to treat this aspect based on the same machinery as for gradient calculation [266–272]. We employed a similar approach to obtain an analytical estimation of NACs with SA-OO-VQE wavefunctions. For sake of conciseness, we will present in the following only the essential equations of our developments (we refer the interested reader to Appendix 4.C where we detail each step of the derivation). In practice, one has to solve the following set of coupled linear equations

$$\begin{pmatrix} \mathcal{H}^{\mathrm{OO}} & \mathcal{H}^{\mathrm{OC}} \\ \mathcal{H}^{\mathrm{CO}} & \mathcal{H}^{\mathrm{CC}} \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\kappa}}^{IJ} \\ \overline{\boldsymbol{\theta}}^{IJ} \end{pmatrix} = - \begin{pmatrix} \boldsymbol{\mathcal{G}}^{\mathrm{O},IJ} \\ 0 \end{pmatrix}, \tag{4.40}$$

to determine the NAC Lagrange multipliers $\overline{\boldsymbol{\kappa}}^{IJ}$ and $\overline{\boldsymbol{\theta}}^{IJ}$. In Eq. (4.40), we retrieve the same Hessian blocks as for the gradient calculation and $\mathcal{G}_{pq}^{\mathrm{O},IJ} = \langle \Psi_I | \left(\partial \mathcal{H} / \partial \kappa_{pq} \right) | \Psi_J \rangle$ represents the interstate orbital coupling gradient whose elements can be easily measured out of a quantum computer (using for example methods provided in Ref. 98). Once the multipliers are determined using a classical computer, the NAC can be evaluated as follows:

$$\begin{aligned} D_{IJ} = \frac{1}{E_J - E_I} \Bigg(& \sum_{pq} \frac{\partial h_{pq}}{\partial x} \gamma_{pq}^{IJ,\mathrm{eff}} + \frac{1}{2} \sum_{pqrs} \frac{\partial g_{pqrs}}{\partial x} \Gamma_{pqrs}^{IJ,\mathrm{eff}} \\ & + \sum_K \sum_n w_K \overline{\theta}_n^{IJ} \mathcal{G}_n^{\mathrm{C},K} \left(\tfrac{\partial H}{\partial x} \right) \Bigg) \\ & - \frac{1}{2} \sum_{pq} \gamma_{pq}^{IJ} \big((\partial_x p | q) - (q | \partial_x p) \big). \end{aligned} \tag{4.41}$$

The effective transition 1- and 2-RDMs introduced here are defined by

$$\boldsymbol{\gamma}^{IJ,\mathrm{eff}} = \boldsymbol{\gamma}^{IJ} + \tilde{\boldsymbol{\gamma}}^{IJ,\mathrm{SA}} \tag{4.42}$$

$$\boldsymbol{\Gamma}^{IJ,\mathrm{eff}} = \boldsymbol{\Gamma}^{IJ} + \tilde{\boldsymbol{\Gamma}}^{IJ,\mathrm{SA}}, \tag{4.43}$$

and $\tilde{\boldsymbol{\gamma}}^{IJ,\mathrm{SA}}$ and $\tilde{\boldsymbol{\Gamma}}^{IJ,\mathrm{SA}}$ are the orbital contributions to the 1- and 2-RDMs, respectively. These state-averaged matrices are defined in a same way as in Eq. (4.36) (where we replace $\overline{\kappa}^I$ by $\overline{\kappa}^{IJ}$). In Eq. (4.41), the terms in parentheses encode the off-diagonal Hellmann–Feynman contribution complemented by additional corrective terms accounting for the non-variational character of the wavefunctions due to orbital and quantum circuit optimization. The contribution outside parenthesis is the so-called "CSF-term" which formally takes into account the variation of the Slater determinants due to nuclear displacement (see Ref. 266 for more details). The elements $(\partial_x p | q)$ represent the half-derivative of MOs' overlap which can be easily calculated analytically with most quantum chemistry packages. Note the presence of transition 1- and 2-RDMs in Eq. (4.41) defined as $\gamma_{pq}^{IJ} = \langle \Psi_I | \hat{E}_{pq} | \Psi_J \rangle$ and $\Gamma_{pqrs}^{IJ} = \langle \Psi_I | \hat{e}_{pqrs} | \Psi_J \rangle$. These matrices can be obtained from a quantum circuit using methods to determine transition matrix elements such as the one provided in Ref. 98.

4.3 Computational details

To test our theoretical developments, we consider the formaldimine molecule CH_2NH, a minimal Schiff base model relevant for the study of the photoisomerization in larger bio-molecules (such as the RPSB molecule whose *cis* to *trans* isomerization plays a key role in the visual cycle process [287–289]). An illustration of the geometry of the molecule is shown in Fig. 4.2a.

In analogy with our previous study [30], we freeze and constrain the N–CH_2 part of the molecule in the same plane. The interatomic distances are $d_{\mathrm{N-C}} = 1.498$ Å, $d_{\mathrm{C-H}} = 1.067$ Å, $d_{\mathrm{N-H}} = 0.987$ Å and the angle $\widehat{\mathrm{N{-}C{-}H}} = 118.36°$. The second H atom is symmetric to the first one with respect to the N–C axis. The two remaining degrees of freedom characterize the out-of-plane bending angle $\alpha \equiv \widehat{\mathrm{H{-}N{-}C}}$ and the dihedral angle $\phi \equiv \mathrm{H{-}}\widehat{\mathrm{N{-}C}}\mathrm{{-}H}$. For practical calculations, the cc-pVDZ basis is used and an active space of four electrons in three orbitals (4,3) is considered. The orbital optimization is realized over the 43 spatial-orbitals of the system (for SA-OO-VQE and SA-CASSCF). Reference quantum chemistry calculations are realized with OpenMolcas [290] (e.g. SA-CASSCF simulation and estimation of the associated gradients and NAC) whereas the Psi4 [291] package is used to provide SA-OO-VQE with initial data about the molecular system.

The noiseless state-vector simulation of the SA-OO-VQE algorithm is realized using the python quantum computing packages OpenFermion [175] and Cirq [292]. The ansatz we employ in the SA-VQE algorithm is a generalized unitary coupled cluster ansatz with spin-free double-excitation operators (GUCCD) such that

$$\hat{U}(\boldsymbol{\theta}) = e^{\hat{T}(\boldsymbol{\theta}) - \hat{T}^\dagger(\boldsymbol{\theta})}, \tag{4.44}$$

$$\hat{T}(\boldsymbol{\theta}) = \sum_{t,v,w,u}^{\text{active}} \theta_{tuvw} \sum_{\sigma,\tau=\uparrow,\downarrow} \hat{a}_{t\sigma}^\dagger \hat{a}_{v\tau}^\dagger \hat{a}_{w\tau} \hat{a}_{u\sigma}. \tag{4.45}$$

Our simulation considers ansätze parameters $\boldsymbol{\theta}$ initialized to zero and optimized using the "Sequential Least Squares Programming" (SLSQP) method from the python Scipy package. For each call of SA-VQE, the SLSQP method is run with a maximum number of 500 iterations and a precision threshold of 10^{-8} Ha. The threshold for the global convergence of SA-OO-VQE is also set to 10^{-8} Ha. A home-made python code has been developed to implement the state-averaged Newton-Raphson algorithm required for the SA-OO subalgorithm. For an active space of four electrons in three orbitals, we have a circuit depth of 2688 (+6 for the initial rotation circuit), and we optimize 12 parameters. We refer the interested reader to our previous work [30] for more details about the GUCCD ansatz (such as gate complexity).

4.4 Numerical results

4.4.1 Illustration of the final state resolution in SA-OO-VQE

As discussed in Sec. 4.2.2, the SA-OO-VQE states obtained after optimizing the $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$ parameters do not correspond to the eigenstates of $H(\boldsymbol{\kappa})$, and an additional rotation between the initial states is required (also called state resolution). In Fig. 4.2, we illustrate the convergence of each step of the SA-OO-VQE algorithm applied to the formaldimine molecule, depicted in panel (a), with geometry parameters set to $\phi = 80°$ and $\alpha = 100°$. The ground, first-excited and state-averaged energies are plotted on panel (b) for each step of the SA-OO-VQE algorithm, where SA-VQE and SA-OO phases are represented by white and grey strips, respectively. The ground and first-excited SA-CASSCF reference energies are also provided for comparison, as well as the state-averaged SA-CASSCF energy which forms a natural lower bound for SA-OO-VQE [see Eq. (4.14)]. As readily seen in panel (b), alternating between the SA-VQE and the SA-OO algorithms progressively lowers the state-averaged energy, requiring three full SA-OO-VQE cycles to reach global convergence. At convergence, this energy has an error of only $\sim 10^{-6}$ Ha with respect to SA-CASSCF, indicating that the subspace spanned by the SA-OO-VQE trial states is a very good approximation to the one spanned by the SA-CASSCF states. However, the converged individual SA-OO-VQE states differ significantly from the SA-CASSCF states (at the end of the third SA-OO-VQE step in panel (b)). Hence, one has to apply to state resolution such as described in Sec. 4.2.2 to recover the correct eigenstates. This final step is symbolized by the blue region in panel (b), where we employ the rotation circuit shown in Fig. 4.1 and optimize the rotation parameter φ such that the energies are effectively pushed as far as possible from each other, thus maximizing the difference between the first-excited and ground-state energies. After this final step, the individual SA-OO-VQE energies are in very good agreement with the SA-CASSCF ones (with an error of $\sim 10^{-6}$ Ha, similar to the state-averaged energy error). Note that such agreement is expected, as single and double excitations are enough to span all the electronic configurations in the case of an active space (4,3). Deviations from the SA-CASSCF results may appear when considering larger active spaces.

In panels (c) and (d) of Fig. 4.2, we show the one-dimensional (1D) PES along the α-angle for SA-OO-VQE and SA-CASSCF with a dihedral angle $\phi = 85°$. More precisely, we compare the 1D-PES of SA-OO-VQE without (panel (c)) and with (panel (d)) the final state-resolution procedure. As readily seen in these

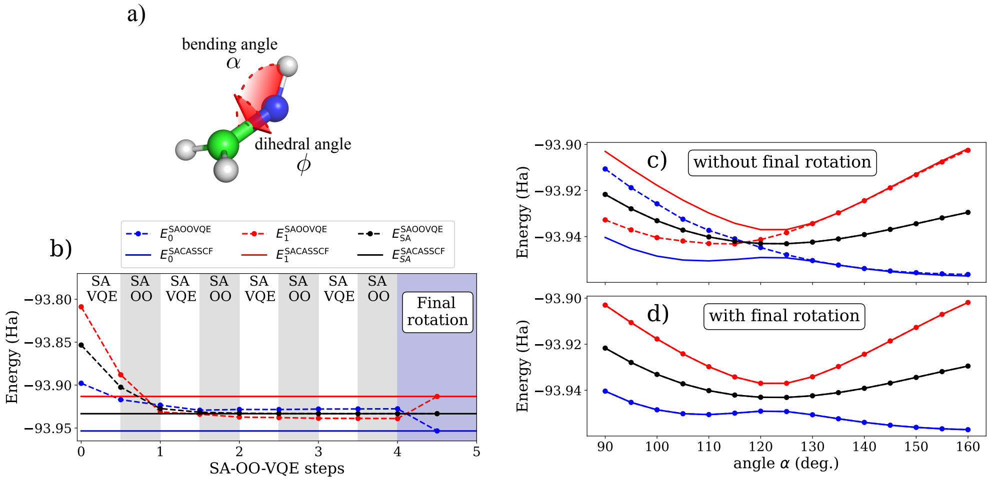


Figure 4.2: Illustration of the convergence of each step of the equi-ensemble SA-OO-VQE algorithm. In all panels, dashed and solid lines represent the SA-OO-VQE and SA-CASSCF energies, respectively. **a)** Geometry of the formaldimine molecule. **b)** Evolution of the state-averaged energy during the different steps of the SA-OO-VQE algorithm for $\phi = 85°$ and $\alpha = 100°$. Converged SA-OO-VQE 1D-PES scans along α with $\phi = 85°$ are shown before (**c**) and after (**d**) the state-resolution procedure.

two panels, the state-averaged SA-OO-VQE energy is in very good agreement with the reference state-averaged SA-CASSCF energy all over the 1D-PES (with an error $< 10^{-6}$ Ha). Without state resolution (see panel (c)), the individual-state energies are globally different from the SA-CASSCF ones, especially for $\alpha < 130°$. For $\alpha > 130°$, the energies of the individual states match the reference ones, showing that the state resolution is not always necessary to capture the eigenstates. Interestingly, the SA-OO-VQE states smoothly cross around $\alpha \sim 118°$. As discussed in Ref. 30, this results from the use of an equi-ensemble where no ordering of the trial states is enforced. In such a case, the converged SA-OO-VQE states will naturally evolve to the state that is the closest to its initial state, *i.e.* with the highest overlap with its initial state (see Ref. 30 for more details). Typically, when $\alpha < 118°$ the singlet single-excited CIS state $|\Phi_B\rangle = -\hat{E}_{\text{hl}}\,|\text{HF}\rangle$ has the largest contribution to the SA-CASSCF ground state, while for $\alpha > 132°$ the $|\Phi_A\rangle = |\text{HF}\rangle$ has the largest contribution (and reciprocally for the first-excited SA-CASSCF state). The state-resolution procedure (see panel (d)) will lift the crossing, thus resulting in an avoided-crossing captured by the adiabatic eigenstates and an excellent agreement between the SA-OO-VQE and SA-CASSCF energies.

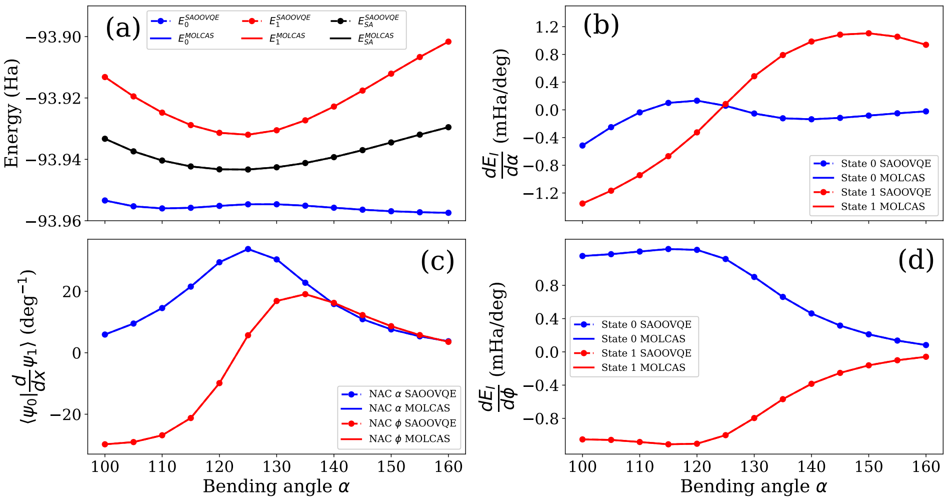


Figure 4.3: Energy gradients and NAC with respect to the bending angle α with $\phi = 80°$. SA-OO-VQE results are shown in dashed lines with dots while solid lines are for the reference SA-CASSCF results. The ground and first-excited state energies are represented by blue and red colors, respectively. **(a)** Potential energy surfaces **(b)** Analytical individual-state energy gradients [Eq. (4.33) for $x = \alpha$] **(c)** non-adiabatic coupling vector $\langle \Psi_0 | (\partial/\partial x) \Psi_1 \rangle$ for $x = \alpha$ (blue) and $x = \phi$ (red) **(d)** Analytical individual-state energy gradients [Eq. (4.33) for $x = \phi$].

4.4.2 Calculation of analytical gradients and non-adiabatic couplings

Let us now turn to the nuclear gradients and non-adiabatic couplings of the adiabatic states obtained after convergence of the SA-OO-VQE algorithm (with state resolution), following Eqs. (4.33) and (4.41). As mentioned in section 4.2.2, the extra cost for the quantum device is in determining the Hessians $\mathcal{H}^{\mathrm{CC}}$ and $\mathcal{H}^{\mathrm{OC}}$. As we have 12 parameters in our simulation, this amounts to 78 and 12 entries of $\mathcal{H}^{\mathrm{CC}}$ and $\mathcal{H}^{\mathrm{OC}}$ to be computed, respectively. With the parameter shift rule, this becomes a total amount of 19968 and 3072 measurements for $\mathcal{H}^{\mathrm{CC}}$ and $\mathcal{H}^{\mathrm{OC}}$, respectively. The rest of the computational work is done on a classical device, given the Hessians and the (transition) 1- and 2-RDM to compute the analytical gradient (non-adiabatic coupling). The results are represented in Figs. 4.3 and 4.4 along the α direction for $\phi = 80°$ and $\phi = 90°$, respectively. As readily seen in Figs. 4.3 and 4.4, both the analytical gradients and the NAC calculated from our SA-OO-VQE implementation cannot be distinguished from the SA-CASSCF results, with a negligible difference of the order of 10^{-3} mHa/degree for the gradients and 10^{-3} degree^{-1} for the NAC amplitudes (all over the PES). This supports the derivations of Eqs. (4.33) and (4.41) and shows

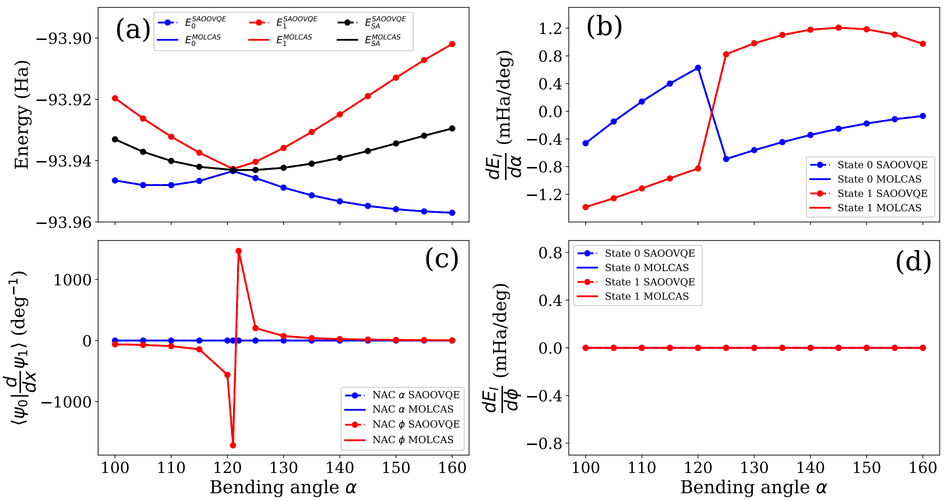


Figure 4.4: Energy gradients and NAC with respect to the bending angle α with $\phi = 90°$. SA-OO-VQE results are shown in dashed lines with dots while solid lines are for the reference SA-CASSCF results. The ground and first-excited state energies are represented by blue and red colors, respectively. **(a)** Potential energy surfaces **(b)** Analytical individual-state energy gradients [Eq. (4.33) for $x = \alpha$] **(c)** non-adiabatic coupling vector $\langle \Psi_0 | (\partial/\partial x) \Psi_1 \rangle$ for $x = \alpha$ (blue) and $x = \phi$ (red) **(d)** Analytical individual-state energy gradients [Eq. (4.33) for $x = \phi$].

that SA-OO-VQE can provide (ideally, *i.e.* without noise) as accurate results as its classical SA-CASSCF analogue. Turning to the energy landscape of Fig. 4.3 (panel (a)) with $\phi = 80°$ (exactly the same as Fig. 4.2, panel (d), but plotted again here for convenience), we observe an avoided crossing between the ground and first-excited states around $\alpha = 125°$. This particular behavior can also be detected by looking at the amplitudes of the NAC (panel (c) of Fig. 4.3), which increase significantly at the avoided crossing position (but without diverging).

In contrast, for $\phi = 90°$ (see panel (a) of Fig. 4.4) the 1D-PES shows a crossing (a conical intersection here). This very different behaviour is manifested by a discontinuity in the gradients $\partial E_0/\partial\alpha$ and $\partial E_1/\partial\alpha$ that suddenly inverse their position at the crossing point $\alpha \approx 121.5°$, while the gradients are smoothly evolving along the 1D-PES when the states do not cross (see panel (b) of Figs. 4.3 and 4.4, respectively). This is the direct consequence of the presence of a degeneracy in the energy profile. In contrast to the α direction (panel (b)), the fact that the gradients $\partial E_0/\partial\phi$ and $\partial E_1/\partial\phi$ are zero for all α (see panel (d) in Fig. 4.4) reveals the presence of extrema for both states in the ϕ direction. This behaviour is consistent with $\phi = 90°$ defining a mirror-plane symmetry σ_v ($\mathcal{C}_\mathrm{s}$ point group) wherein there is no interstate coupling (different irreducible representations) and local extrema are induced for both potential energies at $\phi = 90°$. For a better

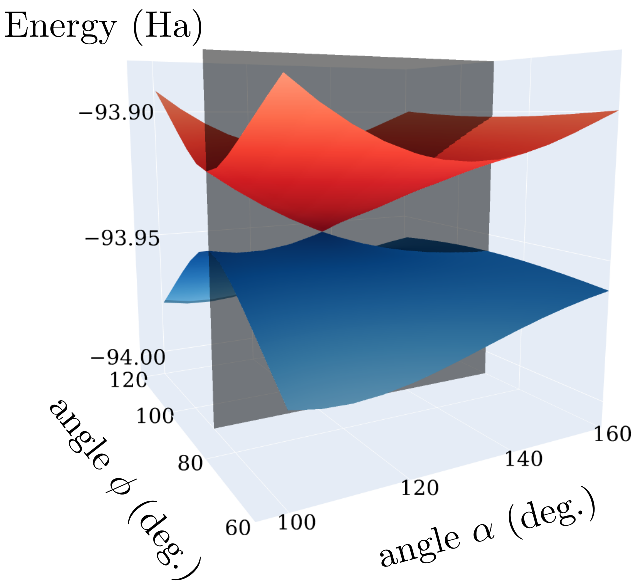


Figure 4.5: (α, ϕ)**-PES for the formaldimine molecule.** The energies are obtained with the SA-OO-VQE algorithm after a full resolution of the states. A conical intersection is observed around the geometry ($\alpha = 121.5°, \phi = 90°$). Grey plane is defined for $\phi = 90°$ and intersects both PESs at extrema values in the ϕ direction.

illustration, these local extrema along the ϕ direction are shown in Fig. 4.5 and correspond to the intersection points between the grey plane (defining $\phi = 90°$) and the two-dimentional (α, ϕ)-PESs computed with SA-OO-VQE.

Considering now the non-adiabatic couplings (panel (c) of Fig. 4.4), we find that it points to a direction perpendicular to the gradients. This is expected, as the NAC together with the gradient of the energy difference span the branching space defined by α and ϕ, as discussed in Sec. 4.2.1. The NAC exhibits an asymptotic discontinuity around the conical intersection, caused by the term $(E_1 - E_0)^{-1}$ which goes to infinity at this degeneracy point. The increase of the NAC amplitude at the avoided crossing or its divergence at the conical intersection is the expected manifestation of its linear dependence on the inverse of the energy difference [see Eq. (4.2)]. This typically results in a break-down of the Born–Oppenheimer approximation, which is consistent with observing radiationless population transfer from the first-excited state back to the ground state.

This will essentially occur around such crossing geometries along the photochemical reaction path. A prototypical example is the photoisomerization process in the retinal chromophore of rhodopsin [256]. In such a situation, quantum dynamics simulations (or their various flavours of quantum-classical approximations) must explicitly account for NAC-terms within the equations of motion for the nuclei

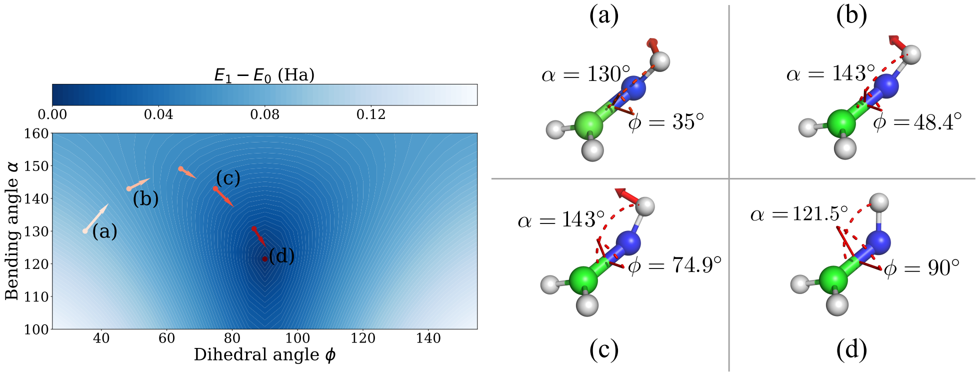


Figure 4.6: (α, ϕ)-Geometry optimization to the conical intersection point. **Left panel:** optimization path of a steepest-descent algorithm to locate the conical intersection of formaldimine. The contourplot shows the energy difference $\Delta E = E_1 - E_0$, and the vectors represent the negative of the gradient of the energy difference $\mathbf{g}_{\Delta E}$ at each point of the optimization. **Right panel:** Four molecular geometries corresponding to four points of the optimization, denoted by (a), (b), (c) and (d) on the left panel. The corresponding (α, ϕ) angles and $\mathbf{g}_{\Delta E}$ vectors are shown.

evolving within a manifold of coupled electronic states.

4.4.3 Geometry optimization to locate formaldimine's conical intersection in the (α, ϕ) space

As an application to the SA-OO-VQE analytical gradients, we perform a geometry optimization inside the (α, ϕ)-plane to find a conical intersection for formaldimine. To do so, we use a steepest-descent optimization algorithm considering as a cost function the energy difference $\Delta E = E_1^{\text{SAOOVQE}} - E_0^{\text{SAOOVQE}}$. At each step of the run, we update the molecular geometry in the (α, ϕ)-plane based on the associated gradient

$$\mathbf{g}_{\Delta E} = \frac{\mathrm{d}E_1^{\text{SAOOVQE}}}{\mathrm{d}\mathbf{x}} - \frac{\mathrm{d}E_0^{\text{SAOOVQE}}}{\mathrm{d}\mathbf{x}} \tag{4.46}$$

which is evaluated using our analytical method described in Sec. 4.2.2 (with $\mathbf{x} = (\alpha, \phi)$).

An example run of the algorithm is shown in Fig. 4.6, where the steepest-descent procedure starts at point (a) for a molecular configuration $(\alpha, \phi) = (130°, 35°)$. The path followed during the geometry optimization is driven by the vector $-\mathbf{g}_{\Delta E}$ which is illustrated at every 4 iterations with arrows on the left panel of Fig. 4.6 (out of a total of 20 iterations required to reach a geometry very near the conical intersection point). The last iteration point is also shown, located exactly on the CI. Similarly, on the right panel of Fig. 4.6 the vector $-\mathbf{g}_{\Delta E}$ is represented with red arrows (starting from the Hydrogen atom) for four molecular geometries

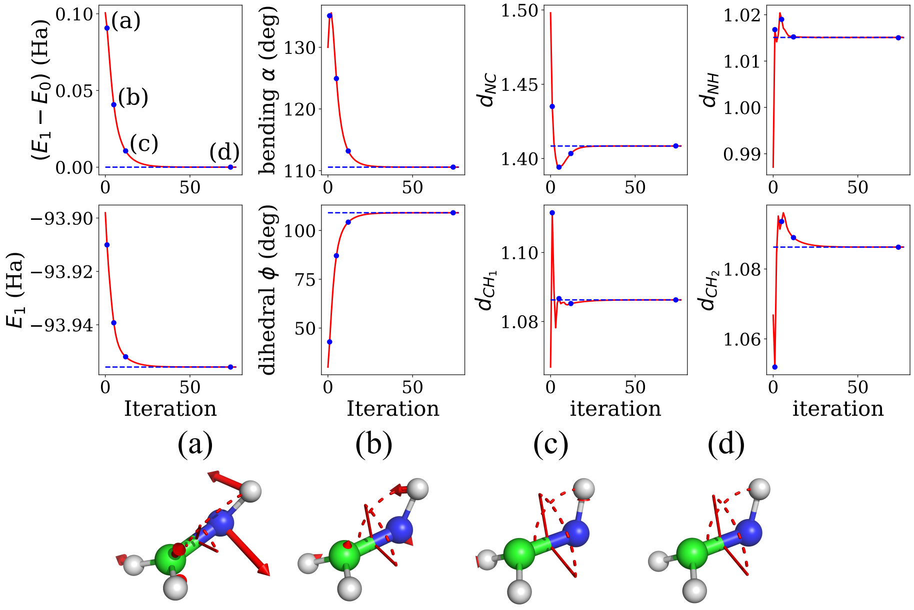


Figure 4.7: Geometry optimization to minimal energy conical intersection (MECI). Upper panel: Energy difference, first-excited state energy, bending angle, dihedral angle, and bond lengths of the molecule are plotted against the number of iterations to convergence of the MECI optimization using the SA-OO-VQE algorithm (red lines). Dashed blue lines indicate the converged values of the MECI optimization using the SA-CASSCF algorithm. **Lower panel:** Four molecular geometries corresponding to four points of the optimization path (blue points on the upper panel). The corresponding gradient vectors $-\mathbf{g}$ are shown for each atom.

(a), (b), (c) and (d) obtained on the path of optimization (also noted on the left panel). From Fig. 4.6, we see that the path stays orthogonal to isolines ($\Delta E = c$ with c constant) of the cost function ΔE. This feature indicates that the SA-OO-VQE algorithm provides consistent pairs of gradients for the two states, thus leading smoothly to the molecular geometry associated to the conical intersection in the (α, ϕ)-plane. At the end of the geometry optimization, the conical intersection is found for a geometry $\phi = 90°$ and $\alpha = 121.47°$. As a comparison, a similar calculation has been realized with the SA-CASSCF method for which a similar path was followed in the (α, ϕ)-plane, thus leading to an equivalent location of the conical intersection with a negligible difference of the final molecular geometry ($\sim 0.001°$ error for both α and ϕ). In practice, note that the non-smooth behavior of the cost function at the conical intersection makes the gradient-descent algorithm hard to converge when approaching this point (as the gradient of the cost function will always have a non-zero component in either direction). As the scope of this chapter is to provide proof-of-principle calculation, we simply stopped the algorithm after a limited number of iterations.

4.4.4 Conical intersection optimization in the full geometry space

The previous section focused on the localization of a conical intersection within the (α, ϕ) 2D-subspace at fixed values of the other internal coordinates. As this subspace is a good description of the branching space, there is a unique point of degeneracy (a conical intersection) within this plane where degeneracy is lifted to first order from it. As such, minimising ΔE, and nothing else, is sufficient, as the complement space is frozen.

In this last section, a step further is realized to characterize the conical intersection of the formaldimine molecule, by determining the so-called *minimal energy conical intersection* (MECI) of the system. In practice, the MECI point corresponds to the optimal geometry of a given system for which the energies of two states get simultaneously degenerated and maximally lowered. Therefore, it encodes the most favorable molecular conformation for the realization of non-radiative photochemical processes but also inter-system (i.e. spin-forbidden) crossings [293, 294]. As a counterpart, the realization of a geometry optimization to precisely determine the MECI position is usually a pretty involved task. The process implies a full relaxation of the internal coordinates of a molecular system, that is driven by the gradients and NAC vectors of each atom.

Using the estimation of NAC and energy gradients with SA-OO-VQE, we realized a geometry optimization to determine the MECI of formaldimine, using the so-called gradient projection method [295–297]. In this algorithm, we simultaneously minimize ΔE^2 – to allow for a smooth minimum – and E_1, using a steepest-descent algorithm where the minimization follows the direction of a composite gradient. To minimize E_1 only in the direction of the seam space, we project out the component of its gradient along the branching space. The composite gradient is:

$$\mathbf{g} = \eta\left(2\Delta E\frac{\mathbf{g}_{\Delta E}}{|\mathbf{g}_{\Delta E}|}\right) + (1-\eta)\mathbf{P}\frac{\mathrm{d}E_1^{\mathrm{SAOOVQE}}}{\mathrm{d}\mathbf{x}}, \tag{4.47}$$

where $\eta \in [0,1]$ is a constant, balancing the two objectives, and

$$\mathbf{P} = \mathbf{1} - \tilde{\mathbf{g}}_{\Delta E}\tilde{\mathbf{g}}_{\Delta E}^{\dagger} - \tilde{\mathbf{h}}_{01}\tilde{\mathbf{h}}_{01}^{\dagger} \tag{4.48}$$

is the projection along the seam space. Here, $\mathbf{h}_{IJ}$ is proportional to the CI-term of the NAC, i.e. the first term in Eq. (4.41). The tilde indicates orthonormalization of $\mathbf{g}_{\Delta E}$ and $\mathbf{h}_{01}$.

Results are laid out in Fig. 4.7. We initialize the algorithm with $\eta = 0.25$ and consider convergence thresholds of 10^{-13} Ha^2 and 10^{-6} Ha for ΔE^2 and

E_1, respectively. As in the previous section, the starting geometry is $(\alpha, \phi) = (130°, 35°)$, but now all the $3N_{\text{atom}}$ nuclear coordinates move in the direction of their respective gradient defined by Eq. (4.47). Our algorithm converges after 78 iterations, where the bending and dihedral angles are $(\alpha, \phi) = (110.6°, 109.1°)$. Note that a dihedral angle of 109.1° here does not break the σ_v mirror symmetry of the molecule, as the hydrogens attached to the carbon bend slightly backwards.

As a reference, in Fig. 4.7 dashed lines are used to represent the final results obtained for a similar MECI optimization realized with the SA-CASSCF method (analytical SA-CASSCF gradients and NAC being evaluated with OpenMolcas to drive the optimization). As readily seen in the different plots, the MECI optimization based of SA-OO-VQE (red lines) converges to the same geometry and energies as in SA-CASSCF, thus showing again the accuracy of our estimations of the NAC and energy gradients within the SA-OO-VQE algorithm.

4.5 Conclusions and perspectives

In this chapter, we introduce several tools to improve our original SA-OO-VQE algorithm [30]. The first improvement consists in introducing a flexible and efficient way to resolve the SA-OO-VQE electronic states. The method, based on equi-ensemble properties, takes advantage of the invariance of the state-averaged energy under rotation of the states involved in the ensemble. Using a simple rotation of the input states (implemented by a short-depth circuit), we show that one can postpone the resolution of the electronic states to the very end of the full SA-OO-VQE scheme, thus avoiding many unnecessary manipulations and quantum measures during intermediate steps of the algorithm. The second improvement is the development of theoretical methods to extract analytical derivatives within the SA-OO-VQE algorithm. These derivatives – the nuclear energy gradients and non-adiabatic couplings – are fundamental for the study of molecular systems, and can be determined using Lagrangian methods that are intimately linked to the so-called coupled-perturbed theory. The accuracy of our derivations is checked against reference results based on SA-CASSCF calculations, for which we obtain very good agreement. Finally, we illustrate the use of these derivatives in practical calculations by performing the geometry optimization towards the conical intersection of the formaldimine molecule. The localization of the spectral degeneracy matches perfectly the predictions from the SA-CASSCF method.

The definition of these new tools opens the way to several new developments. One aspect that we already briefly touched upon above, is the definition of the intermediate diabatic and final adiabatic bases. In our implementation of the SA-OO-VQE algorithm, the procedure starts from reference guess states and has no reason to produce "excessive" transformations, thus making the least-transformed subspace a good candidate for being a quasidiabatic representation (see Ref. 298). At the moment, we have been observing such a property, and we have good incentive but no formal proof. Further work is under way to show that SA-OO-VQE before its final diagonalization (or state-resolution) could indeed be an efficient avenue for the *ab initio* production of relevant quasidiabatic states. Such results would be important as it would facilitate use in molecular quantum dynamics applications. With an appropriate definition of such states and construction of a quantum-classical interface, these tools can benefit from classical implementations [299] of algorithms to perform various forms of molecular dynamics [300, 301], thus giving quantum co-processing a firm place in the toolbox of quantum chemical simulations. A related aspect is to consider different kinds of surface couplings, like those provided by the spin-orbit operator. This is of interest for a range of applications, e.g. the rate of inter-system crossing that is key to technological applications, as the construction of more efficient blue light emitting diodes. Here, we note that it is possible to work with a different set of trial states than the Hartree–Fock and the singlet excited model wave functions chosen in the current work. The latter can be replaced by a triplet excited wave function, while the former can also be a non-ground state determinant (or other simple wave function) if we are interested in excited-state couplings.

4.A Circuit gradient $\mathcal{G}^{\mathrm{C}}$ and Hessian $\mathcal{H}^{\mathrm{CC}}$

Let us consider the estimation of the expectation value M of a generic operator $\hat{\mathcal{M}}$ (in our case the electronic structure Hamiltonian H) with respect to a state $|\Psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta})\,|\Phi_0\rangle$,

$$M(\boldsymbol{\theta}) = \mathrm{Tr}\left[\hat{\mathcal{M}}\rho(\boldsymbol{\theta})\right] = \langle\Psi(\boldsymbol{\theta})|\,\hat{\mathcal{M}}\,|\Psi(\boldsymbol{\theta})\rangle\,, \tag{4.49}$$

where $\rho(\boldsymbol{\theta}) = |\Psi(\boldsymbol{\theta})\rangle\langle\Psi(\boldsymbol{\theta})|$ is the density matrix operator and $\mathrm{Tr}\,[\cdot]$ the trace operation. These matrix elements encode the first order (gradient $\boldsymbol{\mathcal{G}}^{\mathrm{C}}$) and second order (Hessian $\mathcal{H}^{\mathrm{CC}}$) derivatives with respect to the ansatze parameters $\boldsymbol{\theta}$ read as follows

$$\mathcal{G}^{\mathrm{C}}_i = \frac{\partial M(\boldsymbol{\theta})}{\partial\theta_i} \quad \text{and} \quad \mathcal{H}^{\mathrm{CC}}_{ij} = \frac{\partial^2 M(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}, \tag{4.50}$$

and can be evaluated with the parameter-shift rule [66, 285, 302–305].

As a starting point, let us consider the following unitary:

$$\begin{aligned} U(\boldsymbol{\theta}) &= U_1(\theta_1)\times U_2(\theta_2) \\ &= (e^{-i\frac{\theta_1}{2}P_1}e^{-i\frac{\theta_1}{2}\tilde{P}_1})\times(e^{-i\frac{\theta_2}{2}P_2}e^{-i\frac{\theta_2}{2}\tilde{P}_2}) \end{aligned} \tag{4.51}$$

where $[P_1,\tilde{P}_1] = [P_2,\tilde{P}_2] = 0$ with the tilde notation denoting a different Pauli string with the same associated parameters, as this is usually the case in the fermionic-UCC ansatz. We have $P_j = P_j^\dagger$ and

$$\frac{\partial U_j(\theta_j)}{\partial \theta_j} = -\frac{i}{2}(P_j + \tilde{P}_j)U_j(\theta_j) \tag{4.52}$$

and

$$\left(\frac{\partial U_j(\theta_j)}{\partial \theta_j}\right)^\dagger = -\frac{\partial U_j(\theta_j)}{\partial \theta_j}. \tag{4.53}$$

We want to estimate the gradient elements

$$\mathcal{G}_i^{\mathrm{C}} = \frac{\partial}{\partial \theta_i} \mathrm{Tr}\left[\hat{\mathcal{M}}\rho(\boldsymbol{\theta})\right]. \tag{4.54}$$

By taking the derivative with respect to θ_2, we get

$$\begin{aligned}\mathcal{G}_2^{\mathrm{C}} &= \mathrm{Tr}\left[\frac{\partial}{\partial \theta_2}\hat{\mathcal{M}}\rho(\boldsymbol{\theta})\right] \\ &= \hat{\mathcal{M}}U_1(\theta_1)\left[\frac{\partial U_2(\theta_2)}{\partial \theta_2}\rho U_2^\dagger(\theta_2)\right. \\ &\left. + U_2(\theta_2)\rho\left(\frac{\partial U_2(\theta_2)}{\partial \theta_2}\right)^\dagger\right] U_1^\dagger(\theta_1) \\ &= -\frac{i}{2}\hat{\mathcal{M}}U_1(\theta_1)U_2(\theta_2)\left[(P_2 + \tilde{P}_2), \rho\right] U_2^\dagger(\theta_2)U_1^\dagger(\theta_1)\end{aligned} \tag{4.55}$$

where we used the property $[P_i, U_i(\theta_i)] = 0$ and the notation $\rho = |\Phi_0\rangle\langle\Phi_0|$. We then use the property of commutator for an arbitrary operator [see Eq. (2) in Ref. 285, where $U_j(\theta_j) = \exp(-i\theta_j P_j/2)$],

$$[P_j, \rho] = i\left[U_j\left(\frac{\pi}{2}\right)\rho U_j^\dagger\left(\frac{\pi}{2}\right) - U_j\left(-\frac{\pi}{2}\right)\rho U_j^\dagger\left(-\frac{\pi}{2}\right)\right], \tag{4.56}$$

which can be easily demonstrated by considering the property of exponential of Pauli strings

$$e^{i\frac{\theta_j}{2}P_j} = \cos\left(\frac{\theta_j}{2}\right)\mathbf{1} - i\sin\left(\frac{\theta_j}{2}\right)P_j. \tag{4.57}$$

One then separates $(P_j + \tilde{P}_j)$ as follows:

$$[(P_j + \tilde{P}_j), \rho] = [P_j, \rho] + [\tilde{P}_j, \rho] \tag{4.58}$$

such that, by inserting Eq. (4.56) into Eq. (4.55), and subsequently taking the trace, the final expression is:

$$\mathcal{G}_2^{\mathrm{C}} = \frac{1}{2}\left(\left\langle\hat{\mathcal{M}}\right\rangle_{\theta_2^+} - \left\langle\hat{\mathcal{M}}\right\rangle_{\theta_2^-} + \left\langle\hat{\mathcal{M}}\right\rangle_{\tilde{\theta}_2^+} - \left\langle\hat{\mathcal{M}}\right\rangle_{\tilde{\theta}_2^-}\right) \tag{4.59}$$

where the notation $\left\langle\hat{\mathcal{M}}\right\rangle_{\theta_j^\pm}$ refers to the expectation value of the operator $\hat{\mathcal{M}}$ when θ_j has been shifted by $\pm\pi/2$ in front of P_j (and in front of $\tilde{P}_j$ for $\left\langle\hat{\mathcal{M}}\right\rangle_{\tilde{\theta}_j^\pm}$).

Generalizing Eq. (4.59) to any n-fold fermionic excitation generator $\mathcal{G}_j$ associated to the parameter θ_j leads to the parameter-shift rule:

$$\mathcal{G}_j^{\mathrm{C}} = \frac{1}{2}\sum_n^{\forall P_n \in \mathcal{G}_j}\left(\left\langle\hat{\mathcal{M}}\right\rangle_{\theta_{jn}^+} - \left\langle\hat{\mathcal{M}}\right\rangle_{\theta_{jn}^-}\right) \tag{4.60}$$

where θ_{j_n} refers to the parameter associated to the Pauli string P_n coming from the fermionic generator $\mathcal{G}_j$. Note that the parameter-shift rule only applies to generators $\mathcal{G}$ that have at most two distinct eigenvalues [66], which is always the case for any Pauli string but not a linear combination of them. However, each parameter θ_j is associated to two Pauli strings for a single-excitation fermionic operator, to eight Pauli strings for a double-excitation fermionic operator, and to 2^{2n-1} Pauli strings for a n-fold fermionic excitation operator [306, 307], such that

$$e^{\mathrm{i}\frac{\theta}{2}\mathcal{G}} \to \prod_{x=1}^{2^{2n-1}} e^{\mathrm{i}\frac{\theta}{2}P_x}. \tag{4.61}$$

The above formula is actually an equality because Pauli strings resulting from a same n-fold fermionic excitation operator actually commute with each other [306]. Although the fermionic generator $\mathcal{G}$ usually doesn't have two distinct eigenvalues but three [307], they can be decomposed into generators that have only two distinct eigenvalues (for instance, Pauli strings P_x with eigenvalues ±1) and the gradient can be directly calculated by the product rule and the parameter-shift rule [65, 307], necessitating 2^{2n} expectation values. So even for a UCCD ansatz, we would need around $2^4 = 16$ expectation values for a single gradient calculation (to be multiplied by the number of parameters).

Turning to the Hessian estimation, one can derive Eq. (4.60) with respect to another parameter θ_k,

$$\mathcal{H}_{kj}^{\mathrm{CC}} = \frac{1}{2}\sum_n^{\forall P_n \in \mathcal{G}_j}\left(\frac{\partial}{\partial\theta_k}\left\langle\hat{\mathcal{M}}\right\rangle_{\theta_{jn}^+} - \frac{\partial}{\partial\theta_k}\left\langle\hat{\mathcal{M}}\right\rangle_{\theta_{jn}^-}\right) \tag{4.62}$$

and use the parameter-shift rule again, thus leading to [303]

$$\mathcal{H}^{\mathrm{CC}}_{kj} = \frac{1}{4} \sum_{n}^{\forall P_n \in \mathcal{G}_j} \sum_{m}^{\forall P_m \in \mathcal{G}_k} \Big(\left\langle \hat{\mathcal{M}} \right\rangle_{\theta^+_{j_n} \theta^+_{k_m}} - \left\langle \hat{\mathcal{M}} \right\rangle_{\theta^+_{j_n} \theta^-_{k_m}} - \left\langle \hat{\mathcal{M}} \right\rangle_{\theta^-_{j_n} \theta^+_{k_m}} + \left\langle \hat{\mathcal{M}} \right\rangle_{\theta^-_{j_n} \theta^-_{k_m}} \Big). \tag{4.63}$$

According to Eq. (4.63), a single element of the Hessian will require the estimation of 2^{4n} expectation values for a n-fold fermionic excitation operator.

Different strategies have recently been developed to reduce the number of expectation values required to evaluate a ansatze-parameter gradient of a fermionic generator $\mathcal{G}$ with more than two distinct eigenvalues. One can consider an additional ancilla qubit and decompose the derivative into a linear combination of unitaries [66], use stochastic strategies [308, 309] or different generator decomposition techniques [307, 310].

4.B Nuclear derivative of the electronic Hamiltonian operator

In the coupled-perturbed equations, one needs the derivative of the Hamiltonian operator with respect to a nuclear coordinate [267, 284, 311, 312] which is defined by

$$\frac{\partial H}{\partial x} = \sum_{p,q} \frac{\partial h_{pq}}{\partial x} \hat{E}_{pq} + \frac{1}{2} \sum_{p,q,r,s} \frac{\partial g_{pqrs}}{\partial x} \hat{e}_{pqrs} + \frac{\partial E_{nuc}}{\partial x}, \tag{4.64}$$

where the derivative of the electronic integrals are

$$\frac{\partial h_{pq}}{\partial x} = h^{(x)}_{pq} - \frac{1}{2} \left\{ S^{(x)}, h \right\}_{pq} \tag{4.65}$$

$$\frac{\partial g_{pqrs}}{\partial x} = g^{(x)}_{pqrs} - \frac{1}{2} \left\{ S^{(x)}, g \right\}_{pqrs} \tag{4.66}$$

where we retrieve ‘explicit’ and ‘response’ terms with respect to a nuclear coordinate. The explicit terms are the ones super-scripted with $^{(x)}$ indicating a differentiation of the primitive atomic orbitals (MOs coefficients remaining constant). They are defined such as

$$S^{(x)}_{pq} = \sum_{\mu,\nu}^{\mathrm{AOs}} C_{\mu p} C_{\nu q} \frac{\partial S_{\mu\nu}}{\partial x} \tag{4.67}$$

$$h_{pq}^{(x)} = \sum_{\mu,\nu}^{\text{AOs}} C_{\mu p} C_{\nu q} \frac{\partial h_{\mu\nu}}{\partial x} \tag{4.68}$$

$$g_{pqrs}^{(x)} = \sum_{\mu,\nu,\delta,\gamma}^{\text{AOs}} C_{\mu p} C_{\nu q} C_{\delta p} C_{\gamma q} \frac{\partial g_{\mu\nu\delta\gamma}}{\partial x} \tag{4.69}$$

where **C** is the MO coefficient matrix encoding the optimal orbitals that minimize the state-averaged energy. The 'response' terms in curly brackets are defined as

$$\left\{ S^{(x)}, h \right\}_{pq} = \sum_o (S_{po}^{(x)} h_{oq} + S_{qo}^{(x)} h_{po}) \tag{4.70}$$

$$\left\{ S^{(x)}, g \right\}_{pqrs} = \sum_o \Big(S_{po}^{(x)} g_{oqrs} + S_{qo}^{(x)} g_{pors} \tag{4.71}$$

$$+ S_{ro}^{(x)} g_{pqos} + S_{so}^{(x)} g_{pqro} \Big). \tag{4.72}$$

The last term present on the right of Eq. (4.64) is the nuclear derivative of the nuclear repulsion energy which is pretty straightforward to compute in practice.

4.C Analytical derivation of non-adiabatic couplings for SA-OO-VQE

In this section, we introduce the steps to derive the analytical form of Eqs. (4.40) and (4.41) which define the NAC between two states $|\Psi_I\rangle$ and $|\Psi_J\rangle$. Following Ref. 268, one splits the complete derivative in the NAC into two contributions,

$$D_{IJ} = \left\langle \Psi_I \middle| \frac{d}{dx} \Psi_J \right\rangle = \left\langle \Psi_I \middle| \frac{\partial}{\partial x} \Psi_J \right\rangle + D_{IJ}^{\text{CSF}}. \tag{4.73}$$

The first term represents the so-called CI term, and the second one the CSF term (see Ref. 269) that does not appear in the exact theory [see Eq. (4.1)]. The CSF term is readily computed as:

$$D_{IJ}^{\text{CSF}} = -\frac{1}{2} \sum_{pq} \gamma_{pq}^{IJ} \big((\partial_x p|q) - (q|\partial_x p) \big). \tag{4.74}$$

The CI term, however, is more involved. To evaluate this term, we will make use of the off-diagonal Hellmann–Feynman theorem:

$$\left\langle \Psi_I \middle| \frac{\partial}{\partial x} \Psi_J \right\rangle = \Delta E^{-1} \langle \Psi_I | \frac{\partial H}{\partial x} | \Psi_J \rangle . \tag{4.75}$$

Eq. (4.75) is valid if two conditions are met. The first one is that the SA-OO-VQE states $|\Psi_I\rangle$ and $|\Psi_J\rangle$ are good approximations of the exact eigenstates of H (to some negligible errors, which is verified numerically in our work). Second, the NAC has to be variational with respect to the orbital rotation parameters $\boldsymbol{\kappa}$, the ansatze parameters $\boldsymbol{\theta}$ and the final rotation φ implemented for the state resolution. While the SA-OO-VQE states do not satisfy this condition, one can still make the NAC variational with respect to these parameters by introducing the following Lagrangian:

$$\begin{aligned} L_{IJ} &= \overline{S}_{IJ} \\ &+ \frac{1}{\overline{\Delta E}}\left(\sum_{p,q}\overline{\kappa}_{pq}^{IJ}\frac{\partial E_{\text{SA}}}{\partial \kappa_{pq}} + \sum_{n}\overline{\theta}_{n}^{IJ}\frac{\partial E_{\text{SA}}}{\partial \theta_n} + \overline{\varphi}^{IJ}\frac{\partial \Delta E}{\partial \varphi}\right), \end{aligned} \tag{4.76}$$

where $\overline{S}_{IJ} = \left\langle\overline{\Psi_I}\middle|\Psi_J\right\rangle$ is the overlap between the two states, and the leftmost state is kept constant (as denoted by an overbar) because we *only* want to take derivative of the right state in the NAC. Compared to the gradient Lagrangian [Eq. (4.25)], note the presence of the convergence condition $\partial\Delta E/\partial\varphi = 0$ encapsulating the effect of the final state resolution (with $\Delta E \equiv E_J - E_I$). In practice, this condition holds as the final rotation consists in minimizing a given state energy (which is equivalent to maximizing the difference between both individual-state energies). This convergence condition was not needed in the analytical gradient Lagrangian in Eq. (4.25) as, in contrast to the NAC, the individual-state energies are already variational with respect to φ. The factor $\overline{\Delta E}$ in Eq. (4.76) was introduced for convenience (with the overbar meaning that the energy difference is kept constant).

Now, one has to find the Lagrangian multipliers in Eq. (4.76) such that the Lagrangian is fully variational with respect to $\boldsymbol{\kappa}$, $\boldsymbol{\theta}$ and φ,

$$\frac{\partial L_{IJ}}{\partial \kappa_{pq}} = \frac{\partial L_{IJ}}{\partial \theta_n} = \frac{\partial L_{IJ}}{\partial \varphi} = 0. \tag{4.77}$$

From Eq. (4.18), we have

$$\frac{\partial \Delta E}{\partial \varphi} = -4\mathcal{H}_{IJ}, \quad \text{and} \quad \frac{\partial \mathcal{H}_{IJ}}{\partial \varphi} = \Delta E, \tag{4.78}$$

where $\mathcal{H}_{IJ} = \langle\Psi_I|H|\Psi_J\rangle$ (we assume that H is a real operator). We also introduce the two non-zero derivatives of the overlap $\overline{S}_{IJ}$ that read

$$\left\langle\Psi_I\middle|\frac{\partial}{\partial\theta_n}\Psi_J\right\rangle \neq 0, \quad \text{and} \quad \left\langle\Psi_I\middle|\frac{\partial}{\partial\varphi}\Psi_J\right\rangle = -1. \tag{4.79}$$

From these simple relations, one obtains the $\overline{\varphi}^{IJ}$ multiplier as

$$\frac{\partial L_{IJ}}{\partial \varphi} = 0 = -1 - 4\overline{\varphi}^{IJ} \longrightarrow \overline{\varphi}^{IJ} = -\frac{1}{4}. \tag{4.80}$$

Combining this result with Eq. (4.77) and Eq. (4.78) provides the other stationary equations for the orbital parameters,

$$\sum_{rs} \overline{\kappa}^{IJ}_{rs} \mathcal{H}^{\mathrm{CC}}_{pq,rs} + \sum_{n} \overline{\theta}^{IJ}_{n} H^{\mathrm{OC}}_{pq,n} + \langle \Psi_I | \frac{\partial H}{\partial \kappa_{pq}} | \Psi_J \rangle = 0, \tag{4.81}$$

and for the ansatze parameters,

$$\begin{aligned} &\Delta E \left\langle \Psi_I \middle| \frac{\partial}{\partial \theta_n} \Psi_J \right\rangle + \sum_{pq} \overline{\kappa}^{IJ}_{pq} \mathcal{H}^{\mathrm{CC}}_{n,pq} \\ &+ \sum_{m} \overline{\theta}^{IJ}_{m} \mathcal{H}^{\mathrm{CC}}_{n,m} + \frac{\partial \mathcal{H}_{IJ}}{\partial \theta_n} = 0, \end{aligned} \tag{4.82}$$

where we multiplied both sides by ΔE. Let us define the orbital and circuit gradients, respectively, as follows:

$$G^{\mathrm{O},IJ}_{pq} := \langle \Psi_I | \frac{\partial H}{\partial \kappa_{pq}} | \Psi_J \rangle \tag{4.83}$$

$$G^{\mathrm{C},IJ}_{n} := \Delta E \left\langle \Psi_I \middle| \frac{\partial}{\partial \theta_n} \Psi_J \right\rangle + \frac{\partial \mathcal{H}_{IJ}}{\partial \theta_n} = 0. \tag{4.84}$$

Note that the circuit gradient can actually be set to 0. Indeed, we have

$$\begin{aligned} \frac{\partial \mathcal{H}_{IJ}}{\partial \theta_n} &= \left\langle \frac{\partial}{\partial \theta_n} \Psi_I \middle| H \middle| \Psi_J \right\rangle + 0 + \left\langle \Psi_I \middle| H \middle| \frac{\partial}{\partial \theta_n} \Psi_J \right\rangle \\ &= -\Delta E \left\langle \Psi_I \middle| \frac{\partial}{\partial \theta_n} \Psi_J \right\rangle, \end{aligned} \tag{4.85}$$

such that the coupled-perturbed equations read

$$\begin{pmatrix} \mathcal{H}^{\mathrm{OO}} & \mathcal{H}^{\mathrm{OC}} \\ \mathcal{H}^{\mathrm{CO}} & \mathcal{H}^{\mathrm{CC}} \end{pmatrix} \begin{pmatrix} \overline{\boldsymbol{\kappa}} \\ \overline{\boldsymbol{\theta}} \end{pmatrix} = - \begin{pmatrix} \mathbf{G}^{\mathrm{O},IJ} \\ 0 \end{pmatrix}. \tag{4.86}$$

The final form of the CI term of the NAC can now be written as follows,

$$D^{\mathrm{CI}}_{IJ} = \frac{\partial L_{IJ}}{\partial x} = (\Delta E)^{-1} \Bigg(\langle \Psi_I | \frac{\partial H}{\partial x} | \Psi_J \rangle +$$

$$\sum_{p,q} \overline{\kappa}_{pq}^{IJ} \frac{\partial^2 E_{\mathrm{SA}}}{\partial \kappa_{pq} \partial x} + \sum_{n} \overline{\theta}_{n}^{IJ} \frac{\partial^2 E_{\mathrm{SA}}}{\partial \theta_n \partial x} \Bigg), \tag{4.87}$$

where the last term has been set to zero, similarly as in Eq. (4.85):

$$\begin{aligned} \overline{\varphi} \frac{\partial^2 \Delta E}{\partial \varphi \partial x} &= \frac{\partial}{\partial x} \langle \Psi_I | H | \Psi_J \rangle \\ &= -\Delta E \left\langle \Psi_I \middle| \frac{\partial \Psi_J}{\partial x} \right\rangle + \Delta E \left\langle \Psi_I \middle| \frac{\partial \Psi_J}{\partial x} \right\rangle = 0. \end{aligned} \tag{4.88}$$

Similar to the analytical gradient calculation, we end up with:

$$\begin{aligned} D_{IJ} = \frac{1}{E_J - E_I} \Bigg(&\sum_{pq} \frac{\partial h_{pq}}{\partial x} \gamma_{pq}^{IJ,\mathrm{eff}} + \frac{1}{2} \sum_{pqrs} \frac{\partial g_{pqrs}}{\partial x} \Gamma_{pqrs}^{IJ,\mathrm{eff}} \\ &+ \sum_{K} \sum_{n} w_K \overline{\theta}_n^{IJ} \mathcal{G}_n^{\mathrm{C},K} (\tfrac{\partial H}{\partial x}) \Bigg) \\ &- \frac{1}{2} \sum_{pq} \gamma_{pq}^{IJ} \big((\partial_x p | q) - (q | \partial_x p) \big), \end{aligned} \tag{4.89}$$

where the effective transition 1- and 2-RDMs read

$$\boldsymbol{\gamma}^{IJ,\mathrm{eff}} = \boldsymbol{\gamma}^{IJ} + \tilde{\boldsymbol{\gamma}}^{IJ,\mathrm{SA}} \tag{4.90}$$

$$\boldsymbol{\Gamma}^{IJ,\mathrm{eff}} = \boldsymbol{\Gamma}^{IJ} + \tilde{\boldsymbol{\Gamma}}^{IJ,\mathrm{SA}}, \tag{4.91}$$

where $\gamma_{pq}^{IJ} = \langle \Psi_I | \hat{E}_{pq} | \Psi_J \rangle$ and $\Gamma_{pqrs}^{IJ} = \langle \Psi_I | \hat{e}_{pqrs} | \Psi_J \rangle$ are the transition 1- and 2-RDMs, and

$$\tilde{\gamma}_{pq}^{IJ,\mathrm{SA}} = \sum_{o} \big(\gamma_{oq}^{\mathrm{SA}} \overline{\kappa}_{op}^{IJ} + \gamma_{po}^{\mathrm{SA}} \overline{\kappa}_{oq}^{IJ} \big) \tag{4.92}$$

$$\begin{aligned} \tilde{\Gamma}_{pqrs}^{IJ,\mathrm{SA}} = \sum_{o} \big(&\Gamma_{oqrs}^{\mathrm{SA}} \overline{\kappa}_{op}^{IJ} + \Gamma_{pors}^{\mathrm{SA}} \overline{\kappa}_{oq}^{IJ} \\ &+ \Gamma_{pqos}^{\mathrm{SA}} \overline{\kappa}_{or}^{IJ} + \Gamma_{pqro}^{\mathrm{SA}} \overline{\kappa}_{os}^{IJ} \big) \end{aligned} \tag{4.93}$$

are the state-averaged 1- and 2-RDMs (encoding orbital contributions).

CHAPTER 5

A hybrid quantum algorithm to detect conical intersections

5.1 Introduction

Conical intersections (CI) are degeneracy points in the Born-Oppenheimer molecular structure Hamiltonians, where two potential energy surfaces cross. Similar to Dirac cones in graphene [313], these intersections are protected by symmetries of the Hamiltonian which guarantee that any loop in parameter space around a conical intersection has a quantized Berry phase [314]. CIs play an important role in photochemistry [315, 316], as they mediate reactions such as photoisomerization and non-radiative relaxation, which are key steps in processes such as vision [317] and photosynthesis [318]. Therefore, detecting the presence and resolving the properties of CIs is important for computing reaction and branching rates in photochemical reactions [319, 320]. Nevertheless, the study of such processes requires electronic structure methods capable of accurately modelling both the shape and the relative energies of the two intersecting potential energy surfaces, a requirement that poses challenges for the current available methods [321]. Given the need to develop novel methods for identifying and characterizing CIs, quantum computers present themselves as a highly promising option for this task.

Quantum computing has long been driven by the desire to simulate interacting physical systems, such as molecules, as a novel means of investigating their properties [69, 322]. This is typically achieved by preparing eigenstates of molecular Hamiltonians in quantum devices which can natively store and process quantum states. This task would otherwise require an exponentially-scaling classical memory. Recently, with the first noisy and intermediate-scale quantum devices (NISQ) being built, it became increasingly important to research tailored and robust algorithms that minimize the quantum device requirements [53]. Variational quantum algorithms (VQA), such as the variational quantum eigensolver [54, 91] (VQE) and its variations, caught the spotlight in this context, as they allow to prepare and measure quantum states with circuits of relatively low depth. The key feature of VQAs is the repeated execution of short parameterized quantum circuits on the quantum device, from which measurement results are sampled. These results are used to estimate a cost function, which is then minimized by varying the parameters defining the gates of the quantum circuit. Due to the noise introduced by sampling, a relatively large number of circuit runs and measurements are typically needed to estimate the cost function accurately. In chemistry, where VQEs are often proposed as a method to resolve ground state energies to high accuracy, the number of required samples to achieve such accuracy can become prohibitively large [323]. Furthermore, the convergence of the cost function to an optimum is typically only suggested heuristically, and it is proven to be problematic in some cases that lack such heuristic structure [324]. Therefore, it is compelling to suggest VQAs that can access quantities that are less reliant on the precision of both the optimization process and the measurement procedure.

A promising target for VQAs is the computation of the Berry phase $\Pi_{\mathcal{C}}$, which can be used to resolve the existence of CIs. More specifically, $\Pi_{\mathcal{C}}$ is defined as the geometric phase acquired by an eigenstate of a parameterized Hamiltonian over a closed adiabatic path $\mathcal{C}$ in parameter space [314]. Most importantly, it is known that in the presence of certain symmetries, the Berry phase will be quantized to values 0 or π. This quantization is exactly what makes the Berry phase an attractive target for a VQAs, as it implies the final result of the computation need only be accurate to error $< \frac{\pi}{2}$. Quantum algorithms to compute Berry phases have been already proposed, both variational [286, 325] and Hamiltonian-evolution based [326]. Moreover, the long-known effects of Berry phase on nuclear dynamics around a conical intersection [327–329] have been explored recently in analog quantum simulation experiments [330–332]. Nevertheless, previous proposals did not attempt to detect CIs in realistic quantum chemistry problems with an efficient algorithm that can be run on NISQ devices.

In this chapter, we propose a hybrid quantum algorithm to compute the quantized Berry phase for ground states of a family of parameterized real Hamiltonians. We focus on the specific application to molecular Hamiltonians, where we can identify a conical intersection by measuring the Berry phase along a loop in atomic coordinates space. We first review the definition of CIs and Berry phases in Sec. 5.2. Then, we present all the ingredients of the proposed algorithm in Sec. 5.3, which is similar to a VQA in spirit, but it does not require full optimization; rather, the variational parameters are updated by a single Newton-Raphson step for each molecular geometry along a discretization of the loop. In Sec. 5.4, we prove a convergence guarantee for the algorithm under certain assumptions on the ansatz, by providing sufficient condition bounds on the total number of steps and the acceptable sampling noise. Finally, we adapt our algorithm to a specific ansatz in Sec. 5.5 and we benchmark it on a model of the formaldimine molecule $H_2C{=}NH$ in Sec. 5.6. Section 5.7 presents our conclusions, a discussion of potential application cases for our algorithm and an outlook on possible enhancements.

The core code developed for the numerical benchmarks, which provides a flexible implementation of an orbital-optimized variational quantum ansatz, is made available in a GitHub repository [333].

5.2 Background

5.2.1 Conical intersections

Let us consider a molecular electronic structure Hamiltonian $H(\mathbf{R})$ parameterized by the nuclear geometry $\mathbf{R}$ in some configuration space $\mathcal{R}$. A conical intersection is a point $\mathbf{R}^\times \in \mathcal{R}$ where two potential energy surfaces become degenerate, leading to non-perturbatively large non-adiabatic couplings, and thus a breakdown of the Born-Oppenheimer approximation [334, 335]. Conical intersections extend to a manifold of dimension $\dim[\mathcal{R}] - 2$, and lead to the two potential energy surfaces taking the form of cones in the remaining two directions $\hat{x}$, $\hat{z}$. These two potential energy surfaces can be described as eigenstates of the effective Hamiltonian

$$H^{\mathrm{eff}}(\mathbf{R}) = h_x[\mathbf{R} - \mathbf{R}^\times]_x \sigma_x + h_z[\mathbf{R} - \mathbf{R}^\times]_z \sigma_z. \tag{5.1}$$

The Pauli terms σ_x and σ_z form a complete basis for two-dimensional real symmetric matrices. Thus, a single conical intersection cannot be lifted by any real-valued and continuous perturbation of $H^{\mathrm{eff}}(\mathbf{R})$; such a perturbation would only shift the value of $\mathbf{R}^\times$. Moreover, the presence of such an effective Hamiltonian implies that in any direction other than R_x or R_z the energies must be degenerate.

Simulations involving conical intersections are challenging due to the degeneracy of the two states involved, as the character of both states needs to be considered. Active space methods are often used for organic molecules, which involve selecting the chemical bonds that are formed or broken in the reaction pathway as well as the most significant spectator or correlating orbitals [21]. The situation is more complicated when transition metals are involved because the close energetic spacing of d-orbitals typically requires including all five d-orbitals of such a metal into the active space. An additional complication arises if the two crossing states correspond to different atomic configurations: a situation that is not uncommon for the early or late d- (or f-) metals, for which configurations with a different d- (f-) population are energetically close. In such cases, one may need to either work with non-orthogonal orbitals [336] or add an additional d-shell to the active space [337] to qualitatively describe the nature of both states. For cases of practical interest in which one wants to characterize and simulate the internal conversion processes in a complex photo-excited system, the presence of transition metals may easily lead to large active space requirements. These can not be met by classical algorithms and would be highly challenging for quantum algorithms as well.

One way to reduce the complexity of the problem is to first focus on the presence or absence of conical intersections that connect the ground and excited states. The measurement of the Berry phase in chemical systems allows this: without explicitly computing the excited state surface and non-adiabatic couplings, it should be possible to detect whether a loop in the nuclear coordinate space encloses a conical intersection or not. In this manner, one may alleviate the requirements for the active space selection and orbital optimization and quickly establish the region in the potential energy surface that contains an intersection with another surface and needs to be scrutinized further [338]. Information about the location of conical intersections is of interest also for ground-state dynamics; the CIs and the Berry phase they influence the propagation of nuclear wave packets on the adiabatic ground state surface and thereby affect the branching rates and efficiency of reactions or isomerizations [339]. For both types of applications, precise study of dynamics on ground state surfaces as well as characterizing the efficiency of radiationless decay, it is of interest to explore the possibilities offered by quantum algorithms.

5.2.2 Berry phases in real Hamiltonians

The Berry phase $\Pi_{\mathcal{C}}$ is the geometric phase acquired by some eigenstate $|\Phi(\mathbf{R})\rangle$ of a system with parameterized Hamiltonian $H(\mathbf{R})$ as it is adiabatically transported around a closed loop in parameter space $\mathcal{C} \subset \mathcal{R}$. $\Pi_{\mathcal{C}}$ can be defined as the closed line integral of the Berry connection along the loop $\mathcal{C}$

$$\Pi_{\mathcal{C}} = -i \oint_{\mathcal{C}} d\mathbf{R} \cdot \langle \Phi(\mathbf{R}) | \nabla_{\mathbf{R}} | \Phi(\mathbf{R}) \rangle \,. \tag{5.2}$$

The integrand must be imaginary, as

$$\nabla_{\mathbf{R}} \langle \Phi(\mathbf{R}) | \Phi(\mathbf{R}) \rangle = 2\,\mathrm{Re}[\langle \Phi(\mathbf{R}) | \nabla_{\mathbf{R}} | \Phi(\mathbf{R}) \rangle] = 0, \tag{5.3}$$

thus $\Pi_{\mathcal{C}}$ is real. In this work, we assume $|\Phi(\mathbf{R})\rangle$ is the ground state of $H(\mathbf{R})$.

We need to parameterize the loop $\mathcal{C} = \{\mathbf{R}(t), t \in [0,1]\}$ in order to evaluate the integral. Moreover, it is possible to multiply the ground state $|\Phi(\mathbf{R})\rangle$ by a t-dependent phase, resulting in a $U(1)$-gauge a transformation which leaves all physical quantities invariant. We take

$$|\Psi(t)\rangle = e^{i\Theta(t)} |\Phi(\mathbf{R}(t))\rangle, \tag{5.4}$$

which allows us to rewrite the Berry phase as

$$\Pi_{\mathcal{C}} = -i \int_0^1 dt \langle \Psi(t) | \partial_t | \Psi(t) \rangle + \int_0^1 dt\, \partial_t \Theta(t). \tag{5.5}$$

If there is a representation for which each Hamiltonian $H(\mathbf{R})$ is real, it is possible to choose eigenstates that have all real components. In this case, we can choose $\Theta(t)$ such that $|\Psi(t)\rangle$ has real expansion, which implies the first integrand is real; as $\partial_t\langle\Psi(t)|\Psi(t)\rangle = 0$, the integrand must also to be imaginary, therefore $\langle\Psi(t)|\partial_t|\Psi(t)\rangle = 0$. Under this choice, we can evaluate the Berry phase as a boundary term

$$\Pi_{\mathcal{C}} = \int_0^1 dt\, \partial_t\Theta(t) = \Theta(1) - \Theta(0) = \arg\Big[\langle\Psi(0)|\Psi(1)\rangle\Big], \tag{5.6}$$

where the last equality is obtained using the definition in Eq. (5.4). Furthermore, $|\Psi(1)\rangle$ and $|\Psi(0)\rangle$ are real by construction, which implies $\Pi_{\mathcal{C}}$ can only take two values (modulo 2π): 0 or π.

The quantization of $\Pi_{\mathcal{C}}$ implies that it is invariant for topological deformations of $\mathcal{C}$. If $\mathcal{C}$ can be contracted to a point, then $\Pi_{\mathcal{C}} = 0$. A non-trivial $\Pi_{\mathcal{C}} = \pi$ can only occur when $\mathcal{C}$ encircles a degeneracy. One can check with the effective Hamiltonian Eq. (5.1) that any loop encircling $\mathbf{R}^{\times}$ has a Berry phase of π. This extends by continuity to any region of $\mathcal{R}$ around the CI, as long as $\mathcal{C}$ does not enclose a second degeneracy point. Thus, we have a one-to-one correspondence between CIs and the nontrivial Berry phase.

5.2.3 Measuring Berry phase with a variational wavefunction

There have been various proposals in the literature for computing Berry phases using a gate-based quantum device [286, 326]. In this work, we propose to use a variational algorithm to track $|\Psi(t)\rangle$ when parallel-transported around the loop $\mathcal{C}$, in the spirit of the variational adiabatic method described in Ref. [323, 340]. We approximate

$$|\Psi(t)\rangle \approx |\psi(\boldsymbol{\theta}_t^*)\rangle := U(\boldsymbol{\theta}_t^*)\,|\psi_0\rangle\,, \tag{5.7}$$

where $|\psi(\boldsymbol{\theta})\rangle$ is a variational ansatz state, and $\boldsymbol{\theta}_t^*$ continuously tracks a local minimum $[\nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta}_t^*) = 0]$ of the variational energy

$$E(t, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H(\mathbf{R}(t))|\psi(\boldsymbol{\theta})\rangle. \tag{5.8}$$

The angle $\boldsymbol{\theta}_t^*$ is well-defined as long as the Hessian $\nabla^2_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$ remains positive definite in a neighbourhood of $\boldsymbol{\theta}_t^*$ for all t, ensuring the $\boldsymbol{\theta}_t^*$ is continuous in t and non-degenerate. Although our treatment naturally extends to any variational ansatz that continuously parametrizes normalized states $|\psi(\boldsymbol{\theta})\rangle$ (including classical ansätze like e.g. matrix-product states), we assume the operator $U(\boldsymbol{\theta})$ is implemented by a parameterized quantum circuit (PQC) acting on an initial state $|\psi_0\rangle$; this implies that information about the state needs to be extracted from a quantum device through sampling.

5.3 Methods

In this section, we detail all the ingredients needed to implement our hybrid algorithm to resolve quantized Berry phases with a variational quantum ansatz. Initially, in Sec. 5.3.1, we discuss how selecting an ansatz that preserves the Hamiltonian's symmetries establishes a natural gauge, leading to the reduction of the Berry phase integral to the boundary term Eq. (5.6). In Sec. 5.3.2, we introduce our parameter update approach, which employs single Newton-Raphson steps to trace the variational state along a discretization of the loop $\mathcal{C}$. In Sec. 5.3.3 we explain how to employ a basic regularisation technique to handle the potential non-convexity of the cost function and in 5.3.4 we explain how to measure the final overlap using an ancilla-free Hadamard test. Finally, in Sec. 5.3.5, we provide a full overview of the algorithm.

5.3.1 Fixing the gauge with a real ansatz

As discussed in Sec. 5.2, the quantization of the Berry phase is granted by the symmetries of the Hamiltonian family $H(\mathbf{R})$, which ensure the existence of a basis for which each $H(\mathbf{R})$ has a real representation. When it comes to electronic structure Hamiltonians, it is always possible to find a real representation for time-reversal symmetric Hamiltonians with integer total spin [341]. Moreover, real noninteger-spin Hamiltonians are also found throughout nonrelativistic quantum chemistry.

As our variational ansatz state [in Eq. (5.7)] is defined by a family of unitary operators, it inherits a natural gauge from $U(\boldsymbol{\theta})$. In particular, if $U(\boldsymbol{\theta})$ is written as a product of real rotations in the basis in which $H(\mathbf{R})$ is real, then we force $|\psi(\boldsymbol{\theta})\rangle$ to have real components as well, which fixes a global $U(1)$ phase. This can be obtained by constructing the PQC with a sequence of parameterized unitaries such as

$$U_j(\theta_j) = e^{A_j\theta_j} \tag{5.9}$$

generated by antisymmetric operators A_j that are real in the chosen representation. (We choose dimensional units such that $\|A_j\| = 1$ without loss of generality, see Appendix 5.A). Examples from electronic structure include real fermionic (de-)excitations, such as unitary singles ($A_{pq} = \hat{a}_p^\dagger\hat{a}_q - \hat{a}_q^\dagger\hat{a}_p$) and doubles ($A_{pqrs} = \hat{a}_p^\dagger\hat{a}_q\hat{a}_r^\dagger\hat{a}_s - \hat{a}_q^\dagger\hat{a}_p\hat{a}_s^\dagger\hat{a}_r$). Many PQC ansätze commonly proposed for quantum chemistry, such as unitary coupled cluster (UCC) [54, 239, 306] and quantum-number preserving gate fabrics (NPF) [64], are composed from these elementary rotations. Formally, our ansatz state can then be defined as

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{j=1}^{n_p} U_{n_p-j}(\theta_{n_p-j})\,|\psi_0\rangle\,, \tag{5.10}$$

where U_k are the aforementioned parameterized rotations applied in circuit-composition order.

In this case, the Berry phase can be estimated by

$$\Pi_{\mathcal{C}} = \arg\Big[\langle\psi(\boldsymbol{\theta}^*_{t=0})|\psi(\boldsymbol{\theta}^*_{t=1})\rangle\Big], \tag{5.11}$$

which corresponds to the boundary term in Eq. (5.6). This implies that, in the case of a non-trivial Berry phase, the path traced by $\boldsymbol{\theta}^*_t$ will not close up on itself (i.e. $\boldsymbol{\theta}^*_1 \neq \boldsymbol{\theta}^*_0$), highlighting an important difference between the ansatz parameters $\boldsymbol{\theta}$, which fix the gauge of $|\psi(\boldsymbol{\theta})\rangle$, and the Hamiltonian parameters $\mathbf{R}$, which define a ground state $|\Phi(\mathbf{R})\rangle$ up to a $U(1)$ gauge freedom. However, the change in optimal parameters is insufficient to prove the existence of a nontrivial Berry phase; we must both successfully track the minimum $\boldsymbol{\theta}^*_t$ as t traces from 0 to 1, and estimate the final overlap $\langle\psi(\boldsymbol{\theta}^*_0)|\psi(\boldsymbol{\theta}^*_1)\rangle$. While the argument (sign) of the overlap will yield the Berry phase, its absolute value is a proxy of success as it certifies the initial and final states are physically equivalent.

5.3.2 Avoiding full optimization via Newton-Raphson steps

As mentioned above, the Berry phase $\Pi_{\mathcal{C}}$ is a discrete quantity, and therefore we only need to estimate it to accuracy $< \frac{\pi}{2}$. In Appendix 5.A, we show that this implies that we can accept an error on the estimate $\tilde{\boldsymbol{\theta}}_1$ of the final optimum $\boldsymbol{\theta}^*_1$ bounded in 1-norm by $\|\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^*_1\|_1 < 1$. Thus, we are not required to exactly track $|\psi(\boldsymbol{\theta}^*_t)\rangle$ and as a result, the variational energy Eq. (5.8) does not need to be fully re-optimized at every time-step t. Instead, it suffices to keep the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimal parameters within the basis of convergence of the true minimum $\boldsymbol{\theta}^*_t$.

To achieve this, we still need an initial optimum as an input, which is obtained by running one full optimization. If possible, the initial point is selected such that optimization is simplest. Then, we propose to use a single step of the Newton-Raphson algorithm at points $t \in \{\Delta t, 2\,\Delta t, \ldots, 1-\Delta t, 1\}$.

The Newton-Raphson (NR) algorithm determines the update of the estimate of the minimum $\tilde{\boldsymbol{\theta}}$ through the gradient $\mathcal{G}(t,\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} E(t,\boldsymbol{\theta})$ and Hessian $\mathcal{H}(t,\boldsymbol{\theta}) := \nabla^2_{\boldsymbol{\theta}} E(t,\boldsymbol{\theta})$ of the variational energy Eq. (5.8). The derivatives can be computed using either finite-difference methods or parameter-shift rules [66]; either method requires sampling the variational energy $E(t,\boldsymbol{\theta})$ at a number of different parameter points $\boldsymbol{\theta}$. Given the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimum at point t as an initial guess, the NR step with cost $E(t+\Delta t,\boldsymbol{\theta})$ prescribes the update $\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t + d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}$, with

$$d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t} = -\mathcal{H}^{-1}(t+\Delta t, \tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t+\Delta t, \tilde{\boldsymbol{\theta}}_t). \tag{5.12}$$

The Newton-Raphson method is well-known to have a finite-sized basin of quadratic convergence, as long as the cost function is strongly convex at the optimum,

$$m(\boldsymbol{\theta}^*) := \min_{\mathbf{v}} \frac{\mathbf{v}^T\mathcal{H}(\boldsymbol{\theta}^*)\mathbf{v}}{\mathbf{v}^T\mathbf{v}} > 0. \tag{5.13}$$

In other words, the lowest eigenvalue m of the Hessian of the cost function (which we call convexity) at the optimum $\boldsymbol{\theta}^*$ needs to be positive. Details on the convergence properties of NR are given in App. 5.D. In section 5.4, we show that this quadratic convergence is fast enough to keep track of the minimum with only a single step for each t-point.

5.3.3 Regularization and backtracking

To ensure that we successfully track the minimum of the cost function $\boldsymbol{\theta}_t^*$, the estimates $\tilde{\boldsymbol{\theta}}_t$ need to remain in the strongly convex region of optimization space. The existence of such strongly convex region is not always guaranteed since it depends on the ansatz. A common cause of failure of this requirement is exemplified for ansätze with (local) over-parametrization of the state manifold. In fact, if the ansatz has redundant parameters the Hessian of the cost function Eq. (5.8) will always be singular ($m = 0$). Then, arbitrarily small perturbations of the cost function can then cause $m < 0$. When this occurs, the inversion of the Hessian needed for the Newton-Raphson step is ill-defined.

The most direct approach to solve this issue is to select an ansatz with no degeneracies, facilitating a strongly convex cost function at its minima. Nevertheless, this is only possible for very simple problems, and even in this case, quasi-degeneracies can make the convergence region extremely small. Since this is a well-known problem, many alternative solutions have been proposed in the literature. In this subsection, we will explore and implement two of them – back-tracking and regularization.

Back-tracking — Small positive eigenvalues of the Hessian can cause the standard Newton-Raphson step to overshoot along the relative parameter eigenmodes. This effectively reduces the size of the neighborhood of the minimum θ^* in which quadratic convergence is granted (see Appendix 5.D). Since for positive convexity the direction provided by the Newton-Raphson step is guaranteed to be a descent direction, we can mitigate this overshoot by implementing line-search of the minimum on the segment defined by the NR step. In the common variant of back-tracking, the Newton step is iteratively damped by a constant $\beta \in (0, 1)$. At each iteration, the cost function in the new point is measured, until the cost function is reduced enough (the detailed condition is given in Algorithm 5.1). While some additional evaluations of the cost function are needed, the (more expensive) gradient and Hessian are only calculated once. Due to the repeated evaluation, one needs to consider extending line-search methods to cost functions evaluated with sampling noise.

Algorithm 5.1: NR step subroutine with regularization and backtracking

```
Input: Initial estimated parameters $\tilde{\boldsymbol{\theta}}_\tau$
       Estimated Hessian $\tilde{\mathcal{H}}$
       Estimated gradient $\tilde{\boldsymbol{\mathcal{G}}}$
       Cost function $E(\theta)$ given by eq. (5.8)
       Convexity requirement $m_{\text{thr}} > 0$
       Positive constants $\mu$, $\rho$, $\alpha$, $\beta$
Output: Updated estimated parameters $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$.

$\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
// Regularization
if $\lambda_0 < m$ then
    $\mathcal{B} \leftarrow \tilde{\mathcal{H}} + (\rho|\lambda_0| + \mu)\mathbb{1}$;
    $d\boldsymbol{\theta} \leftarrow \mathcal{B}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
else
    $d\boldsymbol{\theta} \leftarrow \tilde{\mathcal{H}}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
// Backtracking
while $E(\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}) > E(\tilde{\boldsymbol{\theta}}_\tau) + \alpha(\tilde{\boldsymbol{\mathcal{G}}} \cdot d\boldsymbol{\theta})$ do
    $d\boldsymbol{\theta} \leftarrow \beta d\boldsymbol{\theta}$
return $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$
```

Regularization — For realistic ansätze and Hamiltonians, it is difficult to avoid (quasi-) redundancies in some regions of parameter space. In this case, the cost function might not be strongly convex around the minimum, or the convexity might be too small to ensure a sufficiently-large convergence region. Furthermore, even for an ideally-convex cost function, noisy evaluation on a quantum device might result in a distorted Hessian with non-positive eigenvalues. To mitigate this issue, we can use a regularization technique that penalizes the change in parameters along the quasi-redundant directions. We propose to use augmentation of the Hessian to regularize the NR step, obtaining a so-called quasi-Newton optimizer. Hessian augmentation is a common practice in quantum chemistry methods that feature orbital optimization, such as self-consistent field methods [21]. If the smallest eigenvalue of the Hessian λ_0 is smaller than a positive threshold convexity m_{thr}, we construct the augmented Hessian as follows

$$\mathcal{B} = \mathcal{H} + \nu\mathbb{1}, \tag{5.14}$$

where we add a constant $\nu > |\lambda_0|$. The augmented Hessian is then positive, and we can realize the NR update as

$$d\boldsymbol{\theta}^{\text{NR}} = -\beta\mathcal{B}^{-1}\boldsymbol{\mathcal{G}}. \tag{5.15}$$

Here, β is the damping constant from back-tracking line search and $\mathcal{G}$ is the gradient as in eq. (5.12). Regularization and back-tracking are typically used in tandem, as quadratic convergence is harder to guarantee when using regularization. The choice of ν is non-trivial: we want it to be large enough to suppress parameter changes along quasi-redundant directions, but we need to avoid exaggerating the damping along relevant directions. Common solutions include choosing $\nu = \rho|\lambda_0| + \mu$ with fixed positive constants ρ and μ, or using a trust-region method [342, 343] where the Newton step is constrained to lie within a ball of some radius h (such that $||d\boldsymbol{\theta}^{\mathrm{NR}}||_2 \leq h$).

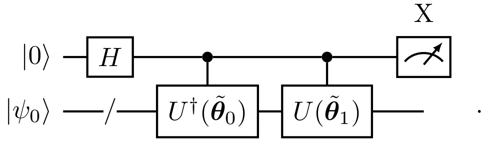
The augmented Hessian method does not require further evaluations of the cost function, although the damping of the parameter updates might imply that more t-steps are needed to successfully resolve the Berry phase.

5.3.4 Measuring the final overlap

For a real ansatz, the overlap of the tracked states at $t = 0$ and $t = 1$ must be real and it can be rewritten as

$$\left\langle \psi(\tilde{\boldsymbol{\theta}}_0) \middle| \psi(\tilde{\boldsymbol{\theta}}_1) \right\rangle = \mathrm{Re}\left[\langle 0| \, U^\dagger(\tilde{\boldsymbol{\theta}}_0) U(\tilde{\boldsymbol{\theta}}_1) \, |0\rangle \right] . \tag{5.16}$$

This quantity can readily be measured by a Hadamard test, implemented as



The required number of samples is small, as we only need to resolve whether the sign of the overlap is $+1$ (trivial $\Pi_{\mathcal{C}} = 0$) or -1 (nontrivial $\Pi_{\mathcal{C}} = \pi$).

Implementing the circuit above requires to realize controlled-$U(\boldsymbol{\theta})$, which might increase significantly the depth of the compiled quantum circuit and make the implementation unfeasible on near-term hardware. However, this requirement can be bypassed by using the control-free echo verification technique [344, 345], in place of the standard Hadamard test, to sample Eq. (5.16). This technique requires access to a reference state $|\psi_{\mathrm{ref}}\rangle$ orthogonal to $|\psi_0\rangle$, which should acquire a known eigenphase φ under the action of the PQC, $U(\boldsymbol{\theta})\,|\psi_{\mathrm{ref}}\rangle = e^{i\varphi}\,|\psi_{\mathrm{ref}}\rangle$. As most of the PQC ansätze used for electronic structure states preserve the total number of electrons (including UCC and NPF), the fully unoccupied state $|0...0\rangle$ can be used as reference. Control-free echo verification circuits only require implementing the non-controlled $U(\boldsymbol{\theta})$, and furthermore provide built-in error mitigation power.

5.3.5 Overview of the algorithm

We are now ready to formalize the proposed algorithm for resolving Berry phases, Algorithm 5.2. The formalization we present here will allow us to bound the number of steps and the sampling cost in the following Section 5.4. Given a path $\mathcal{C}$ and a number of steps $1/\Delta t$, the algorithm attempts to calculate $\Pi_{\mathcal{C}}$ yielding either $\Pi_{\mathcal{C}} = 0$, $\Pi_{\mathcal{C}} = \pi$, or a FAIL state. Again, in Section 5.4 we will bound the probability of the FAIL state occurring. Additional features that extend the practicality of the algorithm and mitigate the failure cases are presented in Sec. 5.5 and later implemented in Sec. 5.6.

Algorithm 5.2: Resolve quantized Berry phase

```
Input: Family of Hamiltonians $H(\mathbf{R})$, $\mathbf{R} \in \mathcal{R}$
       Real ansatz PQC $|\psi(\boldsymbol{\theta})\rangle$
       Set of initial optimal $\boldsymbol{\theta}_0^*$
       Closed path $\mathcal{C} \in \mathcal{R}$, $\mathbf{R}(t) : [0,1] \mapsto \mathcal{C}$
       Number of steps $N$ to discretize $\mathcal{C}$
       Precision requirements $\sigma_{\mathcal{G}} > 0$ and $\sigma_{\mathcal{H}} > 0$
       Convexity requirement $m_{\mathrm{thr}} > 0$
       Regularization reg $\in \{\mathrm{False}, \mathrm{True}\}$
       Final fidelity requirement $F \in (0,1)$
Output: $\Pi_{\mathcal{C}} = 0$ or $\Pi_{\mathcal{C}} = \pi$ or FAIL.

$\Delta t = 1/N$;
$\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0^*$;
for $\tau \in \{0, \Delta t, 2\,\Delta t, ..., 1 - \Delta t\}$ do
    $E(t,\theta) \leftarrow$ define cost function as in Eq. (5.8);
    $\tilde{\boldsymbol{\mathcal{G}}}_j \leftarrow$ sample the gradient to precision $\sigma_{\mathcal{G}}$
        $\left[\tilde{\boldsymbol{\mathcal{G}}}_j = \frac{\partial E}{\partial \theta_j}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
    $\tilde{\mathcal{H}}_{jk} \leftarrow$ sample the Hessian to precision $\sigma_{\mathcal{H}}$
        $\left[\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
    if reg = False then
        $\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
        if $\lambda_0 < m_{thr}$ then return FAIL and exit;
        $d\boldsymbol{\theta}^{\mathrm{NR}} \leftarrow -\tilde{\mathcal{H}}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$ (see Eq. (5.12)); $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow \tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}^{\mathrm{NR}}$;
    if reg = True then
        $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow$ Subroutine 1 $(\tilde{\boldsymbol{\theta}}_\tau, \tilde{\boldsymbol{\mathcal{G}}}, \tilde{\mathcal{H}})$;
$f \leftarrow$ final overlap as in Eq. (5.16) to precision $F$;
if $f^2 < F$ then return FAIL and exit;
return $\Pi_{\mathcal{C}} = \arg\{f\}$
```

If the algorithm fails, it can be re-run with a larger number of steps N and thus a smaller step size Δt. A smaller step size decreases additive NR error bound (the error per step scales as Δt^2, the total bound thus scales as Δt).

5.4 Error analysis and bounding

In this section, we find analytic upper bounds on the cost of estimating a quantized Berry phase $\Pi_{\mathcal{C}}$ on a fixed curve $\mathcal{C}$ using Algorithm 5.2. To simplify the treatment regularization and back-tracking are not considered: instead, we require each estimate $\tilde{\boldsymbol{\theta}}_{j\,\Delta t}$ at the j-th step to be within the region of quadratic convergence of the cost function at the next t-step $E((j+1)\Delta t, \boldsymbol{\theta})$. We prove that this translates to a guarantee of convergence of the algorithm, under three conditions:

1. At the local minimum $\boldsymbol{\theta}^*_t$, where $\psi(\boldsymbol{\theta}^*_t)$ approximates the ground state, the cost function $E(t, \boldsymbol{\theta})$ is strongly convex [as described in Eq. (5.13)];
2. The number of discretization steps N is sufficiently large;
3. The sampling noise on each of the Hessian and gradient elements ($\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$ respectively) is sufficiently small.

The first point entails a requirement on the cost function, defined by the family of Hamiltonians $H(\mathbf{R})$ and the choice of ansatz $|\psi(\boldsymbol{\theta})\rangle$. This requirement is not satisfied if the ansatz state is defined with redundant parameters. We contend that, while strong convexity is a significant assumption, incorporating regularization (or one of the other techniques suggested in the outlook) can alleviate the necessity for such an assumption in practical applications. Our proof provides upper bounds on N and lower bounds on $\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$, which suffice to grant convergence. However, these are not to be considered practical prescriptions, as we do not believe them to be optimal; rather they show which are the relevant factors playing a role in the convergence of the algorithm. As the sampled gradient and Hessian are random variables, the guarantee of convergence for bounded error is to be understood in a probabilistic sense.

We first clarify natural assumptions and notation used in the calculation of the basin of convergence of Newton's method. We require the cost function $E(t, \boldsymbol{\theta})$ to be twice-differentiable by $\boldsymbol{\theta}$, for all t, in a region around the true minima $\boldsymbol{\theta}^*_t$. We require the Hessian to be Lipschitz continuous across this region,

$$\|\mathcal{H}(t, \boldsymbol{\theta}) - \mathcal{H}(t, \boldsymbol{\theta} + d\boldsymbol{\theta})\| < L\|d\boldsymbol{\theta}\|. \tag{5.17}$$

(Here, the Lipschitz constant L can be considered a bound $\|\mathcal{T}\| \leq L$ on the norm of the tensor of third derivatives $T(t, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$.) We also require that the gradient of the t-derivative $\dot{\boldsymbol{\mathcal{G}}} = \nabla_{\boldsymbol{\theta}} \frac{dE}{dt}$ is bounded by $\dot{\mathcal{G}}_{\max}$ in 2-norm. These regularity conditions are satisfied for the PQC ansätze we consider in Sec. 5.5, and in App. 5.B we argue for bounds on L and $\dot{\mathcal{G}}_{\max}$. The strong convexity assumption described in the previous paragraph entails a constant lower bound m_{thr} on the smallest eigenvalue of the Hessian $\mathcal{H}(t, \boldsymbol{\theta}_t^*)$ at the minimizer $\boldsymbol{\theta}_t^*$ for all t.

5.4.1 Bounding the NR error

We will first calculate a lower bound on Δt that ensures the error $\delta\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*$ is bounded by a constant for all values of t. (as shown in Appendix 5.A, a bound on $\|\delta\tilde{\boldsymbol{\theta}}_1\|_1$ is sufficient to ensure $\Pi_{\mathcal{C}}$ can be accurately resolved.) In this calculation, we will allow for an additive error $\sigma_{\boldsymbol{\theta}}$ on $\tilde{\boldsymbol{\theta}}_t$ due to sampling noise; we will simultaneously calculate an upper bound on $\sigma_{\boldsymbol{\theta}}$. We sketch the calculation here and defer details to App. 5.D.

Firstly, it can be shown (Theorem 5.1 in Appendix 5.D) that the Newton-Raphson step Eq. (5.12) with cost function $E(t+\Delta t, \boldsymbol{\theta})$ is guaranteed to converge quadratically [346] to the minimizer $\boldsymbol{\theta}_{t+\Delta t}^*$ as long as the initial guess $\tilde{\boldsymbol{\theta}}_t$ is within a ball centred in the minimizer of radius $\frac{m_{\mathrm{thr}}}{4L}$. Quadratic convergence means that the distance of the updated guess from the minimizer will scale as the square of the distance of the initial guess from the minimizer,

$$\|\delta\tilde{\boldsymbol{\theta}}_{t+\Delta t}\| \leq \frac{L}{m_{\mathrm{thr}}} \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{t+\Delta t}^*\|^2. \tag{5.18}$$

The right-hand side of this equation can be bounded through the triangle inequality as

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| \leq \|\delta\tilde{\boldsymbol{\theta}}_t\| + \|\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_{t+\Delta t}^*\|. \tag{5.19}$$

We can bound the second term in this equation by taking the total t-derivative of the optimality condition $\boldsymbol{\mathcal{G}}(t, \boldsymbol{\theta}_t^*) = 0$, yielding

$$\|\boldsymbol{\theta}_{t+\Delta t}^* - \boldsymbol{\theta}_t^*\| \leq m_{\mathrm{thr}}^{-1} \dot{\mathcal{G}}_{\max} \Delta t. \tag{5.20}$$

If at step t we are within the radius given by Theorem 5.1, the NR step will quadratically converge, suppressing also the error from the previous step, and yielding $\|\delta\tilde{\boldsymbol{\theta}}_t\| \leq \frac{m_{\mathrm{thr}}^2}{16L^2}$.

To account for sampling noise effect, we then consider a small additive error $\sigma_{\boldsymbol{\theta}}$ to $\tilde{\boldsymbol{\theta}}_t$. Maximizing this allowed sampling noise at each step (as we will see, this becomes the bottleneck in our method) then yields the two bounds

$$\sigma_{\boldsymbol{\theta}} \leq \frac{\sqrt{2}-1}{4} \frac{m_{\mathrm{thr}}}{L}, \qquad \Delta t \leq \frac{m_{\mathrm{thr}}^2}{8L\dot{\mathcal{G}}_{\max}}. \tag{5.21}$$

When both bounds are satisfied, we are guaranteed single steps of Newton's method will maintain convergence around the path $\mathcal{C}$.

5.4.2 Bounding the sampling noise

We now translate the bound on $\sigma_{\boldsymbol{\theta}}$ to bounds on the variance of estimates of each element of the gradient and Hessian ($\sigma^2_{\mathcal{G}}$ and $\sigma^2_{\mathcal{H}}$ respectively). This proceeds by simple propagation of variance through Eq. (5.12). We find

$$\begin{aligned}\sigma^2_{\boldsymbol{\theta}} := \mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}] \leq \quad & \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] + \\ & + \|\mathcal{H}^{-1}\|^2 \mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \|d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}\|^2,\end{aligned} \tag{5.22}$$

where $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ are the random variables representing the errors on gradient and Hessian. (a more detailed calculation is given in App. 5.E.) Assuming $\boldsymbol{\theta}$ has n_p elements, each element of the gradient is i.i.d. with variance $\sigma^2_{\mathcal{G}}$, we get

$$\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] = n_p\, \sigma^2_{\mathcal{G}}. \tag{5.23}$$

As $\delta\mathcal{H}$ is a $n_p \times n_p$ real symmetric matrix, assuming its elements are i.i.d. with variance $\sigma^2_{\mathcal{H}}$, we can invoke Wigner's semicircle law [347] to approximate its norm by $\sqrt{n_p}\,\sigma_{\mathcal{H}}$, thus

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \approx n_p\, \sigma^2_{\mathcal{H}}. \tag{5.24}$$

Combining these with Eq. (5.22), and requiring the resulting variance to be small compared to the square allowed additive error $\sigma^2_{\boldsymbol{\theta}}$ we obtain the bound

$$\sigma^2_{\mathcal{G}} + \sigma^2_{\mathcal{H}} \|d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}\|^2 \ll \frac{3-2\sqrt{2}}{16} \frac{m^4_{\mathrm{thr}}}{n_p L^2}. \tag{5.25}$$

We can then bound the norm of the NR update as $\|d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}\| \leq m^{-1}_{\mathrm{thr}} \|\boldsymbol{\mathcal{G}}\|$. Splitting the error budget in half we obtain.

$$\sigma^2_{\mathcal{G}} \ll \frac{3-2\sqrt{2}}{32} \frac{m^4_{\mathrm{thr}}}{n_p L^2} \tag{5.26}$$

$$\sigma^2_{\mathcal{H}} \ll \frac{3-2\sqrt{2}}{32} \frac{m^6_{\mathrm{thr}}}{n_p L^2 \|\boldsymbol{\mathcal{G}}\|^2}. \tag{5.27}$$

Note that these bounds are not tight. For instance, by applying Cauchy-Schwartz inequality to bound $\|\mathcal{H}^{-1} \cdot \mathcal{G}\| \leq \|\mathcal{H}^{-1}\| \|\mathcal{G}\|$, we overlook the fact that the gradient will change more slowly along a lower-eigenvalue eigenmode of the Hessian. We believe further work might allow to define tighter bounds.

5.4.3 Scaling of the total cost

To give an estimate on how many measurements we need to sample gradient and Hessian to sufficient precision, we need to recast the quantities in Eq. (5.27) (the dominant term of the sampling variance) in terms of parameters of the problem. If we use an ansatz without redundancies (or if we can get rid of redundancies through e.g. regularization), and assuming we approximate the ground state well enough, the convexity m_{thr} will be larger than the ground state gap Δ, as every parameterized rotation in the PQC ansatz will introduce a state orthogonal to the ground state. The norm of the gradient and the Lipschitz constant can be bound proportionally to their max norm, as shown in Appendix 5.B, thus $\|\boldsymbol{\mathcal{G}}\| \leq \sqrt{n_p}\|H\|$, $\|\dot{\boldsymbol{\mathcal{G}}}\| \leq \sqrt{n_p}\|\dot{H}\|$ and $L \leq n_p^{3/2}\|H\|$ (where $\|H\|$ is the spectral norm of the Hamiltonian). The number of measurements to sample the Hessian to precision Eq. (5.27) are proportional to the inverse of the bound, with proportionality constant $M_{\mathcal{H}}$ indicating the number of shots required to sample a single element of the Hessian to unit variance (this depends on details such as the decomposition taken to measure the Hamiltonian, and the specifics of the derivative estimation method). Multiplying this by the number of steps $\frac{1}{dt}$ [Eq. (5.21)] gives us the total number of shots required for convergence

$$M_{\text{tot}} = 10^3 \frac{n_p L^3 \|\boldsymbol{\mathcal{G}}\|^2 \dot{\mathcal{G}}_{\max}}{\Delta^8} M_{\mathcal{H}} \tag{5.28}$$

$$< 10^3 \frac{n_p^4 \|H\|^7 \|\dot{H}\|}{\Delta^8} M_{\mathcal{H}}. \tag{5.29}$$

5.5 Adapting to an orbital-optimized PQC ansatz

5

To achieve a good representation of the ground state character while minimizing depth and number of evaluations of quantum circuits, we employ a hybrid ansatz composed of classical orbital rotations and a parameterized quantum circuit (PQC) to represent correlations within an active space. The concept of an orbital-optimized variational quantum eigensolver (OO-VQE) is explored in [30, 348] In this section, we introduce the construction of the OO-PQC ansatz and discuss its specific use in our algorithm, where the orbitals need to continuously track a changing active space depending on the nuclear geometry.

5.5.1 An OO-PQC ansatz with geometric continuity

To represent the electronic structure state, we start by choosing an atomic orbital basis, i.e. a discretization of space defined by a set of N non-orthogonal atomic orbitals $\chi_\mu(\mathbf{R}, \boldsymbol{x})$ (functions of the electronic coordinate $\boldsymbol{x} \in \mathbb{R}^3$, where we make explicit the parametric dependence on the nuclear coordinates $\mathbf{R}$); these orbitals define the overlap matrix

$$S_{\mu\nu}(\mathbf{R}) = (\chi_\mu(\mathbf{R})|\chi_\nu(\mathbf{R})) \tag{5.30}$$

$$:= \int_{\mathbb{R}^3} \chi_\mu^*(\mathbf{R}, \boldsymbol{x}) \chi_\nu(\mathbf{R}, \boldsymbol{x})\, d^3\boldsymbol{x}. \tag{5.31}$$

The atomic orbitals (AOs), along with the overlap matrix, depend on the geometry of the molecule specified by the nuclear coordinates $\mathbf{R}$. (For the sake of simplicity, we limit to considering real AOs.) From these, we could define a set of parameterized orthonormal molecular orbitals (MO)

$$\phi_p(\mathbf{R}, C^{\mathrm{AO}}) = \sum_\mu \chi_\mu(\mathbf{R}) C^{\mathrm{AO}}_{\mu p}, \tag{5.32}$$

which would allow for the definition of a parameterized active space. The downside of this parametrization is that, to ensure MO orthonormality, we need C^{AO} to satisfy the constraint

$$C^{\mathrm{AO}\dagger} S(\mathbf{R}) C^{\mathrm{AO}} = \mathbb{1}, \tag{5.33}$$

which depends nontrivially from $\mathbf{R}$. This implies that we cannot trivially use the same C^{AO} for different geometries $\mathbf{R}$.

In order to address this problem, we have opted to use orthonormalized atomic orbitals (OAO) that are derived from the AOs through symmetric Löwdin orthogonalization [153] as reference in the definition of parameterized MOs. The OAOs are defined as $\phi^{\mathrm{OAO}}_p(\mathbf{R}) = \sum_\mu \chi_\mu(\mathbf{R}) S^{-1/2}_{\mu p}(\mathbf{R})$. Building on these, we can define the MOs as

$$\phi_q(\mathbf{R}, C) = \sum_{\mu,p} \chi_\mu(\mathbf{R}) S^{-1/2}_{\mu p}(\mathbf{R}) C_{pq}, \tag{5.34}$$

where $C = S^{1/2}(\mathbf{R}) \cdot C^{\mathrm{AO}}$. The orthonormality constraint Eq. (5.33) then reduces to requiring C to be orthogonal, and it is independent on $\mathbf{R}$. In summary, Eq. (5.34) defines a set of orthonormal molecular orbitals parameterized by C, well-defined and continuous for changing $\mathbf{R}$.

To start up our algorithm, the matrix C^{AO} can be initialized by a Hartree-Fock (or any other molecular coefficient matrix, e.g. coming from a small CASSCF calculation) at some initial geometry $\mathbf{R}(0)$. From this we recover $C = S^{1/2}(\mathbf{R}(0)) \cdot C^{\mathrm{AO}}$, which is then treated as a variational parameter of the ansatz. Using the parameterized MO Eq. (5.34) we construct the electronic structure Hamiltonian $H(R, C)$ in the (parameterized) molecular basis.

Based on the initial Hartree-Fock orbital energies, we split the N orbitals into a core set with N_{O} doubly-occupied orbitals, an active set with N_{A} orbitals, and a virtual set with N_{V} empty orbitals. Although the split of the orbital indices remains constant throughout the algorithm, the orbitals themselves continuously change through their dependence on $\mathbf{R}$ and C. The correlations are treated only within an active space of η_{A} electrons in N_{A} orbitals. Tracing out the core and virtual orbitals yields the active-space Hamiltonian $H_{\mathrm{A}}(\mathbf{R}, C)$.

The correlated active-space state $|\psi(\boldsymbol{\theta})\rangle$ is represented on a quantum device, using a PQC ansatz of the form Eq. (5.10). The cost function then becomes

$$E(\mathbf{R}, C, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\, H_{\mathrm{A}}(\mathbf{R}, C)\, |\psi(\boldsymbol{\theta})\rangle\,, \tag{5.35}$$

and it can be evaluated by sampling the 1- and 2-electron reduced density matrix (RDM) of the state [61]. (Other efficient sampling schemes, e.g. based on double factorization [11, 59], can be used.)

5.5.2 Measuring boundary terms with the OO-PQC ansatz

When evaluating the final overlap [Eq. (5.16)] with an orbital-optimized ansatz, we have to consider that the states $|\psi(\boldsymbol{\theta}_0)\rangle$ and $|\psi(\boldsymbol{\theta}_1)\rangle$ are defined on different active space orbitals, determined by the MO matrices C_0 and C_1 respectively. The transformation between the two sets of orbitals, $\boldsymbol{\phi}(\mathbf{R}, C_1) = \boldsymbol{\phi}(\mathbf{R}, C_0) \cdot C_{0\to 1}$, is represented by the orthogonal matrix

$$C_{0\to 1} = C_0^\dagger C_1. \tag{5.36}$$

If the algorithms successfully tracked the lowest-energy active space state of the system, the Hilbert spaces spanned by the active orbitals defined by C_0 and C_1 should match. (The same is true for the core space and the virtual space.) This implies the matrix $C_{0\to 1}$ will have block structure, with $[C_{0\to 1}]_{pq} \neq 0$ only if p, q are both in the same set of orbitals (core, active or virtual). The orbital rotations within the core (virtual) subspace will not generate any phase on the state of the system, as all orbitals are doubly-occupied (doubly-unoccupied). The orbital rotation within the active space can then be translated to a unitary transformation on the state by a Bogoliubov transformation

$$G_{0\to 1} = \exp\left\{ \sum_{\{p,q\}\in \mathrm{AS}} [\log(C_{0\to 1})]_{pq} a_p^\dagger a_q \right\}, \tag{5.37}$$

with $a_p^\dagger, a_p$ fermionic creation and annihilation operators on the p orbital. The final overlap

$$\mathrm{Re}\left[\langle 0|\, U^\dagger(\tilde{\boldsymbol{\theta}}_0) G_{0\to 1} U(\tilde{\boldsymbol{\theta}}_1)\, |0\rangle\right] =: \omega_{\mathcal{C}} \tag{5.38}$$

can then be sampled with a Hadamard test, given an quantum circuit implementing the (eventually controlled) operation $G_{0\to 1}$. Under Jordan-Wigner encoding, a quantum circuit for $G_{0\to 1}$ can be implemented as a fabric of parameterized fermionic swap gates of depth N_{A} following a QR decomposition of the orbital rotation generator $[\log(C_{0\to 1})]_{pq}$, also known as a givens rotation fabric [349]. These gates preserve the zero-electrons reference state, allowing to employ the ancilla-free echo verification technique mentioned in section 5.3.4 to measure the final overlap.

5.5.3 Newton-Raphson updates of the OO-PQC ansatz

The proposed OO ansatz has two sets of parameters, C and $\boldsymbol{\theta}$. As the MO matrix C is subject to the constraint Eq. (5.33), its elements cannot be freely updated with NR. Instead, for each NR update with initial MO matrix C, we reparametrize the MOs with a unitary transformation: $C \leftarrow Ce^{-\kappa}$, where κ is any antisymmetric matrix. The derivatives of the energy with respect to any element κ_{pq} can be evaluated analytically (see Appendix 5.C). Furthermore, under this parametrization it can be shown that κ_{pq} where p, q are both core indices or both virtual indices are *redundant* [21] in the definition of the active space orbitals; these $N_{\mathrm{O}}^2 + N_{\mathrm{V}}^2$ parameters are set to zero without reducing the expressivity of the ansatz. We call the unraveled set of remaining parameters $\boldsymbol{\kappa}$. To implement the Newton-Raphson step, the gradient and Hessian with respect to the combined set of parameters $(\boldsymbol{\theta}, \boldsymbol{\kappa})$ is computed. In this manner the gradient splits into two components, and the Hessian into three components

$$\nabla_{(\boldsymbol{\theta},\boldsymbol{\kappa})} E = (\nabla_{\boldsymbol{\theta}} E, \nabla_{\boldsymbol{\kappa}} E) \tag{5.39}$$

$$\nabla^2_{(\boldsymbol{\theta},\boldsymbol{\kappa})} E = \begin{bmatrix} \nabla^2_{\boldsymbol{\theta}} E & \nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E \\ (\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E)^{\intercal} & \nabla^2_{\boldsymbol{\kappa}} E \end{bmatrix}. \tag{5.40}$$

The PQC parameter derivatives $\nabla_{\boldsymbol{\theta}} E$ and $\nabla^2_{\boldsymbol{\theta}} E$ can be evaluated through parameter-shift rule [66, 350] or finite difference, by sampling on the quantum device. The derivatives with respect to the OO parameters $\nabla_{\boldsymbol{\kappa}} E$ and $\nabla^2_{\boldsymbol{\kappa}} E$ are linear functions of the 2-electron RDM, whose coefficients can be computed analytically [21]. The remaining component, the mixed Hessian $\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E$ can be similarly evaluated as a linear function of $\boldsymbol{\theta}$-derivatives of the RDM; for this, we can use the same data sampled from the quantum device to evaluate $\nabla_{\boldsymbol{\theta}} E$. We detail the procedure of estimating these Hessian components in App. 5.C. Thus, evaluating the derivatives with respect to the OO parameters does not require extra sampling on the quantum device.

5.6 Numerical results

In this section, we demonstrate the application of our method to a small model system: the formaldimine molecule $H_2C{=}NH$, an established model in the context of quantum algorithms for excited states in [30, 348, 351]. This molecule is known to have a conical intersection between the singlet ground state and first excited state potential energy surfaces [352]. This CI plays an important role in the photoisomerization process of formaldimine, which in turn can be considered a minimal model for the photoisomerization of the rhodopsin protonated Schiff-base (a key step in the visual cycle process [287, 288]). We consider geometries obtained from the equilibrium configuration by varying the direction of the N−H bond, defined by the bending angle α and the dihedral angle ϕ (see Fig. 5.1d). Varying these angles defines the considered plane in nuclear configuration space $\mathcal{R}$. First, we consider a minimal model of formaldimine (within the minimal basis and a small active space), on which we can test the properties of the algorithm 5.2. Then, we investigate the effects of sampling noise on these results. Finally, we study a more complex model of the same molecule (with a larger basis set and active space), and show that we can achieve similar results by employing regularization and backtracking to deal with the degeneracies of the ansatz manifold.

5.6.1 Numerical simulation details

For our simulations, we use the PennyLane [353] package with the PyTorch backend to construct the hybrid quantum-classical cost function of Eq. (5.35) supporting automatic differentiation (AD) with respect to the PQC parameters. To achieve this, we implement the transformation of the one- and two-body AO basis integrals to the parameterized MO basis [Eq. (5.34)] with AD support. The transformed integrals are projected onto the active space and contracted with the active space one- and two-electron RDM. The RDM elements and their derivatives with respect to the PQC parameters are obtained using PennyLane and its built-in AD scheme based on the parameter shift rule. In summary, gradients and Hessians of the cost function with respect to the PQC parameters are obtained using AD, the orbital gradient and Hessian components are estimated analytically (see Appendix 5.C), and the off-diagonal block of the composite Hessian [Eq. (5.40)] is retrieved by automatic differentiation of the analytical orbital gradient. We use PySCF [222] to generate the molecular integrals (i.e. the full space one- and two-electron integrals and overlap matrices) in the atomic orbital basis.

The core code developed for this project is made available as a python package in the GitHub repository [333]. This code provides a flexible implementation of the orbital-optimized PQC ansatz, which can find many applications in VQAs for chemistry. A tutorial Jupyter notebook showcasing a calculation of Berry phase in the minimal model of Formaldimine is provided in the `examples` folder in the repository.

5.6.2 Minimal model with an degeneracy-free ansatz

We first demonstrate the application of our algorithm to a minimal model of formaldimine, for which we can approximate the ground state with a simple ansatz with no degeneracies. The molecule is described in a minimal STO-3G basis-set, and we select an active space of $\eta_{\mathrm{A}} = 2$ electrons in $N_{\mathrm{A}} = 2$ spatial orbitals [i.e. CAS(2,2)]. As the orbital optimization already allows (spin-adapted) single excitations within the active space, the only parameterized gate we can include in our PQC ansatz is the double-excitation $U(\theta) = e^{\theta(a_0^\dagger a_1^\dagger a_2 a_3 - a_2^\dagger a_3^\dagger a_0 a_1)}$; this corresponds to the unitary coupled-cluster doubles [UCC(S)D] ansatz, where the singles (S) are not explicitly included because they would be redundant with the orbital optimization. This is enough to describe exactly any active space state compatible with the symmetries of the model, without over-parametrizing the ansatz state.

In Fig. 5.1 we demonstrate the application of our algorithm to this model. The minimal basis set is small enough that we can run a full configuration interaction (FCI) calculation to exactly resolve the ground and first excited state energies

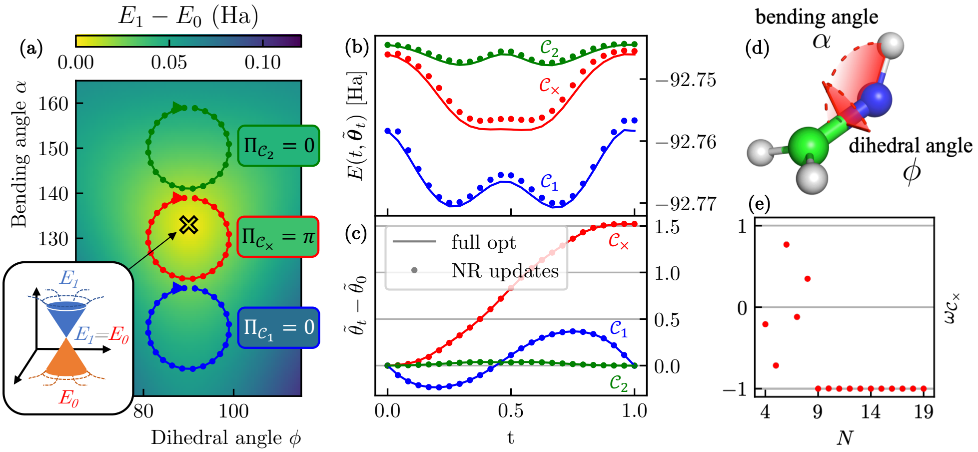


Figure 5.1: (a) Three loops in the nuclear configuration space of formaldimine; $\mathcal{C}_1$ (green), $\mathcal{C}_\times$ (red) and $\mathcal{C}_2$ (blue). $\mathcal{C}_\times$ encircles a CI, resulting in a non-trivial Berry phase. In this representation, the loops are discretized by $N = 25$ points. The color plot indicates the energy gap at the full configuration interaction (FCI) level. **(b)** Energy and **(c)** change in PQC parameter around the three loops, with the same color coding as in (a) and the same $N = 25$. These results refer to a (2,2) active space, and a minimal (STO-3G) basis set description of Formaldimine, with a OO-UCCD ansatz that has a single θ parametrizing the only double excitation. The continuous lines show the true optimum $\boldsymbol{\theta}_t^*$ and the relative energy (obtained by full optimization), while the markers show the progress of the estimate $\tilde{\boldsymbol{\theta}}_t$ from Algorithm 5.2, in absence of sampling noise. The Hessian stays positive throughout the path, no regularization is needed. **(e)** Final overlap computed by Algorithm 5.2 for the red loop containing a CI, for a varying number of total discretization points N. For $N < 9$, the Hessian is not always positive and regularization is needed to invert the Hessian, but no backtracking is used. **(d)** Schematic representation of Formaldimine, indicating the parameters used to define the nuclear geometries in this work.

$E_0(\mathbf{R})$ and $E_1(\mathbf{R})$. Observing the gap $E_1(\mathbf{R}) - E_0(\mathbf{R})$ (portrayed in Fig. 5.1a), we can determine the location of the conical intersection $\phi^\times = 90°$, $\alpha^\times \approx 132°$. We then define three loops in the configuration space $\mathcal{R}$, one loop $\mathcal{C}_\times$ containing the CI and two "trivial" loops $\mathcal{C}_1$, $\mathcal{C}_2$. These loops are centered around $\phi = 90°$ and $\alpha = 110°$ ($\mathcal{C}_1$), $\alpha = 130°$ ($\mathcal{C}_\times$) and $\alpha = 150°$ ($\mathcal{C}_2$), and all have a radius of 10°. Fig. 5.1c shows the progress of the estimate $\tilde{\theta}_t$ (shifted by $\tilde{\theta}_0$) of the optimal PQC parameter θ_t^* throughout the $N = 25$ single-NR-update t-steps. Note that, while we only plot the single PQC parameter relative to the parameterized double excitation, the algorithm updates the MO coefficients $\tilde{C}_t$ as well. We observe that $\tilde{\boldsymbol{\theta}}_1 = \tilde{\boldsymbol{\theta}}_0$ for the trivial loops $\mathcal{C}_1, \mathcal{C}_2$, while $\tilde{\boldsymbol{\theta}}_1 \neq \tilde{\boldsymbol{\theta}}_0$ for the loop $\mathcal{C}_\times$ containing the CI. This is an indication of the effect of the Berry phase, but not the result of the algorithm yet; measuring the overlap Eq. (5.38) yields the correct Berry

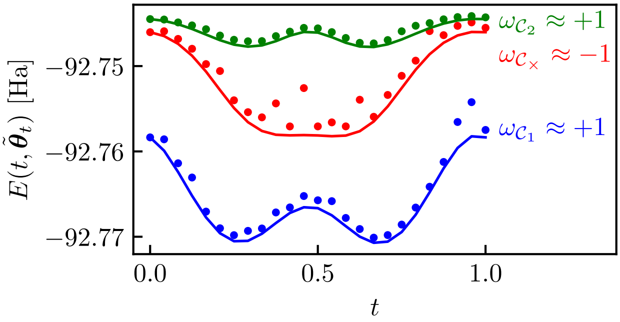


Figure 5.2: Energies throughout the loops in Fig. 5.1 (discretized by $N = 25$ steps), in the presence of sampling noise. Each element of the composite gradient Eq. (5.39) and Hessian Eq. (5.40) are perturbed by random gaussian noise with variance $\sigma^2 = 5 \times 10^{-6}$. In this instance the Hessian stays positive around the path, so no regularization is needed. The full lines, like in Fig. 5.1b, indicate the true optimum $E(t, \boldsymbol{\theta}_t^*)$ obtained by full optimization.

phase $\Pi_{\mathcal{C}} = \arg[\omega_{\mathcal{C}}]$. The estimated energy $E(t, \tilde{\boldsymbol{\theta}}_t)$ and the optimal $E(t, \boldsymbol{\theta}_t^*)$ (obtained by full local optimization) are shown in Fig. 5.1b. We can observe a small deviation from the optimal energy in the region where the character of the state changes faster (along the line $\phi = \phi^\times$, $\alpha < \alpha^\times$), but this does not disrupt the tracking of the minimum. Finally, Fig. 5.1e shows that a number of discretization points $N \geq 9$ is needed to correctly resolve $\Pi_{\mathcal{C}_\times} = \pi$, through the evaluation of the overlap [Eq. (5.38)] $\omega_{\mathcal{C}_\times} = -1$.

5.6.3 Sampling noise

In this section, we explore the robustness of our algorithm with respect to the sampling noise characteristic of VQAs. To avoid defining a specific sampling strategy and keep our results general, we directly add a proxy of sampling noise η to each element of the gradient and Hessian. Each η is an independent Gaussian random variable with variance σ^2; a different η is added to each element of the gradient of (5.39) and Hessian of Eq. (5.40) to get the noisy estimates $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{H}}$. In Fig. 5.2 we show the energy profile of the three loops whose geometry is represented in Fig. 5.1a, for one such random realisation of the sampling noise. (The plotted energy expectation is evaluated exactly, noise is only added to the gradient and Hessian used in the NR updates.) These three loops yield the same Berry phase results as the noiseless case.

The probability P_{success} of Algorithm 5.2 correctly resolving the Berry phase $\Pi_{\mathcal{C}_\times}$ on the nontrivial loop $\mathcal{C}_\times$ is reported in Fig. 5.3, as a function of the number of discretization steps N and of the variance of the added noise on each sampled quantity σ^2. The expected final overlap Eq. (5.38) is -1 for this case, as the loop

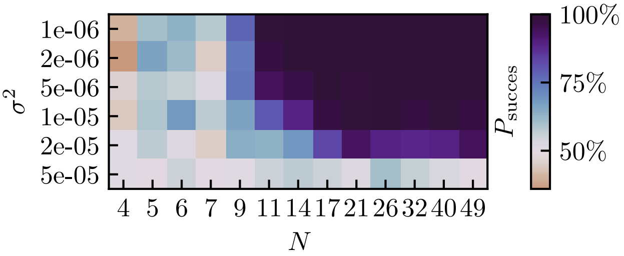


Figure 5.3: Success probability P_{succes} of the algorithm as a function of the number of discretization points N and the sampling noise variance σ^2 for the loop $\mathcal{C}_\times$ containing a CI. The details of the model and geometry of the loop $\mathcal{C}_\times$ match Fig. 5.1 (red loop). Success is defined by resolving a final overlap $\left\langle\psi(\tilde{\boldsymbol{\theta}}_0)\middle|\psi(\tilde{\boldsymbol{\theta}}_1)\right\rangle < 0$, which returns $\Pi_{\mathcal{C}_\times} = \pi$. The success probability is computed over 100 simulated runs. A $P_{\text{success}} \approx 50\%$ indicates the algorithm returns random outcomes. Regularization by Hessian augmentation is enabled in these calculations, while no back-tracking is used.

contains a CI. For each value of N and σ^2, we simulate 100 noisy runs of the algorithm and we declare as successful the ones that yield a negative final overlap (implying $\Pi_{C_\times} = \pi$). Finally we average these outcomes to retrieve the succes probabilities P_{succes}.

From these simulations we conclude that sampling noise reduces the accuracy of tracking the ground state and thus increases the probability of obtaining inaccurate energies. Nevertheless, for a moderate amount of sampling noise our algorithm still resolves the Berry phase correctly. We observe that an error on each gradient and Hessian element with variance of $\sigma^2 = 10^{-5}$ (or smaller) does not compromise the resolution of the Berry phase, as long as the number of discretization points is sufficiently high ($N > 10$, very close to the noiseless case portrayed in Fig. 5.1e). On the other hand, a large enough sampling error ($\sigma^2 \geq 5 \times 10^{-5}$) produces essentially random results ($P_{\text{success}} \approx 50\%$).

5.6.4 Larger basis and active space

To test convergence for a more realistic case where the cost function is not always strongly convex at its minima, we simulate the algorithm on a more challenging model of formaldimine. The model is constructed employing the cc-pVDZ basis set (43 atomic orbitals), and an active space of four electrons in four spatial orbitals [CAS(4,4)]. As a PQC ansatz for the active space state, we use the number-preserving fabric (NPF) ansatz introduced in [64], consisting of a fabric of spin-adapted orbital rotations and double excitations on sets of two spatial orbitals (four spin-orbitals). Four layers of this ansatz are enough to recover the exact CASCI ground state energy inside the active space of 4 orbitals, resulting in 20 PQC parameters. This is an overparameterization of the ground state, implying

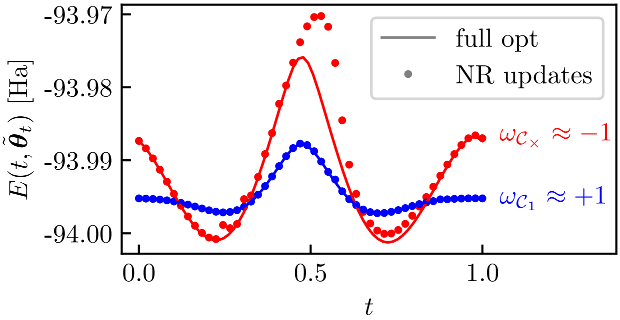


Figure 5.4: Energies throughout two loops, for a more challenging model of Formaldimine, described in the cc-pVDZ basis set and with a (4,4) active space. The (red) blue loop indicates a (non-)trivial Berry phase, centered around ($\alpha = 113°$) $\alpha = 145°$ and $\phi = 90°$ (both), with a radius of 15°, and discretized by $N = 50$ points (note these loops are different from Fig. 5.1). The basis set is too large to run a FCI calculation to locate the CI, and using another approximate method (e.g. state-average CASSCF) might bias the CI location. Instead, we manually choose larger loop geometries and use Algorithm 5.2 to find the value of $\Pi_{\mathcal{C}}$.

a global redundancy in the ansatz and resulting in a singular hessian at every point. The goal of this numerical demonstration is to show that Algorithm 5.2 can still recover the Berry phase, in this case using regularization and backtracking.

In Fig. 5.4 the energies throughout two loops are shown. The location of the conical intersection $(\alpha^\times, \phi^\times)$ in the larger basis set moves compared the case shown in Fig. 5.1 (this is to be expected, as the cc-pVDZ and STO-3G models are effectively different); the basis is now too large to attempt an FCI calculation that would resolve the gap exactly. One could instead resort to a State-Averaged CASSCF calculation to resolve the location of the CI, however, the state-average approach might bias the location of the CI. For this demonstration, we manually select two loops with a slightly larger radius of 15°, centered around $\alpha = 113°$ ($\mathcal{C}_\times$, red line) and around $\alpha = 145°$ ($\mathcal{C}_1$, blue line). These values are chosen based on the location of the CI at the level of theory of large state-average CAS(14,14)SCF calculation, which returns $\alpha^\times \approx 113°$. We choose $\phi = 90°$, as the CI is forced to lie on the $\phi^\times = 90$ hyperplane due to the Cs reflection point-group symmetry. Indeed, we can resolve the correct Berry phase with only $N = 50$ discretiation points, accumulating a small error around the loop which has a minor effect on the final overlap [Eq. (5.38)] of $\omega_{\mathcal{C}_\times} = -0.9994$ (loop containing the CI), and $\omega_{\mathcal{C}_1} = 0.99998$ (trivial loop).

5.7 Conclusion and outlook

In this work, we introduced a hybrid algorithm to resolve conical intersections through the Berry phase they induce. This is achieved by tracking the ground state with a variational quantum ansatz, along a closed path $\mathcal{C}$ in nuclear configuration space. This algorithm only requires approximating the ground state (in contrast to e.g. the state-average VQE [30]) for one nuclear geometry $\mathbf{R}$ at a time, reducing the expressivity requirements of the ansatz. The key requirement of the algorithm is that the ansatz parameters are changed smoothly, tracking a local minimum of the cost function (5.8) and ensuring the $U(1)$-gauge (global phase) of the ansatz state remains well-defined.

As the output is quantized ($\Pi_{\mathcal{C}} \in \{0, \pi\}$), the result only needs to be estimated to a constant precision, and the algorithm is robust to some amount of error; we considered optimization error (the error on estimates of a variational parameter, due to the approximate minimizer) and sampling noise on the quantities measured on the quantum device. We showed that one can update the variational parameters with a single newton step for each geometry in a discretization of the loop $\mathcal{C}$. We proved analytically that the algorithm is ensured to converge for a large enough number of discretization steps N, and small enough additive error, under the assumption that the cost function is strongly convex at its minimum. (We consider sampling noise explicitly, but the robustness results extends to any additive noise, including hardware noise.) We argue this result practically extends to cases where the strong convexity assumption is not satisfied, as long as some regularization technique is employed.

This reasoning is corroborated by numerical demonstrations of CI resolution on a small example system – the formaldimine molecule. Using a minimal description of formaldimine [STO-3G basis, a CAS(2,2), UCCD ansatz], for which we have a strongly convex cost function, we show convergence of Algorithm 5.2 without using regularization for a sufficiently large $N > 11$, as we expect from our analytical results. We also demonstrate the effect of sampling noise in this setting, showing that our algorithm is robust to a sizable amount of noise, achieving convergence for N comparable to the noiseless case. Finally, we demonstrate the application of Algorithm 5.2 with regularization on a more complicated and realistic model of formaldimine [cc-pVDZ basis, CAS(4,4), NPF ansatz]. This case shows that, even with a cost function that is never convex, we can employ regularization to resolve the Berry phase correctly.

5.7.1 Paths towards improving convergence

The key step in our algorithm, where most of the cost in terms of quantum resources is concentrated, is the evaluation of the (NR) parameter update. Ensuring the parameter estimates remain within the basin of convergence of the cost function is crucial, and it is the bottleneck in terms of the cost of our algorithm. As shown in Section 5.4, the size of the basin of convergence depends on the convexity of the cost function at the minimum (5.13). Overparametrizations (local or global) of the cost function are especially disrupting, as they produce singular Hessians ($m = 0$) even at optimal points. In this work, we proposed to use regularization by Hessian augmentation and back-tracking to solve this problem; this technique is practical and, as shown numerically in Section 5.6.4, it can produce convergent results for systems with overparametrized cost functions. However, the success of these techniques may depend on the choice of their hyperparameters (α, β, μ, ρ in Algorithm 5.1), and their application makes the analytical study of the algorithm convergence harder. In this section, we suggest alternative approaches to regularization and resource allocation for the algorithm, which would require further study.

Quantum natural gradients — Quantum natural gradient (QNG) descent is a recently-proposed parameter update technique [354, 355], which takes into account the geometry of the ground state manifold. Natural gradient techniques, already long in use in classical machine learning [356], are invariant with respect to reparametrizations of the cost function; more importantly for our case, they nullify the effect of overparametrizations [357]. The idea of QNG is to transform gradients with respect to the ansatz parameters into gradients with respect to Quantum Information Geometry. This reparametrization is achieved through the Fubini-Study metric tensor $g(\tilde{\boldsymbol{\theta}}_t)$ (to be evaluated at each update). A gradient descent step would then become:

$$\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t - \eta g^+(\tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t+\Delta t, \tilde{\boldsymbol{\theta}}_t), \tag{5.41}$$

where η is the learning rate and $g^+(\tilde{\boldsymbol{\theta}}_t)$ is the pseudo-inverse of the metric tensor. Here $\boldsymbol{\mathcal{G}}(t+\Delta t, \tilde{\boldsymbol{\theta}}_t)$ is just the usual gradient of the energy as in Algorithm 5.2. The resulting QNG step updates the ansatz by a fixed amount in the norm induced by the distance between quantum states, instead of a parameter-space norm, solving the issues connected to overparametrization. Another option would be to use classical natural gradients (NG) [356] defined on the 1- and 2-RDM manifold, which rely on the classical Fisher information matrix.

Adaptive step selection — Choosing a sufficient number of steps N to discretize $\mathcal{C}$ is key to the success of our algorithm, as proven in Section 5.4 and shown in Fig. 5.1 (bottom right). This is because the minimum $\boldsymbol{\theta}_t^*$, along with its basin of convergence, changes between subsequent steps by an amount proportional to the step size $\Delta t = 1/N$. In Algorithm 5.2, we propose to linearly discretize a given parametrization of $\mathcal{C}$ for the sake of simplicity. An adaptive choice of Δt could greatly reduce the cost of the algorithm, letting the steps be larger in the regions where the ground state changes the least, while concentrating more points in the regions where the ground state character sharply shifts. An adaptive step selection technique that preserves the provable convergence could be easily implemented if the gradient and Hessian of the cost function are measured through the 1- and

2-electron RDM. In fact, the cost function can then be written as

$$E(t,\boldsymbol{\theta}) = \sum_{\xi} D(\boldsymbol{\theta})_{\xi} h(t)_{\xi}, \tag{5.42}$$

where $\xi = pq, pqrs \in [N_{\mathrm{A}}]$ are the one- and two-body active-space orbital indices, $D(\boldsymbol{\theta})_{\xi}$ are the RDM elements and $h(t)_{\xi}$ are the one- and two-electron integrals. The derivatives with respect to $\boldsymbol{\theta}$ are then calculated by chain rule from the derivatives of the RDMs, at the current parameter value of $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_t$. This allows to compute energy $E(t', \tilde{\boldsymbol{\theta}}_t)$, gradient $\mathcal{G}(t', \tilde{\boldsymbol{\theta}}_t)$ and Hessian $\mathcal{H}(t', \tilde{\boldsymbol{\theta}}_t)$ for any value t', without further evaluations of the PQC. The step size can be then chosen as the maximum Δt such that the Hessian convexity $m(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)$ remains above some positive threshold value. Further research could quantify the improvement that adaptive step selection would bring to our algorithm, and identify a method to implement this for optimized energy derivatives sampling techniques, such as those using double factorization [358].

5.7.2 Potential applications

The algorithm we propose resolves the Berry phase along a given path $\mathcal{C}$; the description of the loop is an input of the algorithm. This loop construction will depend on the details of the considered application, and might involve chemical intuition and the consideration symmetries of the molecule where present. In realistic applications, we conceive our algorithm as a tool that can help to (1) certify CIs proposed by other methods, (2) determine whether a CI plays a role in a certain reaction, and/or (3) locate a point of the CI manifold in parameter space. In either case, an initial proposal of a path $\mathcal{C}$ that might contain the CI is necessary.

The case 1 is the most direct application of our algorithm. The loop $\mathcal{C}$ is chosen to surround a quasi-degeneracy previously identified by an approximate classical method. The result of our algorithm could then confirm or disprove the presence of the CI. In case 2, given a photochemical reaction whose geometry is approximately known, we can use our algorithm on a set of loops to understand wether a CI plays a role in the reaction. These loops can be constructed by variations of the reaction path along perpendicular coordinates, focusing on the modes that influence the orbitals involved in the reaction, thus greatly reducing the search space for the CI. Finally, to locate the CI (case 3) various search approaches can be considered. For example, starting from a large loop $\mathcal{C}$ that is known to contain the CI, binary-search can be used to iteratively shrink the loop and locate the CI to the desired precision. The considered loops could be defined on a plane, if there is heuristic information about the direction along which the potential energy surfaces split. In alternative, a mesh of orthogonal loops in a subspace of the nuclear configuration space $\mathcal{R}$ can be tested. This, in combination with the data about the ground state energy collected by running the optimization in our algorithm, could also be used to determine the approximate location of the minimal energy crossing point [i.e. the point $\mathbf{R}_{\mathrm{MECP}}$ on the CI manifold with minimum $E_0(\mathbf{R}_{\mathrm{MECP}}) = E_1(\mathbf{R}_{\mathrm{MECP}})$]. Further work is needed to explore these problems, develop procedures to solve them and define and test practical application cases in the three categories.

5.7.3 Outlook

The bounds presented in Section 5.4 have been calculated to provide a guarantee of convergence for our method, which is an atypical feature for variational quantum algorithms. These are not supposed to be tight bounds or resource estimates for a realistic application of our algorithm. Further research is needed to define better bounds. This, along with the choice of a specific method to extract the energy and its derivatives (e.g. RDM sampling and parameter shift rule) could allow estimating the cost of a practical application of this algorithm. Furthermore, a study of the errors due to the ansatz not perfectly reproducing the ground state, and those induced by circuit noise, could help to understand the practical limitations of the algorithm.

The computation of Berry phases is also central when characterizing topological phases of matter [359]. For the specific case of non-interacting Hamiltonians with chiral or inversion symmetry, the winding number is analytically computed through the Zak phase [360], which is related to the Berry phase that is accumulated after a closed loop through the Brillouin Zone. For interacting systems, in which one cannot access momentum space, there is a mechanism in which one introduces an external periodic perturbation to the Hamiltonian [361, 362]. As long as the perturbation does not close the gap and respects the symmetries of the system, the Berry phase can be computed by considering a closed loop in parameter space, similar to what is proposed in this work. As an outlook, one could consider extending the VQE approach to detect topological phases of matter through the computation of the Zak phase. Furthermore, VQA approaches to other topological invariants, such as the Chern number [363] could be considered (possibly inspired by methods related to their experimental detection, such as Thouless pumping [364]).

Finally, classical algorithms to resolve conical intersections from Berry phases inspired by this approach could be designed. If the approximate ground state is represented by a variational classical ansatz that fixes the $U(1)$ gauge, a simple extension of our method could be achievable. For example, this could be achieved with an extension of CASSCF (essentially a CASCI solver on top of an orbital optimization) that implements continuous local optimization of the SCF matrix, to allow enforcing the smoothness constrains that are crucial to keep the gauge of the state fixed.

5.A Bounding overlaps by change in ansatz parameters

The variational Berry phase for a real ansatz state (as introduced in section 5.3.1) is resolved as the argument of the boundary term $\arg[\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle]$. To obtain a nontrivial Berry phase, the initial and final parameters $\tilde{\boldsymbol{\theta}}(0)$ and $\tilde{\boldsymbol{\theta}}(1)$ need to be far enough to allow $\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1$. This implies that the optimal parameters need to change enough along the parametrization of the path (for t going from 0 to 1). In this appendix, we translate this into a lower bound on the one-norm-distance between initial and final parameters $\|\tilde{\boldsymbol{\theta}}(1) - \tilde{\boldsymbol{\theta}}(0)\|_1$. As a consequence, we also find a lower bound on how much error can be allowed without on the final parameter $\tilde{\boldsymbol{\theta}}(1)$ compromising the Berry phase measurement.

We first state a lemma which will be useful in the proof:

Lemma 1

Given two unitary operators U, U', each decomposed as a product of $N \geq 2$ unitary operators $U = U_{N-1}...U_1U_0$, there holds the bound $\|U - U'\| \leq \sum_{j\in[N]}\|U_j - U'_j\|$.

We prove this by induction. For $N = 2$, the proof is by the triangle inequality

$$\begin{aligned}\|U'_1U'_0 - U_1U_0\| &= \|U'_1(U'_0 - U_0) + (U'_1 - U_1)U_0\| && (5.43)\\ &\leq \|U'_1\|\|(U'_0 - U_0)\| + \|(U'_1 - U_1)\|\|U_0\| && (5.44)\\ &= \|(U'_0 - U_0)\| + \|(U'_1 - U_1)\|. && (5.45)\end{aligned}$$

The proof for $N + 1$ can be similarly be reduced to the proof for N:

$$\|U'_NU'_{N-1}...U'_0 - U_NU_{N-1}...U_0\| \leq \|U'_N - U_N\| + \|U'_{N-1}...U'_0 - U_{N-1}...U_0\|. \tag{5.46}$$

Let us consider the case of real Ansatz Eq. (5.10). We remind $U_i = e^{A_i\theta_i}$ for antisymmetric real A_i. We can then bound the overlap between ansatz states using Lemma 1,

$$\begin{aligned}1 - \langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta}')\rangle &= 1 - \langle\psi_0|\,U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta}')\,|\psi_0\rangle && (5.47)\\ &= \langle\psi_0|\,U^\dagger(\boldsymbol{\theta})[U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}')]\,|\psi_0\rangle && (5.48)\\ &\leq \|U(\boldsymbol{\theta}) - U(\boldsymbol{\theta}')\| && (5.49)\\ &\leq \sum_{j\in[N]}\|U_j(\theta_j) - U_j(\theta'_j)\| && (5.50)\\ &= \sum_{j\in[N]}\|e^{A_j\theta_j} - e^{A_j\theta'_j}\| && (5.51)\\ &= 2\sum_{j\in[N]}\left\|\sin\left[\frac{A_j}{2}(\theta_j - \theta'_j)\right]\right\| \leq \sum_{j\in[N]}\|A_j\||\theta_j - \theta'_j| && (5.52)\end{aligned}$$

The difference of parametrers $|\theta_j - \theta'_j|$ is always rescaled by the respective $\|A_j\|$; we can interpret this by considering θ_j and A_j as dimensionful quantities, with inverse dimension to each other. We can always redefine units rescaling θ_j and A_j – without loss of generality we choose units for which $\|A_j\| = 1$ (the only assumption being the boundedness of A_j). Under this choice,

$$\langle\psi(\tilde{\boldsymbol{\theta}}(0))\big|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1 \implies \|\tilde{\boldsymbol{\theta}}(0) - \tilde{\boldsymbol{\theta}}(1)\|_1 \geq 2 \tag{5.53}$$

Thus the parameters need to change (in 1-norm) by at least 2 along the path to achieve the same state and nontivial Berry phase. By the same reasoning, an error on the final parameters $\delta\tilde{\boldsymbol{\theta}}(1)$ bounded by $\|\delta\tilde{\boldsymbol{\theta}}(1)\|_1 < 1$ will not change the argument (i.e. the sign) of the overlap, thus allowing to resolve the correct Berry phase.

5.B Bounding the norm of energy derivatives

Suppose we have a variational state parameterized by $\boldsymbol{\theta}$

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|0\rangle = \prod_{k=0}^{n_p-1} U_k(\theta_k)V_k\,|0\rangle \tag{5.54}$$

on a n_p-dimensional manifold (within a larger Hilbert space), where the product is assumed to be taken in unitary composition order (right to left). Assume each U_k is generated by anti-hermitian operators iA_k with unit norm

$$U_k(\theta) = e^{iA_k\theta}, \quad A_k = A_k^\dagger, \quad \|A_k\| = 1. \tag{5.55}$$

Given H is an observable of known norm $\|H\|$, define

$$E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|\,H\,|\psi(\boldsymbol{\theta})\rangle\,. \tag{5.56}$$

Define the tensors of derivatives,

$$\mathcal{G}_j(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}E(\boldsymbol{\theta}) \tag{5.57}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j\partial\theta_k}E(\boldsymbol{\theta}) \tag{5.58}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}E(\boldsymbol{\theta}) \tag{5.59}$$

$$\tag{5.60}$$

our goal is to bound their (vector-induced) 2-norms $\|\mathcal{G}\|$, $\|\mathcal{H}\|$, $\|\mathcal{T}\|$.

We first notice that

$$\frac{\partial}{\partial \theta_j} U(\theta) = \left(\prod_{k=j}^{n_p - 1} U_k(\theta_k) V_k \right) i A_j \left(\prod_{k=0}^{j-1} U_k(\theta_k) V_k \right) = i \tilde{A}_j U(\theta) \tag{5.61}$$

where $\|\tilde{A}_j\| = \|A_j\| = 1$, as conjugation by a unitary preserves norm. Using this, we can get expressions for the tensors of derivatives in terms of commutators of Hermitian operators of known norm

$$\mathcal{G}_j(\boldsymbol{\theta}) = i \langle \psi(\boldsymbol{\theta}) | [H, \tilde{A}_j] | \psi(\boldsymbol{\theta}) \rangle \tag{5.62}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = - \langle \psi(\boldsymbol{\theta}) | [[H, \tilde{A}_j], \tilde{A}_k] | \psi(\boldsymbol{\theta}) \rangle \tag{5.63}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = -i \langle \psi(\boldsymbol{\theta}) | [[[H, \tilde{A}_j], \tilde{A}_k], \tilde{A}_l] | \psi(\boldsymbol{\theta}) \rangle \tag{5.64}$$

A trivial bound on this involves bounding each commutator by its norm, e.g.

$$\|\mathcal{G}(\boldsymbol{\theta})\|^2 = \sum_{k=0}^{n_p+1} \| \langle \psi(\boldsymbol{\theta}) | [H, \tilde{A}_j] | \psi(\boldsymbol{\theta}) \rangle \|^2 \leq 4 n_p \|H\|^2 . \tag{5.65}$$

It is an open question wether we can improve on this bound.

We can get a similar result for the derivatives with respect to the orbital rotations, considering the reparametrization described in Sec. 5.5.3. We call $|\psi\rangle$ the PQC ansatz state in the full (active + core + virtual) space, padded with virtual (core) registers of qubits in the state $|0\rangle$ ($|1\rangle$). We drop explicit dependence on C and $\boldsymbol{\theta}$, and we make explicit the differential rotation parameters κ. The cost function is then

$$E(\kappa) = \langle \psi | \, e^{\sum_{pq} \kappa_{pq} E_{pq}} H e^{-\sum_{rs} \kappa_{rs} E_{rs}} \, | \psi \rangle \tag{5.66}$$

where E_{pq} is the generator of a spin-adapted orbital rotation. Its derivatives at $\kappa = 0$ [note that the index pairs (pq) are collected in one index for the purpose of rotating higher order derivatives] are easily calculated to be

$$\mathcal{G}_{(pq)}(\kappa = 0) = \langle \psi | \, [H, E_{pq}] \, | \psi \rangle \, , \tag{5.67}$$

$$\mathcal{H}_{(pq),(rs)}(\kappa = 0) = \langle \psi | \, [[H, E_{pq}], E_{rs}] \, | \psi \rangle \, , \tag{5.68}$$

$$\mathcal{T}_{(pq),(rs),(tu)}(\kappa = 0) = \langle \psi | \, [[[H, E_{pq}], E_{rs}], E_{tu}] \, | \psi \rangle \, . \tag{5.69}$$

Observing that $\|E_{pq}\| = 2$ we obtain the same result as above (up to constant factors 2, 4, 8 respectively, coming from this norm).

5.C Analytical orbital gradient and Hessian

In this section, we expand on the estimation of analytic orbital gradient [right block of the vector in Eq. (5.39)] and orbital-orbital and orbital-circuit Hessian [bottom right and top right blocks of the matrix in Eq. (5.40), respectively]. We show how after the reparametrization $C \leftarrow C \cdot e^{-\boldsymbol{\kappa}}$, the $\boldsymbol{\kappa}$-derivatives of the cost function $E(C \cdot e^{-\boldsymbol{\kappa}}, \boldsymbol{\theta})$ can be expressed as a linear function of the 1- and 2-electron RDM, and the mixed Hessian $\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E(C \cdot e^{-\boldsymbol{\kappa}}, \boldsymbol{\theta})$ can be expressed in terms of $\boldsymbol{\theta}$-derivatives of the same RDM.

We first define the 1- and 2-electron reduced density matrix (RDM) in the spin-restricted formalism

$$\gamma_{pq}(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| E_{pq} |\psi(\boldsymbol{\theta})\rangle \tag{5.70}$$

$$\Gamma_{pqrs}(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| e_{pqrs} |\psi(\boldsymbol{\theta})\rangle \,, \tag{5.71}$$

Here, p, q, r, s are meant to be general indices (either occupied, active or virtual), where the state $|\psi(\boldsymbol{\theta})\rangle$ is to be intended as padded by two registers of qubits in the $|0\rangle^{\otimes 2N_\mathrm{V}}$ ($|1\rangle^{\otimes 2N_\mathrm{O}}$) state for the virtual (occupied) orbitals. In the molecular orbital basis defined by C [orbitals in Eq. (5.34)], we can write the Hamiltonian as

$$H = \sum_{pq} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} e_{pqrs} \tag{5.72}$$

where h_{pq} and g_{pqrs} are the one- and two-electron integrals (with spatial orbital indices p, q, r, s ordered according to the chemists' convention), and they implicitly depend on C through the MOs. The expectation value of the Hamiltonian can then be written as a contraction of the integrals with the RDM,

$$E(C, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| H(C) |\psi(\boldsymbol{\theta})\rangle = \sum_{pq} [h(C)]_{pq} \gamma_{pq} + \frac{1}{2} \sum_{pqrs} [g(C)]_{pqrs} \Gamma_{pqrs} \tag{5.73}$$

where we made explicit the dependence on C.

To derive analytical orbital rotation derivatives, we closely follow Ref. [21]. We start by separating the dependence on κ of the reparametrized cost function $E(C \cdot e^{-\kappa}, \boldsymbol{\theta})$, by using the equivalent state transformation formalism provided by Thouless theorem [28]

$$E(C \cdot e^{-\kappa}, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})| H(C \cdot e^{-\kappa}) |\psi(\boldsymbol{\theta})\rangle = \langle\psi(\boldsymbol{\theta})| e^{\hat{\kappa}} H(C) e^{-\hat{\kappa}} |\psi(\boldsymbol{\theta})\rangle \tag{5.74}$$

where $\hat{\kappa} = \sum_{pq} \kappa_{pq} E^-_{pq}$ is the operator that generates a unitary on the Hilbert state space equivalent to the orbital rotation. We know that the rotations where p, q are both virtual indices form a redundant subgroup, so we can freeze the corresponding $\kappa_{pq} = 0$; the same is true for p, q both core space indices. In other terms, $\kappa_{pq} = 0$ if $p, q \in V$ or $p, q \in O$, with V and O the sets of virtual and core indices. The remaining elements of κ_{pq} satisfy $\kappa_{pq} = -\kappa_{qp}$. We define a vector of unique non-redundant orbital rotation parameters

$$\boldsymbol{\kappa} = \{\kappa_{pq}, \forall p \in O \cup A, \forall q \in A \cup V : q > p\}, \tag{5.75}$$

and we redefine the cost function with respect to this vector,

$$E(C, \boldsymbol{\kappa}, \boldsymbol{\theta}) \equiv E(C \cdot e^{-\kappa}, \boldsymbol{\theta}). \tag{5.76}$$

We are interested in the derivative with respect to this vector; we can always switch from the matrix κ to the unraveled vector of unique non-redundant parameters $\boldsymbol{\kappa}$, and vice versa. By comparing the Taylor series in κ with the Baker-Campbell-Hausdorff expansion:

$$E(\boldsymbol{\theta}, \boldsymbol{\kappa}) = \langle\psi(\boldsymbol{\theta})|\, H\, |\psi(\boldsymbol{\theta})\rangle + \langle\psi(\boldsymbol{\theta})|\, [\hat{\kappa}, H]\, |\psi(\boldsymbol{\theta})\rangle + \frac{1}{2}\, \langle\psi(\boldsymbol{\theta})|\, [\hat{\kappa}, [\hat{\kappa}, H]]\, |\psi(\boldsymbol{\theta})\rangle + \ldots \tag{5.77}$$

One can readily verify that the analytical orbital derivatives at $\kappa_{pq} = 0$ are given by:

$$[\nabla_{\boldsymbol{\kappa}} E]_{pq} := \left.\frac{\partial E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq}}\right|_{\kappa=0} = \langle\psi(\boldsymbol{\theta})|\, [E^{-}_{pq}, H]\, |\psi(\boldsymbol{\theta})\rangle \tag{5.78}$$

$$[\nabla^2_{\boldsymbol{\kappa}} E]_{pqrs} := \left.\frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq} \partial \kappa_{rs}}\right|_{\kappa=0} = \frac{1}{2}(1 + P_{pq,rs})\, \langle\psi(\boldsymbol{\theta})|\, [E^{-}_{pq}, [E^{-}_{rs}, H]]\, |\psi(\boldsymbol{\theta})\rangle \tag{5.79}$$

where $P_{pq,rs}$ permutes the pair of indices pq with rs. The calculation of the commutators in Eq. (5.78) and (5.79) can be found in common quantum chemistry textbooks [21], and they all one- or two-body operators; thus their expectation value can be written as a linear form in the RDM (γ, Γ). The gradient evaluates to

$$[\nabla_{\boldsymbol{\kappa}} E]_{pq} = 2(F_{pq}(\boldsymbol{\theta}) - F_{qp}(\boldsymbol{\theta})) \tag{5.80}$$

where F is the generalized Fock matrix,

$$F_{pq}(\boldsymbol{\theta}) = \sum_{m} \gamma_{pm}(\boldsymbol{\theta}) h_{qm} + \sum_{mnk} \Gamma_{pmnk}(\boldsymbol{\theta}) g_{qmnk} \tag{5.81}$$

The Hessian evaluates to

$$[\nabla^2_{\boldsymbol{\kappa}} E]_{pqrs} = (1 - P_{pq})(1 - P_{rs}) \left[2\gamma_{pr} h_{qs} - (F_{pr} + F_{rp})\delta_{qs} + 2Y_{pqrs}\right], \tag{5.82}$$

where we introduced

$$Y_{pqrs} = \sum_{mn} \Gamma_{pmrn} g_{qmns} + \Gamma_{pmnr} g_{qmns} + \Gamma_{prmn} g_{qsmn} \tag{5.83}$$

and dropped the explicit dependence on $\boldsymbol{\theta}$. For the composite Hessian, we simply take the gradient of Eq. (5.80) with respect to $\boldsymbol{\theta}$, by using the chain rule

$$[\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E]_{pq} := \left.\frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq} \partial \boldsymbol{\theta}}\right|_{\kappa=0} = 2\left(\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial F_{qp}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \tag{5.84}$$

where

$$\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{m} \frac{\partial \gamma_{pm}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} h_{qm} + \sum_{mnk} \frac{\partial \Gamma_{pmnk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} g_{qmnk}. \tag{5.85}$$

Thus, once we have the derivatives of the 1- and 2-RDM to sufficient precision, we can evaluate the orbital gradient, Hessian and composite Hessian analytically, recovering all terms in Eq. (5.39) and Eq. (5.40) without any additional quantum cost.

5.D Bounding the cumulative error due to Newton-Raphson updates

In this section, we prove that using a single Newton-Raphson (NR) parameter update per Δt-step is sufficient to achieve an error on the estimate of the minimizer scaling as $O(\Delta t^2)$ after any number of Δt-steps, as long as the cost function is strongly-convex at the minimum and Δt is small enough. First, we recall sufficient conditions for quadratic convergence of the Newton-Raphson step. We then use these to bound the error of a single NR-step when the cost function is changed from $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$. We translate this into an upper bound on Δt which guarantees the error stays bounded throughout the optimization path. Finally, we show that we can allow a sufficiently small additive error on the Newton-Raphson update and retain the bounded error throughout the path.

Quadratic convergence of NR — Consider a cost function $E(\boldsymbol{\theta})$ with gradient $\mathcal{G}_j = \frac{\partial E}{\partial \theta_j}$ and Hessian $\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}$, and an initial guess of a minimizer $\boldsymbol{\theta}^{(0)}$. The Newton-Raphson step prescribes the update $\boldsymbol{\theta}^{(0)} \mapsto \boldsymbol{\theta}^{\mathrm{NR}} = \boldsymbol{\theta}^{(0)} + d\boldsymbol{\theta}^{\mathrm{NR}}$ with $d\boldsymbol{\theta}^{\mathrm{NR}} = \mathcal{H}^{-1}(\boldsymbol{\theta}^{(0)}) \mathcal{G}(\boldsymbol{\theta}^{(0)})$. Theorem 3.5 from Nocedal and Wright [346] gives sufficient conditions under which quadratic convergence of the NR update is guaranteed. We simplify these conditions, and obtain the following

Theorem 5.1
Consider a cost function $E(\boldsymbol{\theta})$ with Lipschitz-continuous Hessian $\|\mathcal{H}(\boldsymbol{\theta}) - \mathcal{H}(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\| \leq L\|\delta\boldsymbol{\theta}\|$ and a local minimizer $\boldsymbol{\theta}^*$ with positive convexity $m := \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\|^{-1} >$

0. *Given an initial guess* $\boldsymbol{\theta}^{(0)}$ *which is close enough to the minimum, i.e.*

$$\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\| \leq \frac{m}{4L}, \tag{5.86}$$

the NR update will converge quadratically towards the minimum with

$$\|\boldsymbol{\theta}^{NR} - \boldsymbol{\theta}^*\| \leq \frac{L}{m}\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\|^2. \tag{5.87}$$

To prove this, we only need to show that a strong convexity condition is satisfied within a r-ball centered in $\boldsymbol{\theta}^*$ including all close-enough possible initial guesses ($r = \frac{m}{4L}$), i.e.

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq 2m^{-1}, \quad \forall\|\delta\boldsymbol{\theta}\| \leq \frac{m}{4L}. \tag{5.88}$$

To prove this we expand $\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})$ using Taylor's theorem,

$$\exists 0 < s < 1: \quad \mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta}) = \mathcal{H}^{-1}(\boldsymbol{\theta}^*) + \delta\boldsymbol{\theta} \cdot \frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta}) \tag{5.89}$$

$$\frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}} = -\mathcal{H}^{-1}\frac{\partial \mathcal{H}}{\partial \boldsymbol{\theta}}\mathcal{H}^{-1} \tag{5.90}$$

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\| + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2 L\|\delta\boldsymbol{\theta}\| \tag{5.91}$$

$$\leq m^{-1} + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2\frac{m}{4}, \tag{5.92}$$

where we used the Lipschitz constant L as a bound on the derivative of the Hessian. This last condition holds if

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\| \leq 2m^{-1}. \tag{5.93}$$

As we can choose $s < 1$ and the result is clearly true at $\delta\boldsymbol{\theta} = 0$, the result holds recursively.

Single NR update with a changing cost function — We now consider a family of cost functions $E(t, \boldsymbol{\theta})$ continuously parameterized by t. Suppose we have an approximation $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}^*_t$ of $E(t, \boldsymbol{\theta})$, with error $\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\|$. In each step of our method (Algorithm 5.2), we shift the cost function $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$ by Δt and we use the current minimizer estimate $\tilde{\boldsymbol{\theta}}_t$ as initial guess for the next step; $\boldsymbol{\theta}^{(0)}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t$. We can bound the error of this initial guess by the triangle inequality,

$$\|\boldsymbol{\theta}^{(0)}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + \|\boldsymbol{\theta}^*_{t+\Delta t} - \boldsymbol{\theta}^*_t\|. \tag{5.94}$$

While the first term is the (given) error on the estimate, the second can be obtained by taking the total t-derivative of the minimum condition $\mathcal{G}(t, \boldsymbol{\theta}^*(t)) = 0$, and applying Taylor's theorem

$$\exists\tau \in [t, t+\Delta t]: \quad \|\boldsymbol{\theta}^*_{t+\Delta t} - \boldsymbol{\theta}^*_t\| = \left\| \left.\frac{d\boldsymbol{\theta}^*_t}{dt}\right|_{t=\tau} \right\| \Delta t = \|\mathcal{H}^{-1}(\tau, \boldsymbol{\theta}^*_\tau)\dot{\boldsymbol{\mathcal{G}}}(\tau, \boldsymbol{\theta}^*_\tau)\|\Delta t \tag{5.95}$$

with $\dot{\boldsymbol{\mathcal{G}}} = \nabla \frac{\partial}{\partial t} E(\boldsymbol{\theta}, t)$.

We now assume that the convexity at the minimum is bounded from below throughout the whole t-path by a constant $m \geq 0$,

$$m(t, \boldsymbol{\theta}^*_t) := \|\mathcal{H}^{-1}(t, \boldsymbol{\theta}^*_t)\|^{-1} > m \quad \forall t \in [0, 1], \tag{5.96}$$

and that the gradient of the change is never larger than $\dot{\mathcal{G}}_{\max}$. (while the first assumption imposes the nontrivial condition of strong convexity at the minimum, the second is always granted for cost functions from a continuous family of bounded Hamiltonians). We can then write

$$\|\boldsymbol{\theta}^{(0)}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t. \tag{5.97}$$

To ensure this initial guess is within the quadratic convergence region of the NR step, we require

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t \leq \frac{m}{4L}, \tag{5.98}$$

choosing an $\alpha, \beta \in (0, 1]$ this condition can be written as

$$\Delta t = \frac{m^2}{4L\dot{\mathcal{G}}_{\max}}\alpha\beta, \tag{5.99}$$

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| = (1-\alpha)\beta\frac{m}{4L}. \tag{5.100}$$

This allows to apply Theorem 5.1, and bound the error after a single NR step,

$$\|\tilde{\boldsymbol{\theta}}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \frac{L}{m}\left[\frac{\beta m}{4L}\right]^2 = \beta^2\frac{m}{16L}. \tag{5.101}$$

Multiple steps — We want to ensure the error on the minimizer estimate remains bounded for t taking subsequent values is $[0, \Delta t, 2\,\Delta t, ..., 1]$, while taking a single NR step at a time. We can do this by imposing the error after each step [Eq. (5.101)] is not larger than the error on the previous step estimate,

$$\beta^2\frac{m}{16L} \leq (1-\alpha)\beta\frac{m}{4L}. \tag{5.102}$$

This is granted for any $\beta \in (0, 1]$ by choosing $\alpha = 1 - \frac{\beta}{4}$. The maximum $\Delta t = \frac{3}{4}\frac{m^2}{4L\dot{\mathcal{G}}_{\max}}$ is achieved by picking $\beta = 1$, and yields an error bounded by the constant $\frac{m}{16L}$.

Allowing an additive error — To account for sampling noise, it is useful to consider an additive error of magnitude $\sigma_{\boldsymbol{\theta}}$ on the estimate $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}_t^*$ at each t-point, modifying Eq. (5.100) into

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*\| = (1-\alpha)\beta\frac{m}{4L} + \sigma_{\boldsymbol{\theta}}. \tag{5.103}$$

This yields the condition

$$\frac{L}{m}\left[\frac{\beta m}{4L} + \sigma_{\boldsymbol{\theta}}\right]^2 \leq (1-\alpha)\beta\frac{m}{4L}. \tag{5.104}$$

If we define γ by $\sigma_{\boldsymbol{\theta}} = \frac{\gamma\beta m}{4L}$, we can write

$$(1+\gamma)^2\frac{m}{16L}\beta^2 \leq (1-\alpha)\beta\frac{m}{4L}, \tag{5.105}$$

which is saturated by $\beta = 4\frac{1-\alpha}{(1+\gamma)^2}$. We then get

$$\Delta t = \frac{m^2}{L\dot{\mathcal{G}}_{\max}}\frac{\alpha(1-\alpha)}{(1+\gamma)^2}, \tag{5.106}$$

which is maximised (while keeping $\beta \leq 1$) by $\alpha = \max[\frac{1}{2}, 1 - \frac{(1+\gamma)^2}{4}]$. The allowed additive noise is then

$$\sigma_{\boldsymbol{\theta}} = \frac{\gamma}{(1+\gamma)^2}\frac{m}{L}(1-\alpha) = \frac{\gamma m}{2L}\min[\frac{1}{2}, \frac{1}{(1+\gamma)^2}], \tag{5.107}$$

maximised for the choice $\gamma = \sqrt{2}-1$ yielding

$$\sigma_{\boldsymbol{\theta}} = \frac{\sqrt{2}-1}{4}\frac{m}{L}, \quad \Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}. \tag{5.108}$$

5.E Bounding the sampling cost

We call $\sigma_{\mathcal{G}}^2$ and $\sigma_{\mathcal{H}}^2$ the variances of each element of the gradient and Hessian respectively, due to sampling noise. To compute the error on the parameter updates, we propagate these variances through the definition of the NR update Eq. (5.12). The first-order differential change (here denoted with δ) of the NR update $d\boldsymbol{\theta}^{\mathrm{NR}}$ with respect to changes in the gradient and Hessian is

$$\delta[d\boldsymbol{\theta}^{\mathrm{NR}}] = \mathcal{H}^{-1}\cdot[-\delta\boldsymbol{\mathcal{G}} + \delta\mathcal{H}\cdot d\boldsymbol{\theta}^{\mathrm{NR}}], \tag{5.109}$$

where we use the ordered matrix-product notation, with vectors in boldface. When $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ are the random variables representing the errors on the gradient and Hessian, the expected mean square error on the NR update defining the norm of the covariance matrix

$$\left\|\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}]\right\| := \mathbb{E}\left[\|\delta[d\boldsymbol{\theta}^{\mathrm{NR}}]\|^2\right] = \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\boldsymbol{\mathcal{G}}\|^2\right] + \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\mathcal{H}\cdot d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right], \tag{5.110}$$

where we used the zero-average property of $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ to drop the expectation values of mixed terms. This can further be bounded as

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] + \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right]\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2. \tag{5.111}$$

Assuming the same variance $\sigma_{\mathcal{G}}^2$ on each of the n_p elements $\mathcal{G}_j$ of the gradient, we get

$$\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] = n_p\,\sigma_{\mathcal{G}}^2. \tag{5.112}$$

As the Hessian is a random real symmetric matrix with i.i.d. elements, each with a variance $\sigma_{\mathcal{H}}$, we can invoke Wigner's semicircle law [347] to bound the spectral norm as

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \leq (\sqrt{n_p}\,\sigma_{\mathcal{H}})^2. \tag{5.113}$$

Combining these with the strong convexity bound $\|\mathcal{H}^{-1}\| \leq m^{-1}$, we get

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq m^{-2}\left[n_p\,\sigma_{\mathcal{G}}^2 + n_p\,\sigma_{\mathcal{H}}^2\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \tag{5.114}$$

The calculations in Appendix 5.D conclude that, for each Δt-step, we can afford an additive error on the NR update of at most $\sigma_{\boldsymbol{\theta}} \leq \frac{\gamma}{4}\frac{m}{L}$ with $\gamma = \sqrt{2}-1$. Comparing this result to the variance just calculated, we can formulate the requirement on the variance

$$m^{-2}n_p\left[\sigma_{\mathcal{G}}^2 + \sigma_{\mathcal{H}}^2\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \ll \frac{\gamma^2}{16}\frac{m^2}{L^2}. \tag{5.115}$$

This gives conditions on the elementary sampling variances

$$\sigma_{\mathcal{G}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p} \tag{5.116}$$

$$\sigma_{\mathcal{H}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p\,\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2} < \frac{\gamma^2}{16}\frac{m^6}{L^2 n_p\|\dot{\mathcal{G}}_{\max}\|^2\Delta t^2} \tag{5.117}$$

To recast this bound in terms of variables of the problem, we use the following relations derived in Appendix 5.B: $L = \max\|\mathcal{T}\| < n_p^{3/2}\|H\|$ (the norm of the third derivative tensor $\mathcal{T}$ is bounded by $n_p^{3/2}$ times by its infinity norm), $\|\dot{\mathcal{G}}_{\max}\|\Delta t < \sqrt{n_p}\|\frac{dH}{dt}\|\Delta t \approx \sqrt{n_p}\|H\|$ (same infinity norm bound). Furthermore, we assume the convexity is larger than the ground state gap, $m > \Delta$; this holds if the ansatz approximates the ground state well enough, and changes in any ansatz parameter

θ_k introduce a different excited state. Substituting these relations we obtain

$$\sigma_{\mathcal{G}}^2 \ll 0.01\frac{\Delta^4}{\|H\|^2 n_p^4}, \tag{5.118}$$

$$\sigma_{\mathcal{H}}^2 \ll 0.01\frac{\Delta^6}{\|H\|^4 n_p^5}. \tag{5.119}$$

The number of total required shots to sample the Hessian (gradient) for all the N steps will thus scale as $\sigma_{\mathcal{H}}^{-2}\Delta t^{-1}$ $(\sigma_{\mathcal{G}}^{-2}\Delta t^{-1})$. Picking the maximal $\Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}$, and considering only the dominant term (relative to sampling the Hessian) we can write

$$\#\text{shots} \propto n_p^7 \frac{\|H\|^7 \|\frac{dH}{dt}\|}{\Delta^8} \tag{5.120}$$

CHAPTER 6

Machine Learning Density Functionals from Noisy Quantum Data

6.1 Introduction

Following the rapid development of quantum hardware, the last decade saw an explosion in the research on quantum algorithms for noisy intermediate-scale quantum (NISQ) devices [53]. An important goal of this research is identifying avenues towards achieving useful quantum advantage: the application of quantum devices to relevant, classically-intractable problems. A prominent example is the electronic structure problem, which is central to chemistry and material science, and has been a major focus for quantum algorithms due to its inherent quantum nature and significant scientific and commercial importance [7–9]. Much of the recent research on quantum algorithms for this problem has centered on the variational quantum eigensolver (VQE), a NISQ-tailored algorithm designed to approximate ground states by optimizing a heuristic quantum ansatz [54, 92]. However, several significant challenges hinder the achievement of practical solutions to the electronic structure problem. Accurately estimating energies and other properties from a state prepared on a quantum computer is made prohibitively expensive by sampling costs [63]. This is exacerbated by the necessity of applying error mitigation techniques to reduce bias due to circuit noise, which increases the sampling overhead [365]. Additionally, VQE introduces errors intrinsic to the algorithm, related to the expressibility of the ansatz and the complexity of the optimization landscape. In practical electronic structure calculations, it is often necessary to solve for multiple system configurations, which increases the overall

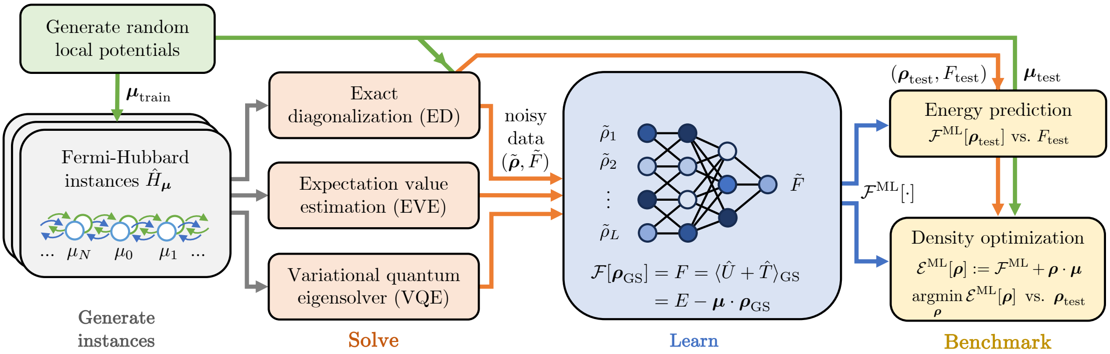


Figure 6.1: Overview: Workflow for machine learning a DFT functional from noisy quantum data. The process begins with generating instances of the Hubbard model $\hat{H}_{\boldsymbol{\mu}}$ from random local potentials $\boldsymbol{\mu}_{\text{train}}$. For each of these instances, estimates of the ground state density $\tilde{\boldsymbol{\rho}}$ and the potential-interaction energy $\tilde{F}$ are obtained using one of three different methods: ED, EVE or VQE. The data extracted from EVE and VQE are noisy, representing two distinct types of errors — handled separately in this study — that are characteristic of NISQ quantum algorithms: sampling noise for EVE and expressibility and optimization errors for VQE. The pairs $(\tilde{\boldsymbol{\rho}}, \tilde{F})$, split in training and validation sets using 5-fold cross validation, are used to train a machine learning model to learn the universal functional $\mathcal{F}$. This, in turn can be used to predict the ground state energy $E_{\text{GS}} = F[\boldsymbol{\rho}_{\text{GS}}] + \boldsymbol{\mu} \cdot \boldsymbol{\rho}$, and to perform density optimization for new potentials (in a Kohn-Sham-like scheme). The ML models are benchmarked on a test set of new problem instances defined by random potentials $\boldsymbol{\mu}_{\text{test}}$ and solved with exact diagonalization. The trained models are benchmarked on both of these tasks using a separate test set from ED to investigate the robustness to errors.

cost.

Density Functional Theory (DFT) is the workhorse of modern quantum chemistry and material science, used to investigate the electronic structure of many-body systems at a low computational cost [366]. Central to DFT is the definition of a *universal functional*, which maps the ground state electronic density to the corresponding kinetic and interaction energy for a given problem family [15, 18]. Although exact universal functionals theoretically exist, their general explicit form is unknown, and their evaluation would necessarily be computationally intractable [16]. Instead, DFT practitioners rely on a vast array of approximate functionals, developed over the decades through mathematical assumptions and physical intuition. Recently, machine learning (ML) approaches have emerged as powerful tools for designing new DFT functionals [367, 368]. In particular, deep-learning-based functionals have demonstrated remarkable performance on benchmark problems in chemistry [369–371]. These models can be trained on mixed datasets generated through various approaches, including DFT based, on other functionals, expensive first-principles methods like coupled-cluster, and experimental data [372, 373]. The synergies between DFT and quantum algorithms have been explored in the context of learning functionals on fault-tolerant quantum computers [374], and in using quantum algorithms to supplement results from approximate DFT functionals [375].

As quantum hardware continues to improve, the prospect of using quantum devices to generate training data that is inaccessible to classical methods is becoming increasingly realistic. Such data, derived from the quantum simulation of complex systems, could significantly enhance the capabilities of classical ML models [376–378]. In turn, these models could generalize from the limited, expensive-to-generate quantum data. Moreover, by leveraging the underlying structure, physically-motivated ML models could combine information from all the training data points, filtering out noise and leading to predictions that are more accurate than any individual training point. While not much research has focused on practically training classical models with quantum-generated data, related concepts exist in the literature. In quantum science, ML is often applied to learning properties of a system from measurement outcomes, in tasks like phase classification [379]. Classical shadow tomography [243] constructs classical models that can predict many properties of a given quantum state from a limited set of measurements. Similarly, recent work demonstrates a technique to extract of classical surrogates from QML models trained on quantum devices [380]. Moreover, classical ML models have been employed for noise mitigation in quantum systems [381].

In this work, we explore the application of NISQ devices to generate noisy training data aimed at learning DFT functionals for the Fermi-Hubbard model, with the goal of generalizing and enhancing these datasets. We focus on limited, noisy datasets of 1,000 data points, which, though small from a machine learning perspective, can already be expensive to generate on a quantum device. We separately examine the effects of sampling noise typical of NISQ algorithms and the algorithmic errors associated with VQE, analyzing how these factors impact the learned functionals. The models are benchmarked on both an energy prediction task and a Kohn-Sham-like density optimization task.

The remainder chapter is structured as follows: Sec. 6.2 summarizes the necessary background on DFT, the physical system and the learning task. Sec. 6.3 discusses the dataset generation process, including details on the considered quantum algorithms. Sec. 6.4 describes the machine learning model, training process, and compares various ML regression methods. Sec. 6.5 presents the benchmarking results of the trained ML models. Finally, Sec. 6.6 discusses the findings and provides an outlook for future research.

6.2 Background

6.2.1 Density Functional Theory

Central to DFT are the Hohenberg–Kohn theorems [15], which assert two fundamental principles: (1) the many-body ground state of a system of interacting electrons in an external potential is uniquely determined by its electron density, and (2) the ground state energy can be obtained variationally by minimizing a functional of the electronic density.

To formalize this and set the notation, consider a family of electronic Hamiltonians

$$\hat{H} = \hat{U} + \hat{T} + \hat{V} \tag{6.1}$$

where $\hat{U}$ represents a fixed electron-electron interaction, $\hat{T}$ is a fixed kinetic energy term, and $\hat{V}$ is a free external potential term. The expectation value of the potential $\langle \hat{V} \rangle$ is defined as a linear functional of the electronic density $\boldsymbol{\rho}$.

The first Hohenberg-Kohn theorem ensures the existence of the *universal functional*

$$\mathcal{F} : \boldsymbol{\rho}_{\mathrm{GS}} \mapsto F := \langle \psi_{\mathrm{GS}} | \, (\hat{U} + \hat{T}) \, | \psi_{\mathrm{GS}} \rangle \,, \tag{6.2}$$

which maps the electron density $\boldsymbol{\rho}_{\mathrm{GS}}$ of the ground state $|\psi_{\mathrm{GS}}\rangle$ of any $\hat{H}$ to the respective expectation value of the kinetic and interaction energy F. The functional is termed *universal* because it is independent of the external potential acting on the system. The total ground state energy $E = \langle \psi_{\mathrm{GS}} | \, \hat{H} \, | \psi_{\mathrm{GS}} \rangle$ can be reconstructed from the ground state density as

$$E = \mathcal{E}[\boldsymbol{\rho}_{\mathrm{GS}}] := \mathcal{F}[\boldsymbol{\rho}_{\mathrm{GS}}] + \langle \psi_{\mathrm{GS}} | \, \hat{V} \, | \psi_{\mathrm{GS}} \rangle \,, \tag{6.3}$$

where $\langle \psi_{\mathrm{GS}} | \, \hat{V} \, | \psi_{\mathrm{GS}} \rangle$ is a linear functional of $\boldsymbol{\rho}_{\mathrm{GS}}$ by definition. The second Hohenberg-Kohn theorem states that, for a given $\hat{V}$, the ground state energy and density of the corresponding $\hat{H}$ can be obtained by minimizing this functional:

$$E = \min_{\boldsymbol{\rho}} \mathcal{E}[\boldsymbol{\rho}] \,, \quad \boldsymbol{\rho}_{\mathrm{GS}} = \arg\min_{\boldsymbol{\rho}} \mathcal{E}[\boldsymbol{\rho}]. \tag{6.4}$$

While DFT approaches have been developed to investigate a wide variety of many-body systems, here we focus on a particular lattice model. In lattice models, electrons occupy discrete, localized sites, and the electron density $\boldsymbol{\rho}$ is represented as a vector where each component ρ_j denotes the occupation of the j-th site. The functional $\mathcal{F}[\boldsymbol{\rho}]$ is the reformulated as a function of this vector, simplifying the computational treatment of the system and making it more amenable to a machine learning scheme.

6.2.2 The Fermi-Hubbard model

In this work, we focus on the Fermi-Hubbard model, a paradigmatic lattice spin-$\frac{1}{2}$ Fermion model with on-site interaction. We take this model on a one-dimensional ring of L sites. The Hamiltonian of this model,

$$\hat{H}_{\boldsymbol{\mu}} = \hat{T} + \hat{U} + \hat{V}_{\boldsymbol{\mu}}, \tag{6.5}$$

comprises a kinetic energy term $\hat{T}$, ab interaction term $\hat{U}$, and a potential term $\hat{V}_{\boldsymbol{\mu}}$. All of these can be expressed in terms of fermionic creation and annihilation operators, $\hat{a}_{j,\sigma}$ and $\hat{a}^{\dagger}_{j,\sigma}$. From here on, $j \in \{1, ..., L\}$ denotes the site index, and $\sigma = \{\uparrow, \downarrow\}$ represents the electron spin. The number operator is defined as $\hat{n}_{j,\sigma} = \hat{a}^{\dagger}_{j,\sigma}\hat{a}_{j,\sigma}$. The kinetic term describes the hopping of electrons between adjacent sites and is given by:

$$\hat{T} := -t \sum_{j}^{L} \sum_{\sigma \in \{\uparrow,\downarrow\}} (\hat{a}^{\dagger}_{j,\sigma}\hat{a}_{j+1,\sigma} + \hat{a}^{\dagger}_{j+1,\sigma}\hat{a}_{j,\sigma}), \tag{6.6}$$

where t is the hopping parameter. We choose without loss of generality $t = 1$, thereby fixing the units of all energies. We consider periodic boundary conditions, meaning ($\hat{a}_{L+1,\sigma} := \hat{a}_{1,\sigma}$). The interaction term accounts for the on-site repulsion between electrons of opposite spins and is expressed as:

$$\hat{U} := u \sum_{j}^{L} \hat{n}_{j,\uparrow}\hat{n}_{j,\downarrow}, \tag{6.7}$$

where u is the interaction strength. The potential $\hat{V}_{\boldsymbol{\mu}}$ defines a specific instance of the Hubbard model within the broader family of models characterized by the chain length L and the interaction strength u/t. The local chemical potential $\boldsymbol{\mu} = (\mu_0, ..., \mu_L)$ introduces an additional term at each site to the Hamiltonian:

$$\hat{V}_{\boldsymbol{\mu}} = \boldsymbol{\mu} \cdot \boldsymbol{n} := \sum_{j}^{L} \mu_j (\hat{n}_{j,\uparrow} + \hat{n}_{j,\downarrow}). \tag{6.8}$$

Each choice of $\boldsymbol{\mu}$ defines a unique instance $\hat{H}_{\boldsymbol{\mu}}$ within the model family, with ground state $|\psi_{\mathrm{GS}}\rangle$, and the corresponding ground state energy E. The elements of the electronic density vector $\boldsymbol{\rho}_{\mathrm{GS}}$ are the expected number of particle at each site, summed over the two spin species:

$$\rho_j = \langle\psi_{\mathrm{GS}}|(\hat{n}_{j,\uparrow} + \hat{n}_{j,\downarrow})|\psi_{\mathrm{GS}}\rangle\,. \tag{6.9}$$

The universal functional $\mathcal{F}$ is then a function mapping the vector $\boldsymbol{\rho}_{\mathrm{GS}}$ to the scalar $F = \langle\psi_{\mathrm{GS}}|\,(\hat{U} + \hat{T})\,|\psi_{\mathrm{GS}}\rangle$; we call this the *Hubbard functional.* Note that the Hubbard functional depends on the value of the interaction strength u/t, the number of sites L and the geometry of the system.

In our numerical study, we focus on a periodic Hubbard chain of $L = 8$ sites; this system is small enough to allow effective exact diagonalization benchmarks. We choose an interaction strength of $u/t = 4$, corresponding to the most challenging regime to simulate when scaling to larger systems [382]. Furthermore, we restrict all solutions to the quarter-filling ($L/4 = 2$ particles of each spin species) and total-spin singlet subspace, based on the symmetries of the Hubbard model. A more general functional can in principle be obtained by stitching together functionals defined on different symmetry sectors.

6.2.3 The learning task

In recent years, data-driven approaches for learning approximate DFT functionals have emerged, where a machine learning model is trained to map electron densities to energies [367]. For the Hubbard model, Nelson *et al.* [86] demonstrated a method for training a model to learn the functional $\mathcal{F}$ using 105,000 density-ground state energy pairs obtained via exact diagonalization. However, the reliance on exact diagonalization limits the feasible system size due to its high computational demands. In this work, we consider the task of learning an approximation $\mathcal{F}^{\mathrm{ML}}$ of the same functional, but from noisy data generated by a quantum algorithm. Generating such data is expensive, with the cost rising with both the size of the dataset and the desired accuracy. By training a model on a limited dataset, we aim to enable generalization and reduce noise, potentially leading to predictions with greater accuracy than any individual data point.

Our overall approach to learning a DFT functional from noisy quantum data is illustrated in Fig. 6.1: The process begins by initializing the Hubbard model with random values of $\boldsymbol{\mu}$ drawn from a specified distribution (Sec. 6.3). The ground state problem is then solved for 1000 potentials to obtain a dataset of noisy density-energy pairs $(\tilde{\boldsymbol{\rho}}_X, \tilde{F}_X)$, where X indicates the method used to generate the data. To represent the noise typical of NISQ algorithms, we consider and simulate two methods: expectation value estimation (EVE), which isolates the effect of sampling noise, and the variational quantum eigensolver (VQE), which isolates expressibility and optimization errors. Exact diagonalization results are used for benchmarking, while the noisy data are employed to train a neural network-based regression model. This approach allows us to explore how the inherent noise characteristic of quantum data from current near-term devices impacts the training of a machine-learned DFT functional.

6.3 The dataset

Training and testing data sets are constructed by solving configurations of the Hubbard problem with a random on-site potential $\boldsymbol{\mu}$. The same configurations are solved with each of the considered methods: exact diagonalization (ED), expectation value estimation (EVE), and variational quantum eigensolver (VQE). The training set and the test set are generated using 1000 independent random potentials each, the training set will be further split to implement 5-fold cross-validation. Note that this number of configurations is smaller compared to the 105,000 potentials used in [86]. We focused on this small-data regime due to the anticipated computational expense of generating data points with quantum algorithms.

To sample the random potential $\boldsymbol{\mu}$ we follow a modified version of the approach in [86]. For each data point, we first sample a strength parameter $W \in [0.005t, 2.5t]$ uniformly at random. We then sample $\boldsymbol{\mu}$ uniformly at random and calculate its standard deviation $\sigma(\boldsymbol{\mu}) = \sqrt{\sum_j \mu_j^2 - (\sum_j \mu_j)^2}$. If the standard deviation $\sigma(\boldsymbol{\mu}) < 0.4t$ we accept the potential and add it to our dataset, otherwise we reject and repeat the sampling procedure from the beginning. This procedure produces a representative distribution of potential of varied strengths avoiding too large energy fluctuations, without requiring the solution of the Hubbard problem instance at this stage.

We construct an exact dataset to act as a baseline, solving all the problem instances (training and test set) with exact diagonalization. For every problem instance $\hat{H}_{\boldsymbol{\mu}}$, we construct the Hamiltonian matrix block corresponding to quarter filling and spin-singlet. We diagonalize it, obtaining the ground state $|\psi_{\mathrm{GS}}(\boldsymbol{\mu})\rangle$ and the energy $E(\boldsymbol{\mu})$. From the state we calculate the expected particle density vector $\boldsymbol{\rho}_{\mathrm{GS}}(\boldsymbol{\mu})$. We then compute the kinetic-interaction energy $F(\boldsymbol{\mu}) = E(\boldsymbol{\mu}) - \boldsymbol{\mu} \cdot \boldsymbol{\rho}_{\mathrm{GS}}(\boldsymbol{\mu})$. Thus, from each random configuration $\boldsymbol{\mu}$, we obtain a pair $(\boldsymbol{\rho}_{\mathrm{GS}}, F)$ representing the input and output of the exact Hubbard functional $\mathcal{F}$.

6.3.1 Expectation value estimation

Assuming the exact ground state $|\psi_{\mathrm{GS}}\rangle$ of the Fermi-Hubbard Hamiltonian Eq. (6.5) can be prepared on a quantum device, measuring F and $\boldsymbol{\rho}_{\mathrm{GS}}$ will still incur an error called *sampling noise*. To reproduce this effect we simulate expectation value estimation, by drawing samples from the measurement outcomes distribution defined by the Born rule on the ground state vector $|\psi_{\mathrm{GS}}\rangle$ produced by exact diagonalization.

To define the measurements for expectation value estimation, we note that all the terms we want to measure are diagonal either in real space (density and Coulomb energy) or in Fourier space (kinetic energy) [63]. For estimating the real-space-diagonal terms, we can sample all the number operators $\hat{n}_{j,\sigma}$ at the same time. Averaging M samples we obtain estimates of each $\langle \hat{n}_{j,\sigma} \rangle$, and thus reconstruct an estimate $\tilde{\rho}_j$ for the density $\rho_j = \langle \hat{n}_{j,\uparrow} \rangle + \langle \hat{n}_{j,\downarrow} \rangle$. The expectation value of the Coulomb energy $\langle \hat{U} \rangle = u \sum_j \langle \hat{n}_{j,\uparrow} \hat{n}_{j,\downarrow} \rangle$ can be estimated from the same samples by averaging the on-site correlators $\hat{n}_{j,\uparrow} \hat{n}_{j,\downarrow}$. The kinetic energy operator $\hat{T}$ is diagonalized by expressing it in terms of the Fourier-space fermionic operators $\hat{a}_{k,\sigma} = 1/N \sum_{j=0}^{N-1} e^{i\,jk2\pi/N} \hat{a}_{j,\sigma}$,

$$\hat{T} = 2t \sum_{k=0}^{N-1} \sum_{\sigma=\uparrow,\downarrow} \cos\left(\frac{2\pi k}{N}\right) \hat{n}_{k,\sigma} \tag{6.10}$$

where $\hat{n}_{k,\sigma} = \hat{a}^{\dagger}_{k,\sigma}\hat{a}_{k,\sigma}$. The operators $\hat{n}_{k,\sigma}$ can be sampled at the same time; another M samples are taken and averaged to obtain $\langle \hat{n}_{k,\sigma} \rangle$, and their results combined to estimate $\langle \hat{T} \rangle$. The estimate $\tilde{F}_{\mathrm{EVE}}$ of F is finally calculated by summing the kinetic and Coulomb contributions. For the central limit theorem the estimate of $\tilde{F}$ is unbiased and asymptotically normal, with variance proportional to $1/M$.

Assuming we have a qubit-based quantum device and Jordan-Wigner Fermion-to-qubit encoding [55, 124], the real-space samples are simply obtained by measuring all qubits in the computational basis as $\hat{n}_{j,\sigma} = (1 - Z_{j,\sigma})/2$ (where $Z_{j,\sigma}$ is the Pauli Z-operator on the qubit that encodes the spin-orbital j, σ). In order to sample the plane wave number operators $\hat{n}_{k,\sigma}$, we need to perform a basis rotation circuit implementing the fast fermionic Fourier transform [383]: a circuit of depth $O(N)$ acting on the two spin sectors independently.

6.3.2 Variational quantum eigensolver

The VQE is a hybrid quantum-classical method that combines a quantum subroutine with a classical optimizer, aiming to estimate the ground state energy of a Hamiltonian [54]. The quantum subroutine prepares on a quantum device an ansatz state $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})\,|\psi_{\mathrm{GS}}\rangle$ dependent on a set of classical parameters $\boldsymbol{\theta}$; and measures its expected energy $E(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}) | H | \psi(\boldsymbol{\theta}) \rangle$ through EVE. This energy is then minimized over the parameters $\boldsymbol{\theta}$ using a classical optimizer which calls the quantum subroutine.

The VQE is a heuristic method, which means that convergence to the true ground state energy is not guaranteed [384]. The manifold of states $\psi(\boldsymbol{\theta})$ that can be represented by the chosen ansatz is, in general, much smaller than the Hilbert space of the problem. This is a feature of VQE: restricting the number of parameters is the only way to make the problem feasible for the classical optimizer. The minimum energy ansatz state will thus only approximate the ground state, with its energy $\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta})$ guaranteed to be larger than the ground state energy by the variational principle. The quality of the results depends on the choice of ansatz, the performance of the classical optimizer, and is further affected by hardware noise and sampling noise. In this work, we focus on isolating the effect of the errors intrinsic to the VQE due to ansatz expressibility and optimization. We use a classical state-vector simulator to extract expectation values without sampling noise, which is studied separately through EVE. Both the energy and the electronic density of the optimal state are affected by the VQE errors.

Variational ansatz families fall into two main categories: hardware-efficient and physically inspired. Physically inspired ansätze, like the Unitary Coupled Cluster (UCC) [54, 62] consider essential system properties but often require high circuit depths, making their implementation challenging for near-term quantum devices. Hardware-efficient ansätze instead maximize the number of parameters per circuit layer, but may still require many layers to approximate the ground state and do not take into consideration the problem symmetries. We will consider two ansätze in this work: the well-known Variational Hamiltonian Ansatz (VHA) [63] and the Number Preserving Fabric (NPF) [64] ansatz, which combine the advantages of both categories.

The VHA is one of the earliest proposed ansätze for VQE study of the Fermi-Hubbard model [63, 323], and has found success in hardware experiments [385, 386]. This ansatz is inspired by the adiabatic algorithm [387, 388] and formally equivalent to the quantum alternating operator ansatz (QAOA) [389, 390]. It distinguishes itself from the usual VQE ansätze by incorporating the values of $\boldsymbol{\mu}$ defining the problem instance into the ansatz definition. The circuit structure for the Hubbard model, for parameters $\boldsymbol{\theta} = \{\theta_{i,j}\}$, is

$$|\psi \{\theta_{i,j}\}\rangle = \prod_{i=1}^{p} \left[e^{-i\hat{T}_e\theta_{i,0}} e^{-i\hat{T}_o\theta_{i,1}} e^{-i\hat{V}_\mu\theta_{i,2}} e^{-i\hat{U}\theta_{i,3}} \right] |\psi_0\rangle \,, \tag{6.11}$$

thus able to represent a first order trotter decomposition in p slices of the adiabatic evolution to H from H^0, commonly taken as the non-interacting Hamiltonian of which $|\psi_0\rangle$ is the ground state. In Eq. (6.11), $\hat{T}_e$ ($\hat{T}_o$) are all hopping terms on even (odd) sites in Eq. (6.6).

The NPF ansatz takes another strategy which is, in contrast to VHA, independent of the problem. It composes the two fundamental interactions contained in a spin-preserving two-body Hamiltonian; a spatial orbital rotation and a double excitation. The parameterized composition of these operations we call $Q(\theta, \phi)$, and they are arranged in a brick-wall pattern to maximize the gate density, resulting an expressive ansatz with only local gates. For the exact form of the ansatz, see Ref. [64].

While both the VHA and NPF ansatz are physically inspired and preserve key symmetries like total spin and number of particles, we highlight here some differences between the two. The number of parameters in the VHA equals 4 per layer by inspection of Eq. (6.11). The NPF ansatz has two parameters per Q-block, of which there are $\frac{N_q}{2} - 1$ per layer. In our case $N_q = 2N = 16$, resulting 14 parameters per layer. Although the NPF ansatz has more parameters per layer, its higher gate density means it does not necessarily result in deeper circuits. VQE cost functions are often highly irregular and non-convex, with numerous local minima. Hence, the choice of classical optimizer is crucial [68]. The VHA, being a more structured ansatz, may have a more irregular cost landscape, making it prone to getting stuck in local minima. To address this, we use the gradient-free Constrained Optimization by Linear Approximation (COBYLA) optimizer. For the NPF, the abundance of parameters offers more flexibility but less structure, and empirically, following the gradient in this cost landscape tends to work better. Therefore, we employ the Sequential Least Squares Programming (SLSQP) optimizer for the NPF ansatz.

6.4 Learning the functional

In this section we present the chosen machine learning models and the details of the training procedure. Alongside the main neural-network model, for which we provide a detailed result analysis in Sec. 6.5, we present a comparative study of a set of out-of-the box models in Sec. 6.4.2.

6.4.1 Method

The main ML model we consider is an adaptation of the convolutional neural network (CNN) model proposed by Nelson *et al.* [86]. The model consists of a sequence of one 1-dimensional periodic convolutional layer with 8 local features and kernel size 3, two dense layers with 128 features each, and a final dense layer with a single output.

Before being processed by the CNN, the inputs $\boldsymbol{\rho}_X$ are normalized. Each element ρ_j of the density $\boldsymbol{\rho}$ is normalized using the mean $\bar{\rho}$ and standard deviation σ_ρ of the whole training set:

$$\rho_j \mapsto (\rho_j - \bar{\rho})/\sigma_\rho =: \mathrm{input}_j, \tag{6.12}$$

(where we dropped the method index X for convenience of notation). Conversely, the output of the model is rescaled by the standard deviation of the training energies σ_F and shifted by their mean $\bar{F}$:

$$\mathrm{output} \mapsto \sigma_F \cdot \mathrm{output} + \bar{F} := \mathcal{F}[\boldsymbol{\rho}]. \tag{6.13}$$

In potential real-world application of our method to noisy data from systems too large for exact benchmarking, the training data becomes the sole source of information for model evaluation. Thus, model selection, hyperparameter optimization and validation must be conducted exclusively using the available (noisy) dataset. To address this, we use 5-fold cross-validation [391]. This involves partitioning the noisy training dataset into five equally sized subsets (folds). During each iteration, one fold is set aside as the validation set, while the remaining four folds are used to train the model. The performance metrics obtained from each fold are then averaged to provide a comprehensive evaluation of the model performance.

The target functional $\mathcal{F}[\boldsymbol{\rho}]$ is invariant under translations, mirror symmetry, and their combination. This means that $\mathcal{F}[\boldsymbol{\rho}]$ does not change under the transformation

$$\rho_j \mapsto \rho_{(\pm j+k) \bmod N}\,, \quad \forall k \in \{0, ..., N-1\}. \tag{6.14}$$

We exploit this to augment the training dataset: from each pair $(\tilde{\boldsymbol{\rho}}_X, \tilde{F}_X)$ we construct 16 data points, applying the $N = 8$ shifts and $N = 8$ mirror-and-shifts to the density and copying the energy [86]. This data augmentation is performed after the cross-validation split, on the training and validation splits separately. The data points within the training split are then scrambled before dividing in batches for training.

The model is trained with the Adam optimizer [392]. To avoid overfitting, we implement an early stopping strategy which monitors the validation loss during training, halting when this loss stops decreasing. The model and its training are implemented using the TensorFlow-Keras library [393, 394].

6.4.2 Alternative model selection

Given the changes in the underlying statistical structure of the data in this work compared to Ref. [86], we evaluate the suitability of the CNN model by comparing its performance with other machine learning models known for their robust regression capabilities and ability to handle noise and outliers. The objective of this section is to determine if any of these alternative models can enhance the learning task with our noisy data. Our findings indicate that the CNN model is the best performing model. Nonetheless, we present our comparative study for completeness.

The ensemble of models to be evaluated comprises 11 regression models, including the convolutional neural network (CNN) model from Ref. [86], three linear regression-based approaches known for their outlier robustness (Huber regression [395], Theil-Sen regression (TS) [396], and random sample consensus (RANSAC) regression [397, 398]), five decision tree models (AdaBoost (AB) [399], random forest (RF) [400], gradient boosting (GB) [401]), support vector regression (SVR) [402], nearest neighbor regression (KNN) [403], and XGBoost (XGB) [404]. For the CNN model, we adopted the hyperparameters from Nelson *et al.* [86], while for the other models, we used standard baseline hyperparameters.

To evaluate model performance, we present results for models trained on data generated from both the VQE (using an NPF ansatz with depth $d = 5$) and EVE ($M = 2000$ shots) methods. These two instances are chosen to represent the typical characteristics of VQE and EVE datasets, and we focus exclusively on them for clarity and readability. For benchmarking purposes, we also include performance metrics for models trained on exact data. In scenarios where our method is applied to noisy data from systems that are too large for exact benchmarking, the cross-validation of training data will be the only resources available. Consequently, model selection and hyperparameter optimization must be based solely on these datasets: For this, we employ 5-fold cross-validation, as introduced in Sec. 6.4.1.

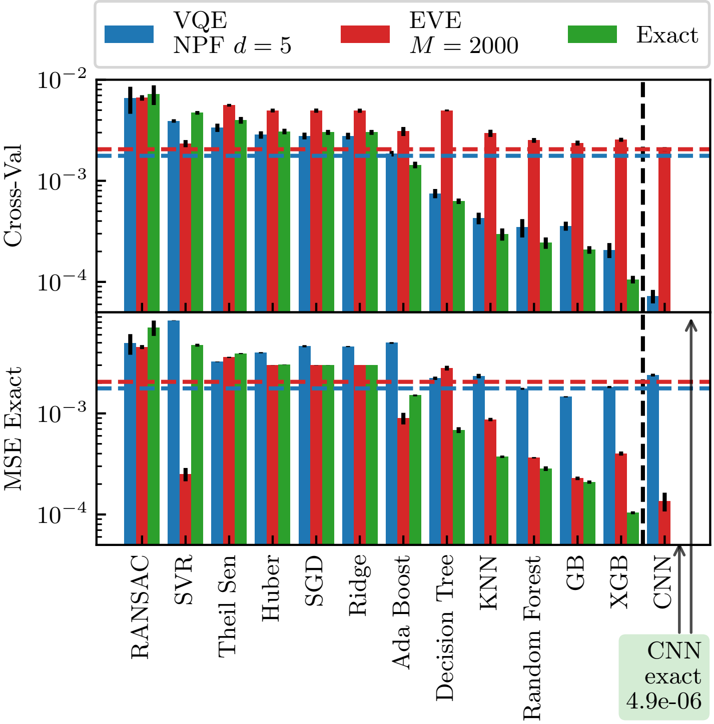


Figure 6.2: Model Selection: Comparison of the performance of various regression models, trained and validated using a variational quantum eigensolver (VQE) (NPF ansatz with depth $d = 5$), expectation value estimation (EVE) with $M = 2000$ shots, and exact data. The top panel displays the performance based on 5-fold cross validation MSE, computed by comparing the predictions made by models trained on individual folds against the (noisy) data across their respective validation sets. The bottom panel benchmarks the model performance by evaluation on exact densities from a test dataset and comparing them with the corresponding exact energies. Dashed lines indicate the baseline MSE of the noisy datasets used for training and validation against the exact data. Based on the cross-validation scores in the top panel, the CNN model from Ref. [86], separated by the black dashed line, outperforms all the out-of-the-box regression models we tested for the learning task at hand.

The comparative performance of the models is illustrated by the mean and standard deviation of their cross-validation scores, as shown in the top panel of Fig. 6.2. Additionally, both panels display the baseline mean-squared error (MSE) of the noisy data used for training and validation, indicated by dashed lines. The bottom panel of Fig. 6.2 illustrates the models' performance compared to the exact benchmark on a separate test set, generated with a distinct set of potentials not used during training. While this information is not directly used in the model selection process, it provides insights into the models' ability to generalize to noiseless data.

The cross-validation scores for models trained on EVE data never fall below the baseline error of the training data. This is because the validation dataset is also noisy, and with EVE's noise being randomly distributed with a zero mean, the model cannot effectively learn or predict the noise pattern. Among the models tested, those achieving the lowest validation scores on this dataset, aside from the optimized CNN model, are Support Vector Regression (SVR), Gradient Boosting (GB), Random Forest, and XGBoost (XGB). These models also perform the best in predicting noiseless data. We observe the characteristic performance that a cross-validation score close to the baseline of unbiased noise indicates that the underlying model has learned patterns that extrapolate to some extent to noiseless data.

The performance of models trained on VQE data differs noticeably when tested against exact data. As discussed in Section 6.3.2, the optimization error inherent in our VQE data can lead to an overestimation of the ground state energy for a given model instance. This results in biased noise that affects the training process. Consequently, the model inadvertently learns to fit this biased noise, which does not exist in the exact data, leading to a worse performance when testing the generalizability. When the MSE between the VQE-generated energies and the exact energies (indicated by the dashed blue line) is less than or equal to the predicted MSE from cross-validation, it suggests that the model has absorbed the bias present in the VQE data. Specifically, for models 7-13 trained on VQE data, the validation MSE is smaller than the baseline error of the training data. This observation implies that (1) there is an (expected) bias in the error of VQE and (2) these models learn to predict this bias.

Based on the cross validation data, the CNN model from Ref. [86] performs best on all datasets. While gradient boosting performs slightly better on the presented VQE dataset (NPF ansatz with $d = 5$), general performance across VQE datasets is at best comparable to the CNN model. We thus continue further analysis with the CNN model only.

6.5 Results

In this section, we benchmark the CNN model trained on noisy datasets, analyzing its performance as a function of both the type and magnitude of noise in the training data. We test the models on two tasks: energy prediction, discussed in Sec. 6.5.1, and density optimization, discussed in Sec. 6.5.2. All benchmarks are conducted using the same test set, consisting of 1,000 new problem instances with random potentials $\boldsymbol{\mu}_{\text{test}}$. These are solved with exact diagonalization obtaining a set of exact test densities $\boldsymbol{\rho}_{\text{test}}$ and energies F_{test}.

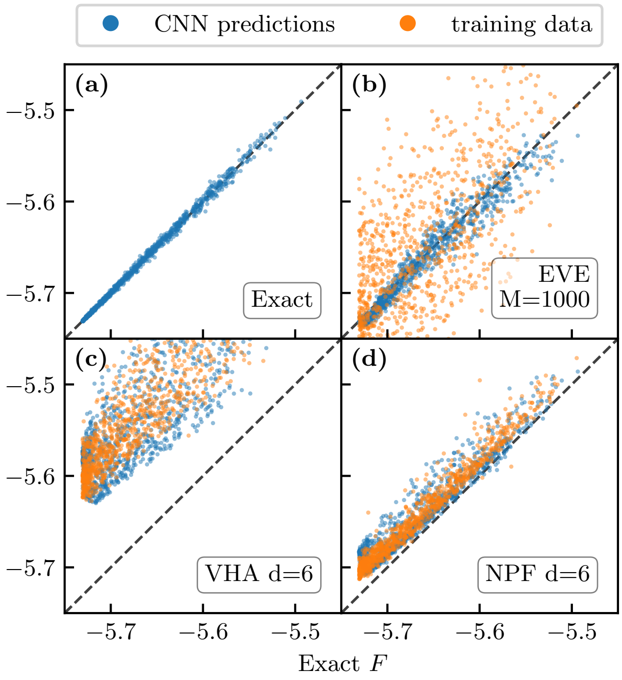


Figure 6.3: Noisy data and model predictions: Comparison of exact ground state kinetic-interaction energies F with training data and model predictions across four different methods for generating density-energy pairs. (a) Illustrates the model's ability to learn a functional from exact data, accurately capturing the relationship between density and ground state energies obtained from exact diagonalization. (b-d) Demonstrates the model's capacity to learn a functional which generalizes to the exact data, resulting in predictions that are on average closer to the ground truth than the raw data. These panels include: expectation value estimation (EVE) with $M = 1000$ shots and variational quantum eigensolver (VQE) with Variational Hamiltonian (VHA) / Number-Preserving Fabric (NPF) Ansatz of depth $d = 6$.

6.5.1 Energy prediction

In the energy prediction task, we assess the model's accuracy by comparing its predictions on the exact densities, $\mathcal{F}^{\mathrm{ML}}[\boldsymbol{\rho}_{\mathrm{test}}]$, against the exact energies F_{test}. Figure 6.3 summarizes this comparison for four models, each trained on datasets constructed with a different method: exact diagonalization, EVE ($M = 1000$), VHA-VQE ($d = 6$), and NPF-VQE ($d = 6$). The figure also reports the noisy training energies F, compared to the respective exact value. The model trained on exact data serves as a benchmark, representing an upper limit of performance. Models trained on the same amount of noisy data points are expected to perform less accurately.

The EVE data are spread uniformly around the diagonal line, as sampling noise is unbiased. The model trained on this data demonstrates improved energy predictions, which cluster closer to the diagonal line compared to the noisy training data.

Conversely, the VQE data suffer from a positive bias due to the limited expressive power of the ansatz. Variational methods always approximate the ground state energy from above, $\tilde{E}_{\text{VQE}} \geq E$. As a consequence, $\tilde{F}_{\text{VQE}}$ typically (albeit not always) overestimates F. We observe this bias is larger for smaller values of F, particularly for the $d = 6$ NPF ansatz. This effect is likely due to the variation in the strength of $\boldsymbol{\mu}$, which influences the efficiency of VQE – larger potentials cause stronger localization, thereby increasing kinetic and interaction energy as electrons are more confined. Localization also reduces ground state correlations, which can enhance the performance of limited-depth VQEs. The predictions of the models trained on VQE data are distributed similarly to their respective training data, indicating that the model learns the inherent bias.

To systematically compare models trained on datasets with a varying noise strengths, we summarize the results of the energy prediction task with the mean squared error (MSE)

$$\text{MSE}_{\mathcal{F}^{\text{ML}}} = \frac{1}{\text{test size}} \sum_{\text{test set}} |\mathcal{F}^{\text{ML}}[\boldsymbol{\rho}_{\text{test}}] - F_{\text{test}}|^2. \tag{6.15}$$

This is compared to the raw MSE of the method used to generate the training data, calculated directly from the training set as

$$\text{MSE}_X = \frac{1}{\text{train size}} \sum_{\text{train set}} |\tilde{F}_X - F|^2. \tag{6.16}$$

Figure 6.4 shows the performance of the model trained on EVE data as a function of the number of samples M. The squared error in the training data decreases proportionally to $1/M$, consistent with the standard sampling limit. (The factor 4 arises from the average ground-state variance of the two operators $\hat{T}$ and $\hat{U}$, which are sampled separately as discussed in Sec. 6.3.1.) The model's predictions improve on this MSE by up to factor 20 in the case of large sampling noise. As the sampling noise decreases, the model's performance saturates to that of the model trained on exact data. It is important to note that the total number of samples needed to generate the training dataset is $1000\,M$, uniformly distributed over the 1000 different problem instances. The improvement is due to the model's ability to synthesize the information from all these training instances, distilling the underlying trend while filtering out the unbiased noise.

Figure 6.5 presents the same performance benchmark for the model trained on VQE data. The MSE of the VQE, for both ansätze considered, decreases exponentially with the depth of the ansatz. Here, the depth d represents the

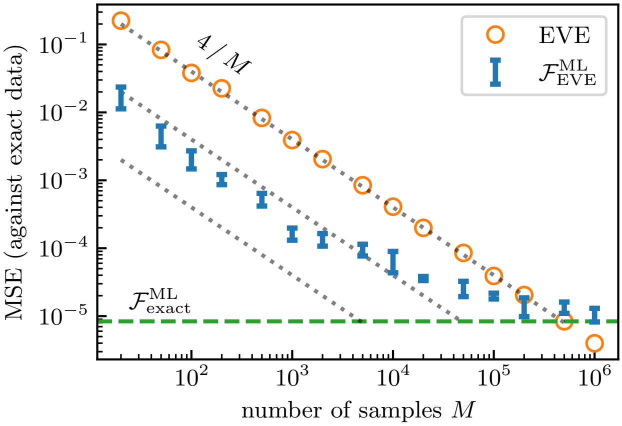


Figure 6.4: Learning improvement for sampling noise: Scaling of the Mean Squared Error (MSE) on F as a function of M, the number of sampling shots used in the data generation for the training set. The errors are evaluated against exact diagonalization data. The orange circles represent the MSE of the raw EVE data, demonstrating how the estimate variance decreases proportionally to $1/M$. The top dotted black line indicate this scaling, the others indicate a MSE reduction by one and two orders of magnitude. The dashed green line marks the baseline of the model trained and evaluated on exact data, providing a reference point for assessing the relative performance of the models. Blue markers show the MSE of the CNN model trained with EVE data, evaluated on exact densities from a test set and compared to the corresponding energies. The error bars indicate the standard deviation over the 5 cross-validation folds. The model predictions improve on the raw EVE error by more than an order of magnitude in the high-noise region, flattening to the baseline provided by the model trained on noiseless data when the noise becomes small enough.

number of repeated layers in the ansatz, while the upper-axis labels indicate the number of parameters. For depths larger than $d = 6$ the optimization of the VHA becomes significantly harder, thus we explore the larger-parameter-number regime with the less-structured NPF ansatz. The learned functional generalizes the training dataset successfully, showing a MSE comparable to the respective training data for all depths and both ansätze. Unlike the models trained on EVE, those trained on VQE data do not manage to learn the underlying functional while filtering out noise. This limitation is due to the bias intrinsic to the VQE errors, which is learned by the model. Although there is some quasi-stochastic error in the VQE results, caused by the optimization converging to different local minima of the energy, the dominant factor appears to be the biased error due to the limited expressibility of the ansatz.

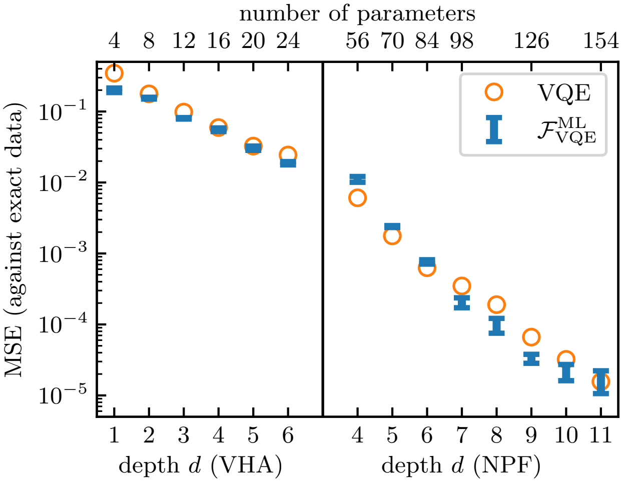


Figure 6.5: Models learn the bias from VQE data: Scaling of the Mean Squared Error (MSE) for different models evaluated against exact data as a function of d, the depth of the circuit used in the data generation for the training set, and the type of Ansatz VHA/NPF. The orange line represent the MSE of the raw VQE data, demonstrating how the error decreases as d increases. Red markers show the mean of 5-fold cross-validation performance (black bars represent the individual model performance) of machine learning models trained to predict VQE data and evaluated against exact data, highlighting some improvements against the raw data.

6.5.2 Density optimization

The main application of DFT functionals lie in solving new problem instances through density optimization. The goal of density optimization is equivalent to usual Kohn-Sham self-consistent field, but rather than optimizing a set of single-particle orbitals together with an exchange-correlation functional, we directly optimize the total energy functional in Eq. (6.3) with respect to the density $\boldsymbol{\rho}$ (subject to appropriate constraints). In this section we explore the performance in this application of the ML functional learned from noisy data using the CNN model.

For each test set potential $\boldsymbol{\mu}_{\text{test}}$ and each considered learned functional F^{ML}, we construct the total energy functional

$$\mathcal{E}^{\mathrm{ML}}[\boldsymbol{\rho}] = \mathcal{F}^{\mathrm{ML}}[\boldsymbol{\rho}] + \boldsymbol{\mu}_{\text{test}} \cdot \boldsymbol{\rho}. \tag{6.17}$$

We then minimize $\mathcal{E}^{\mathrm{ML}}[\boldsymbol{\rho}]$ under the constraints of $\rho_j \in [0, 2]$ for each site j (positive and bounded occupation) and $\sum_j \rho_j = 4$ (fixed total number of particles); this yields predictions for the ground state density $\boldsymbol{\rho}^*_{\mathrm{GS}}$ and energy $E^* = \mathcal{E}[\boldsymbol{\rho}^*]$.

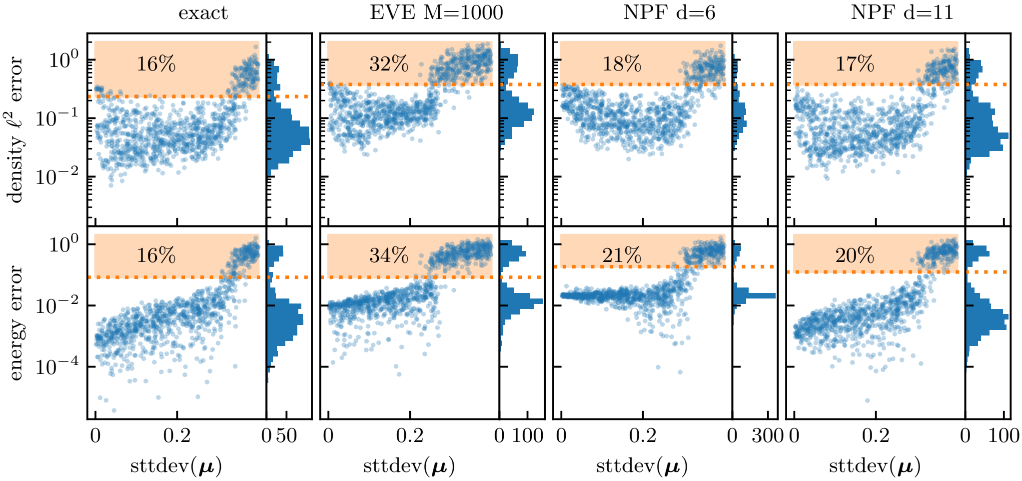


Figure 6.6: Application of the trained model to density optimization: Results of Kohn-Sham-like density optimization on a test set of 1000 new instances defined by random potentials, with the Hubbard functional defined by the CNN trained on various datasets. The performance is characterized by the density error in ℓ^2 norm $\|\boldsymbol{\rho}_{\text{opt}} - \boldsymbol{\rho}_{\text{exact}}\|_2$ (top row) and by the absolute error on the kinetic-interaction energy $|f[\boldsymbol{\rho}_{\text{opt}}] - F_{\text{exact}}|$ (bottom row). The scatter plots show the error dependence on the standard deviation of the potential $\|\boldsymbol{\mu} - \bar{\mu}\|_2$, and the histograms summarize the errors. The histograms show the errors tend to follow a bimodal distribution; the percentage of points in the higher-error mode of the distribution (shaded area above the dotted line) is annotated on each plot.

The minimization is performed using the Sequential Least Squares Programming (SLSQP) method implemented in the SciPy package [173], exploiting automatic differentiation of the CNN model to obtain analytical gradients of $\mathcal{E}[\boldsymbol{\rho}]$.

Figure 6.6 shows the resulting errors, for a selection of models trained on datasets constructed with different methods with different accuracy: exact diagonalization, EVE ($M = 1000$), NPF-VQE ($d = 6$), and NPF-VQE ($d = 11$). For energy, we consider the absolute deviation $|E^* - E_{\text{test}}|$, and for density we consider the error in ℓ^2-norm

$$\|\boldsymbol{\rho}^* - \boldsymbol{\rho}_{\text{test}}\|_2 = \sqrt{\sum\nolimits_j |\rho_j^* - \rho_{\text{test}\,j}|^2}. \tag{6.18}$$

In the scatter plots we show these errors for each of the test set instances versus the strength of the potential $\boldsymbol{\mu}$ — as the average value of the potential $\bar{\mu} = \sum_j \mu_j$ shifts the functional Eq. (6.17) by a constant and does not affect optimization, the strength of the potential is characterized as its standard deviation $\text{sttdev}(\boldsymbol{\mu}) = \|\boldsymbol{\mu} - \bar{\mu}\|_2$. To the right of each scatter plot, the distribution of the errors in summarized in a histogram. As all histograms display bimodal distributions, with two clearly separated lobes at lower- and higher-errors. On each histogram, we show the line that separates the two modes, and report the fraction of training data in the higher-error mode.

The results in the left panel of Fig. 6.6 show that, even when using the model trained on exact data, density optimization does not always converge to the correct result. In particular, for the largest potential strengths the optimization yields densities and energies with a large error. We attribute this to the limited availability of training points near the edges of the potential distribution. In the region of smaller potential strengths, at the center of the distribution of $\boldsymbol{\mu}$, has more data available; there the model can learn and generalize correctly. Farther from the center of the distribution, where the potential strength is larger, few data points are available for a wider area of possible cases to cover. We estimate 16% of the test set potentials are in this edge area; and a similar fraction of training points are available there. While this unavoidable for a continuous and unbounded distribution of problem instances, one can mitigate the problem by training the model on a set of potentials wider than the application range.

Conversely, for the smallest potentials, the density optimization sometimes yields correct energies and wrong densities. This occurs because, for very small potentials, the considered Hubbard model can exhibit two quasi-degenerate states with similar energies but significantly different densities. As a consequence $\mathcal{E}[\boldsymbol{\rho}]$ has multiple local minima for very distant densities, and the local optimizer we use sometimes finds a wrong density with a good associated energy.

Both of these effects, for the smallest and largest potential strengths, affect the results of the models trained using noisy data in a qualitatively similar way. Noisy training data causes an overall increase of error on the densities and energies, as well as shrinks the range of potentials on which the results of energy optimization are reliable. In the case of VQE with lower depth (NPF $d = 6$), we can see a concentration of the energies due to the large positive bias of the training data, which dominates the energy error.

6.6 Discussion and outlook

In this work, we demonstrated the application of machine learning models motivated by density functional theory to generalize and enhance a set of data subject to noise characteristic of near-term quantum algorithms. We showed that meaningful density functionals can be learned from a small amount of noisy training data and benchmarked the performance of these models in their usual DFT applications. Comparing the performance of the learned models to the respective training data, we showed a significant improvement is only achieved when the noise in the input data is unbiased (Fig. 6.4). This implies that learning techniques are well-suited to improve results from quantum algorithms that suffer from sampling noise. However, even when the trained models learn the inherent dataset bias, the proposed technique could be useful to generalize the dataset to new problem instances (Fig. 6.6). In the near future, quantum devices might be able to generate approximate solutions for quantum simulation problems that are beyond the reach of classical computation but remain too noisy to be directly useful. Nonetheless, this data could be valuable for supplementing the training of classical machine learning models, possibly in combination with other classically-generated data. In this context, it is important to characterize the effect of noise in quantum data on the trained models.

Our study focused on analyzing the (biased) VQE error and (unbiased) sampling noise separately. However, in practical implementations on quantum hardware, both sampling and optimization noise critically influence the performance and accuracy of VQE algorithms. Additionally, hardware noise is a significant factor. Investigating the combined effects of these noise sources presents a promising direction for future research.

Furthermore, we concentrated on learning from a very limited dataset, comprising only 1,000 training and validation instances. To better understand the trade-offs involved in achieving functionals with useful precision, it would be relevant to study the performance of learning as a function of both dataset size and problem size. Additionally, advancing towards realistic applications will require the development of techniques that enable learning from mixed datasets, combining results from approximate classical algorithms and quantum data. In this context, transfer learning might offer valuable insights [405].

In this study, we focused on learning Hohenberg-Kohn density functionals, a cornerstone of density functional theory with extensive literature support; however, the choice of learning targets can generally vary depending on the specific application. In the future, we envision beyond-classical quantum computations contributing valuable data to further enhance a wider range of existing machine learning-based functionals. Another potential learning target is the exchange-correlation functional used in Kohn-Sham density functional theory, which differs from our functional by a classical estimate of the kinetic energy expectation value, based on a product state. Alternatively, learning targets could be drawn from 1-particle reduced density matrix functional theory, where the underlying compressed representation of the quantum state is its 1-particle reduced density matrix [406]. Finally, one could consider directly learning the mapping $\boldsymbol{\mu} \to E$, also known as the Hohenberg-Kohn mapping [407]. A comparative study could help determine which of these learning targets is most resilient to noise in the data, offering better adaptability in real-world scenarios.

Bibliography

[1] C. E. Dykstra. *Theory and Applications of Computational Chemistry : The First Forty Years*. Elsevier, Amsterdam, 1st ed edition (2005). ISBN 0-444-51719-7 978-0-444-51719-7 0-444-51904-1 978-0-444-51904-7.

[2] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Reviews of Modern Physics*, **87**, 897–923 (2015).

[3] F. Neese, M. Atanasov, G. Bistoni, D. Maganas & S. Ye. Chemistry and Quantum Mechanics in 2019: Give Us Insight and Numbers. *Journal of the American Chemical Society*, **141**, 2814–2824 (2019).

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Quantum supremacy using a programmable superconducting processor. *Nature 2019 574:7779*, **574**, 505–510 (2019).

[5] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer et al. Quantum computational advantage with a programmable photonic processor. *Nature 2022 606:7912*, **606**, 75–81 (2022).

[6] H. S. Zhong, H. Wang, Y. H. Deng, M. C. Chen, L. C. Peng, Y. H. Luo et al. Quantum computational advantage using photons. *Science*, **370**, 1460–1463 (2020).

[7] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin & X. Yuan. Quantum computational chemistry. *Reviews of Modern Physics*, **92**, 015003 (2020).

[8] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker & M. Troyer. Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences*, **114**, 7555–7560 (2017).

[9] B. Bauer, S. Bravyi, M. Motta & G. K.-L. Chan. Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. *Chemical Reviews*, **120**, 12685–12717 (2020).

[10] J. J. Goings, A. White, J. Lee, C. S. Tautermann, M. Degroote, C. Gidney, T. Shiozaki, R. Babbush & N. C. Rubin. Reliably assessing the electronic structure of cytochrome P450 on today's classical computers and tomorrow's quantum computers. *Proceedings of the National Academy of Sciences*, **119**, e2203533119 (2022).

[11] V. Von Burg, G. H. Low, T. Häner, D. S. Steiger, M. Reiher, M. Roetteler & M. Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, **3**, 033055 (2021).

[12] M. Born & R. Oppenheimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, **389**, 457–484 (1927).

[13] K. N. Houk & F. Liu. Holy Grails for Computational Organic Chemistry and Biochemistry. *Accounts of Chemical Research*, **50**, 539–543 (2017).

[14] T. J. Lee & P. R. Taylor. A diagnostic for determining the quality of single-reference electron correlation methods. *International Journal of Quantum Chemistry*, **36**, 199–207 (1989).

[15] P. Hohenberg & W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, **136**, B864–B871 (1964).

[16] N. Schuch & F. Verstraete. Computational Complexity of interacting electrons and fundamental limitations of Density Functional Theory. *Nature Physics*, **5**, 732–735 (2009).

[17] S. Lehtola, C. Steigemann, M. J. T. Oliveira & M. A. L. Marques. Recent developments in libxc — A comprehensive library of functionals for density functional theory. *SoftwareX*, **7**, 1–5 (2018).

[18] W. Kohn & L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, **140**, A1133–A1138 (1965).

[19] E. R. Davidson. The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices. *Journal of Computational Physics*, **17**, 87–94 (1975).

[20] K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi & W. A. de Jong. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *The Journal of Chemical Physics*, **147**, 184111 (2017).

[21] T. Helgaker, P. Jørgensen & J. Olsen. *Molecular Electronic-Structure Theory*. Wiley, first edition (2000). ISBN 978-0-471-96755-2 978-1-119-01957-2.

[22] D. S. Levine, D. Hait, N. M. Tubman, S. Lehtola, K. B. Whaley & M. Head-Gordon. CASSCF with Extremely Large Active Spaces Using the Adaptive Sampling Configuration Interaction Method. *Journal of Chemical Theory and Computation*, **16**, 2340–2354 (2020).

[23] R. J. Bartlett & M. Musiał. Coupled-cluster theory in quantum chemistry. *Reviews of Modern Physics*, **79**, 291–352 (2007).

[24] A. Baiardi & M. Reiher. The density matrix renormalization group in chemistry and molecular physics: Recent developments and new challenges. *The Journal of Chemical Physics*, **152**, 040903 (2020).

[25] B. M. Austin, D. Y. Zubarev & W. A. J. Lester. Quantum Monte Carlo and Related Approaches. *Chemical Reviews*, **112**, 263–288 (2012).

[26] H. Ruocheng & S. Luber. Complete active space analysis of a reaction pathway: Investigation of the oxygen-oxygen bond formation. *Journal of computational chemistry*, **41** (2020).

[27] B. O. Roos et al. The complete active space self-consistent field method and its applications in electronic structure calculations. *Advances in chemical physics*, **69**, 399–445 (2007).

[28] D. J. Thouless. Stability conditions and nuclear rotations in the Hartree-Fock theory. *Nuclear Physics*, **21**, 225–232 (1960).

[29] B. O. Roos, P. R. Taylor & P. E. M. Sigbahn. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chemical Physics*, **48**, 157–173 (1980).

[30] S. Yalouz, B. Senjean, J. Günther, F. Buda, T. E. O'Brien & L. Visscher. A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states. *Quantum Science and Technology*, **6**, 024004 (2021).

[31] P. E. Siegbahn, J. Almlöf, A. Heiberg & B. O. Roos. The complete active space SCF (CASSCF) method in a Newton–Raphson formulation with application to the HNO molecule. *J. Chem. Phys.*, **74**, 2384–2396 (1981).

[32] Z.-L. Cai & J. R. Reimers. The Low-Lying Excited States of Pyridine. *The Journal of Physical Chemistry A*, **104**, 8389–8408 (2000).

[33] H. Lischka, D. Nachtigallová, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado & M. Barbatti. Multireference Approaches for Excited States of Molecules. *Chemical Reviews*, **118**, 7293–7361 (2018).

[34] Chr. Møller & M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review*, **46**, 618–622 (1934).

[35] Kerstin. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej & Krzysztof. Wolinski. Second-order perturbation theory with a CASSCF reference function. *The Journal of Physical Chemistry*, **94**, 5483–5488 (1990).

[36] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger & J. P. Malrieu. Introduction of n-electron valence states for multireference perturbation theory. *The Journal of Chemical Physics*, **114**, 10252–10264 (2001).

[37] C. Angeli, R. Cimiraglia & J. P. Malrieu. N-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants. *The Journal of Chemical Physics*, **117**, 9138–9153 (2002).

[38] B. M. Hoffman, D. Lukoyanov, Z.-Y. Yang, D. R. Dean & L. C. Seefeldt. Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical Reviews*, **114**, 4041–4062 (2014).

[39] R. Bjornsson, F. Neese, R. R. Schrock, O. Einsle & S. DeBeer. The discovery of Mo(III) in FeMoco: Reuniting enzyme and model chemistry. *JBIC Journal of Biological Inorganic Chemistry*, **20**, 447–460 (2015).

[40] J. Kempe, A. Kitaev & O. Regev. The Complexity of the Local Hamiltonian Problem. *SIAM Journal on Computing*, **35**, 1070–1097 (2006).

[41] S. Arora & B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 1 edition (2009). ISBN 978-0-521-42426-4 978-0-511-80409-0.

[42] G. K.-L. Chan. Quantum chemistry, classical heuristics, and quantum advantage, arXiv:2407.11235 (2024).

[43] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt et al. Evidence for the utility of quantum computing before fault tolerance. *Nature*, **618**, 500–505 (2023).

[44] D. Aharonov, M. Ben-Or, R. Impagliazzo & N. Nisan. Limitations of Noisy Reversible Computation, arXiv:quant-ph/9611028 (1996).

[45] D. Aharonov. A Quantum to Classical Phase Transition in Noisy Quantum Computers. *Physical Review A*, **62**, 062311 (2000).

[46] P. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 124–134 (1994).

[47] C. Gidney & M. Ekerå. How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. *Quantum*, **5**, 433 (2021).

[48] A. Yu. Kitaev. Quantum measurements and the Abelian Stabilizer Problem. *arXiv e-prints*, pages quant–ph/9511026 (1995).

[49] M. A. Nielsen & I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press, Cambridge (2010).

[50] L. Egan, D. M. Debroy, C. Noel, A. Risinger, D. Zhu, D. Biswas et al. Fault-tolerant operation of a quantum error-correction code. *ArXiv:2009.11482* (2020).

[51] R. Acharya, I. Aleiner, R. Allen, T. I. Andersen, M. Ansmann, F. Arute et al. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, **614**, 676–681 (2023).

[52] R. Acharya, L. Aghababaie-Beni, I. Aleiner, T. I. Andersen, M. Ansmann, F. Arute et al. Quantum error correction below the surface code threshold, arXiv:2408.13687 (2024).

[53] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, **2**, 79 (2018).

[54] A. Peruzzo, J. R. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik & J. L. O'Brien. A Variational Eigenvalue Solver on a Photonic Quantum Processor. *Nature Communications*, **5**, 4213 (2014).

[55] P. Jordan & E. Wigner. Über das paulische äquivalenzverbot. *Zeitschrift für Physik*, **47**, 631–651 (1928).

[56] J. T. Seeley, M. J. Richard & P. J. Love. The Bravyi-Kitaev transformation for quantum computation of electronic structure. *The Journal of Chemical Physics*, **137**, 224109 (2012).

[57] A. Tranter, S. Sofia, J. Seeley, M. Kaicher, J. McClean, R. Babbush, P. V. Coveney, F. Mintert, F. Wilhelm & P. J. Love. The Bravyi–Kitaev transformation: Properties and applications. *International Journal of Quantum Chemistry*, **115**, 1431–1441 (2015).

[58] Z. Jiang, A. Kalev, W. Mruczkiewicz & H. Neven. Optimal fermion-to-qubit mapping via ternary trees with applications to reduced quantum states learning. *Quantum*, **4**, 276 (2020).

[59] J. Cohn, M. Motta & R. M. Parrish. Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians. *PRX Quantum*, **2**, 040352 (2021).

[60] N. C. Rubin, R. Babbush & J. McClean. Application of fermionic marginal constraints to hybrid quantum algorithms. *New Journal of Physics*, **20**, 053020 (2018).

[61] X. Bonet-Monroig, R. Babbush & T. E. O'Brien. Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States. *Physical Review X*, **10**, 031064 (2020).

[62] A. G. Taube & R. J. Bartlett. New perspectives on unitary coupled-cluster theory. *International Journal of Quantum Chemistry*, **106**, 3393–3401 (2006).

[63] D. Wecker, M. B. Hastings & M. Troyer. Progress towards Practical Quantum Variational Algorithms. *Physical Review A*, **92**, 042303 (2015).

[64] G.-L. R. Anselmetti, D. Wierichs, C. Gogolin & R. M. Parrish. Local, expressive, quantum-number-preserving VQE ansätze for fermionic systems. *New Journal of Physics*, **23**, 113010 (2021).

[65] G. E. Crooks. Gradients of parameterized quantum gates using the parameter-shift rule and gate decomposition. *arXiv:1905.13311 [quant-ph]* (2019).

[66] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac & N. Killoran. Evaluating Analytic Gradients on Quantum Hardware. *Physical Review A*, **99**, 032331 (2019).

[67] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes & M. Cerezo. A Review of Barren Plateaus in Variational Quantum Computing, arXiv:2405.00781 (2024).

[68] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko & T. E. O'Brien. Performance comparison of optimization methods on variational quantum algorithms. *Physical Review A*, **107**, 032407 (2023).

[69] A. Aspuru-Guzik, A. D. Dutoi, P. J. Love & M. Head-Gordon. Simulated Quantum Computation of Molecular Energies. *Science*, **309**, 1704–1707 (2005).

[70] S. Deffner & S. Campbell. Quantum speed limits: From Heisenberg's uncertainty principle to optimal quantum control. *Journal of Physics A: Mathematical and Theoretical*, **50**, 453001 (2017).

[71] S. Lee, J. Lee, H. Zhai, Y. Tong, A. M. Dalzell, A. Kumar et al. Is there evidence for exponential quantum advantage in quantum chemistry? (2022).

[72] W. Kohn. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Reviews of Modern Physics*, **71**, 1253–1266 (1999).

[73] S. Fomichev, K. Hejazi, M. S. Zini, M. Kiser, J. F. Morales, P. A. M. Casares et al. Initial state preparation for quantum chemistry on quantum computers, arXiv:2310.18410 (2024).

[74] A. Delgado, P. A. M. Casares, R. dos Reis, M. S. Zini, R. Campos, N. Cruz-Hernández et al. Simulating key properties of lithium-ion batteries with a fault-tolerant quantum computer. *Physical Review A*, **106**, 032428 (2022).

[75] D. W. Berry, C. Gidney, M. Motta, J. R. McClean & R. Babbush. Qubitization of Arbitrary Basis Quantum Chemistry Leveraging Sparsity and Low Rank Factorization. *Quantum*, **3**, 208 (2019).

[76] J. Lee, D. W. Berry, C. Gidney, W. J. Huggins, J. R. McClean, N. Wiebe & R. Babbush. Even More Efficient Quantum Computations of Chemistry Through Tensor Hypercontraction. *PRX Quantum*, **2**, 030305 (2021).

[77] Y. Su, D. W. Berry, N. Wiebe, N. Rubin & R. Babbush. Fault-Tolerant Quantum Simulations of Chemistry in First Quantization. *PRX Quantum*, **2**, 040332 (2021).

[78] D. W. Berry, N. C. Rubin, A. O. Elnabawy, G. Ahlers, A. E. DePrince III, J. Lee, C. Gogolin & R. Babbush. Quantum Simulation of Realistic Materials in First Quantization Using Non-local Pseudopotentials (2023).

[79] E. Campbell. Random Compiler for Fast Hamiltonian Simulation. *Physical Review Letters*, **123**, 70503 (2019).

[80] A. M. Childs & N. Wiebe. Hamiltonian simulation using linear combinations of unitary operations. *Quantum Information and Computation*, **12**, 901–924 (2012).

[81] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari & R. D. Somma. Exponential Improvement in Precision for Simulating Sparse Hamiltonians. STOC '14, pages 283–292. Association for Computing Machinery, New York, NY, USA (2014). ISBN 978-1-4503-2710-7.

[82] G. H. Low & I. L. Chuang. Hamiltonian Simulation by Qubitization. *Quantum*, **3**, 163 (2019).

[83] G. Knizia & G. K.-L. Chan. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Physical review letters*, **109**, 186404 (2012).

[84] G. Knizia & G. K.-L. Chan. Density matrix embedding: A strong-coupling quantum embedding theory. *Journal of chemical theory and computation*, **9**, 1428–1432 (2013).

[85] S. Wouters, C. A. Jiménez-Hoyos, Q. Sun & G. K.-L. Chan. A practical guide to density matrix embedding theory in quantum chemistry. *Journal of chemical theory and computation*, **12**, 2706–2719 (2016).

[86] J. Nelson, R. Tiwari & S. Sanvito. Machine Learning Density Functional Theory for the Hubbard Model. *Physical Review B*, **99**, 075132 (2019).

[87] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová et al. Quantum Chemistry in the Age of Quantum Computing. *Chemical Reviews*, **119**, 10856–10915 (2019).

[88] D. S. Abrams & S. Lloyd. Simulation of many-body Fermi systems on a universal quantum computer. *Phys. Rev. Lett.*, **79**, 2586 (1997).

[89] D. S. Abrams & S. Lloyd. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.*, **83**, 5162 (1999).

[90] T. E. O'Brien, B. Tarasinski & B. M. Terhal. Quantum phase estimation of multiple eigenvalues for small-scale (noisy) experiments. *New Journal of Physics*, **21**, 23022 (2019).

[91] J. R. McClean, J. Romero, R. Babbush & A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, **18**, 023023 (2016).

[92] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii et al. Variational quantum algorithms. *Nature Reviews Physics*, **3**, 625–644 (2021).

[93] S. Endo, Z. Cai, S. C. Benjamin & X. Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan*, **90**, 032001 (2021).

[94] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, **94**, 015004 (2022).

[95] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter & W. A. de Jong. Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states. *Physical Review A*, **95**, 042308 (2017).

[96] J. Lee, W. J. Huggins, M. Head-Gordon & K. B. Whaley. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. *Journal of Chemical Theory and Computation*, **15**, 311–324 (2019).

[97] P. J. Ollitrault, A. Kandala, C. F. Chen, P. K. Barkoutsos, A. Mezzacapo, M. Pistoia, S. Sheldon, S. Woerner, J. M. Gambetta & I. Tavernelli. Quantum equation of motion for computing molecular excitation energies on a noisy quantum processor. *Physical Review Research*, **2**, 043140 (2020).

[98] K. M. Nakanishi, K. Mitarai & K. Fujii. Subspace-search variational quantum eigensolver for excited states. *Phys. Rev. Res.*, **1**, 033062 (2019).

[99] R. M. Parrish, E. G. Hohenstein, P. L. McMahon & T. J. Martinez. Hybrid Quantum/Classical Derivative Theory: Analytical Gradients and Excited-State Dynamics for the Multistate Contracted Variational Quantum Eigensolver. *arXiv:1906.08728* (2019).

[100] R. M. Parrish, E. G. Hohenstein, P. L. McMahon & T. J. Martínez. Quantum Computation of Electronic Transitions Using a Variational Quantum Eigensolver. *Physical Review Letters*, **122**, 230401 (2019).

[101] Y. Ibe, Y. O. Nakagawa, T. Yamamoto, K. Mitarai, Q. Gao & T. Kobayashi. Calculating transition amplitudes by variational quantum eigensolvers. *arXiv:2002.11724* (2020).

[102] O. Higgott, D. Wang & S. Brierley. Variational Quantum Computation of Excited States. *Quantum*, **3**, 156 (2019).

[103] T. Jones, S. Endo, S. McArdle, X. Yuan & S. C. Benjamin. Variational quantum algorithms for discovering Hamiltonian spectra. *Phys. Rev. A*, **99**, 062304 (2019).

[104] P. Jouzdani, S. Bringuier & M. Kostuk. A Method of Determining Excited-States for Quantum Computation. *arXiv:1908.05238* (2019).

[105] I. Kassal & A. Aspuru-Guzik. Quantum algorithm for molecular properties and geometry optimization. *J. Chem. Phys.*, **131**, 224102 (2009).

[106] T. E. O'Brien, B. Senjean, R. Sagastizabal, X. Bonet-Monroig, A. Dutkiewicz, F. Buda, L. DiCarlo & L. Visscher. Calculating energy derivatives for quantum chemistry on a quantum computer. *npj Quantum Information*, **5**, 1–12 (2019).

[107] K. Mitarai, Y. O. Nakagawa & W. Mizukami. Theory of analytical energy derivatives for the variational quantum eigensolver. *Phys. Rev. Res.*, **2**, 013129 (2020).

[108] I. O. Sokolov, P. K. Barkoutsos, L. Moeller, P. Suchsland, G. Mazzola & I. Tavernelli. Microcanonical and finite-temperature ab initio molecular dynamics simulations on quantum computers. *Phys. Rev. Research*, **3**, 013125 (2021).

[109] L. Bittel & M. Kliesch. Training Variational Quantum Algorithms Is NP-Hard. *Physical Review Letters*, **127**, 120502 (2021).

[110] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow & J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, **549**, 242–246 (2017).

[111] A. F. Izmaylov, T.-C. Yen & I. G. Ryabinkin. Revising the measurement process in the variational quantum eigensolver: Is it possible to reduce the number of separately measured operators? *Chem. Sci.*, **10**, 3746–3755 (2019).

[112] V. Verteletskyi, T.-C. Yen & A. F. Izmaylov. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *J. Chem. Phys.*, **152**, 124114 (2020).

[113] A. Jena, S. Genin & M. Mosca. Pauli partitioning with respect to gate sets. *arXiv:1907.07859* (2019).

[114] A. F. Izmaylov, T.-C. Yen, R. A. Lang & V. Verteletskyi. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method. *J. Chem. Theory Comput.*, **16**, 190–195 (2019).

[115] T.-C. Yen, V. Verteletskyi & A. F. Izmaylov. Measuring all compatible operators in one series of single-qubit measurements using unitary transformations. *Journal of chemical theory and computation*, **16**, 2400–2409 (2020).

[116] P. Gokhale, O. Angiuli, Y. Ding, K. Gui, T. Tomesh, M. Suchara, M. Martonosi & F. T. Chong. Minimizing state preparations in variational quantum eigensolver by partitioning into commuting families. *arXiv:1907.13623* (2019).

[117] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley & R. Babbush. Efficient and Noise Resilient Measurements for Quantum Chemistry on Near-Term Quantum Computers. *npj Quantum Information*, **7**, 1–9 (2021).

[118] A. Zhao, A. Tranter, W. M. Kirby, S. F. Ung, A. Miyake & P. J. Love. Measurement reduction in variational quantum algorithms. *Phys. Rev. A*, **101**, 062322 (2020).

[119] G. Wang, D. E. Koh, P. D. Johnson & Y. Cao. Minimizing Estimation Runtime on Noisy Quantum Computers. *PRX Quantum*, **2**, 010346 (2021).

[120] G. García-Pérez, M. A. Rossi, B. Sokolov, F. Tacchino, P. K. Barkoutsos, G. Mazzola, I. Tavernelli & S. Maniscalco. Learning to measure: Adaptive informationally complete POVMs for near-term quantum algorithms. *arXiv:2104.00569* (2021).

[121] J. R. McClean, R. Babbush, P. J. Love & A. Aspuru-Guzik. Exploiting Locality in Quantum Computation for Quantum Chemistry. *The Journal of Physical Chemistry Letters*, **5**, 4368–4380 (2014).

[122] R. Babbush, D. W. Berry, I. D. Kivlichan, A. Y. Wei, P. J. Love & A. Aspuru-Guzik. Exponentially more precise quantum simulation of fermions in second quantization. *New J. Phys.*, **18**, 033032 (2016).

[123] M. Motta, E. Ye, J. R. McClean, Z. Li, A. J. Minnich, R. Babbush & G. K. Chan. Low rank representations for quantum simulation of electronic structure. *arXiv:1808.02625* (2018).

[124] J. D. Whitfield, J. Biamonte & A. Aspuru-Guzik. Simulation of Electronic Structure Hamiltonians Using Quantum Computers. *Molecular Physics*, **109**, 735–750 (2011).

[125] Y. Ouyang, D. R. White & E. T. Campbell. Compilation by stochastic Hamiltonian sparsification. *Quantum*, **4** (2020).

[126] K. Kuroiwa & Y. O. Nakagawa. Penalty methods for a variational quantum eigensolver. *Phys. Rev. Research*, **3**, 013197 (2021).

[127] S. Saebo & P. Pulay. Fourth-order Mo/ller–Plessett perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.*, **86**, 914–922 (1987).

[128] S. Saebo & P. Pulay. Local treatment of electron correlation. *Ann. Rev. Phys. Chem.*, **44**, 213–236 (1993).

[129] M. Schütz, G. Hetzer & H.-J. Werner. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.*, **111**, 5691–5705 (1999).

[130] V. Weijo, P. Manninen, P. Jørgensen, O. Christiansen & J. Olsen. General biorthogonal projected bases as applied to second-order Møller-Plesset perturbation theory. *J. Chem. Phys.*, **127**, 074106 (2007).

[131] C. Hampel & H.-J. Werner. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.*, **104**, 6286–6297 (1996).

[132] M. Schütz & H.-J. Werner. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.*, **114**, 661–681 (2001).

[133] T. Korona & H.-J. Werner. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.*, **118**, 3006–3019 (2003).

[134] O. Christiansen, P. Manninen, P. Jørgensen & J. Olsen. Coupled-cluster theory in a projected atomic orbital basis. *J. Chem. Phys.*, **124**, 084103 (2006).

[135] D. Maynau, S. Evangelisti, N. Guihéry, C. J. Calzado & J.-P. Malrieu. Direct generation of local orbitals for multireference treatment and subsequent uses for the calculation of the correlation energy. *J. Chem. Phys.*, **116**, 10060–10068 (2002).

[136] C. Angeli, C. J. Calzado, R. Cimiraglia, S. Evangelisti, N. Guihéry, T. Leininger, J.-P. Malrieu, D. Maynau, J. V. P. Ruiz & M. Sparta. The use of local orbitals in multireference calculations. *Mol. Phys.*, **101**, 1389–1398 (2003).

[137] N. Ben Amor, F. Bessac, S. Hoyau & D. Maynau. Direct selected multireference configuration interaction calculations for large systems using localized orbitals. *J. Chem. Phys.*, **135**, 014101 (2011).

[138] C. Chang, C. J. Calzado, N. B. Amor, J. S. Marin & D. Maynau. Multi-scale multireference configuration interaction calculations for large systems using localized orbitals: Partition in zones. *J. Chem. Phys.*, **137**, 104102 (2012).

[139] A. S. P. Gomes & C. R. Jacob. Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems. *Annual Reports Section" C"(Physical Chemistry)*, **108**, 222–277 (2012).

[140] S. J. Bennie, M. Stella, T. F. Miller III & F. R. Manby. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *J. Chem. Phys.*, **143**, 024105 (2015).

[141] B. Hégely, P. R. Nagy, G. G. Ferenczy & M. Kállay. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.*, **145**, 064107 (2016).

[142] F. Libisch, M. Marsman, J. Burgdörfer & G. Kresse. Embedding for bulk systems using localized atomic orbitals. *J. Chem. Phys.*, **147**, 034110 (2017).

[143] D. V. Chulhai & J. D. Goodpaster. Improved accuracy and efficiency in quantum embedding through absolute localization. *J. Chem. Theory Comput.*, **13**, 1503–1508 (2017).

[144] E. R. Sayfutyarova, Q. Sun, G. K.-L. Chan & G. Knizia. Automated construction of molecular active spaces from atomic valence orbitals. *Journal of Chemical Theory and Computation*, **13**, 4063–4078 (2017).

[145] S. J. R. Lee, M. Welborn, F. R. Manby & T. F. Miller. Projection-Based Wavefunction-in-DFT Embedding. *Accounts of Chemical Research*, **52**, 1359–1368 (2019).

[146] X. Wen, D. S. Graham, D. V. Chulhai & J. D. Goodpaster. Absolutely Localized Projection-Based Embedding for Excited States. *Journal of Chemical Theory and Computation*, **16**, 385–398 (2020).

[147] D. Claudino & N. J. Mayhall. Automatic partition of orbital spaces based on singular value decomposition in the context of embedding theories. *J. Chem. Theory Comput.*, **15**, 1053–1064 (2019).

[148] D. Claudino & N. J. Mayhall. Simple and efficient truncation of virtual spaces in embedded wave functions via concentric localization. *J. Chem. Theory Comput.*, **15**, 6085–6096 (2019).

[149] M. R. Hermes & L. Gagliardi. Multiconfigurational Self-Consistent Field Theory with Density Matrix Embedding: The Localized Active Space Self-Consistent Field Method. *Journal of Chemical Theory and Computation*, **15**, 972–986 (2019).

[150] M. R. Hermes, R. Pandharkar & L. Gagliardi. The Variational Localized Active Space Self-Consistent Field Method. *J. Chem. Theory Comput.*, **16**, 4923–4937 (2020).

[151] I. Mayer. On Löwdin's method of symmetric orthogonalization*. *International Journal of Quantum Chemistry*, **90**, 63–65 (2002).

[152] D. Szczepanik & J. Mrozek. On several alternatives for Löwdin orthogonalization. *Computational and Theoretical Chemistry*, **1008**, 15–19 (2013).

[153] P.-O. Löwdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics*, **18**, 365–375 (1950).

[154] B. C. Carlson & J. M. Keller. Orthogonalization Procedures and the Localization of Wannier Functions. *Phys. Rev.*, **105**, 102–103 (1957).

[155] J. Pipek & P. G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *The Journal of Chemical Physics*, **90**, 4916–4926 (1989).

[156] J. M. Foster & S. F. Boys. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.*, **32**, 300–302 (1960).

[157] C. Edmiston & K. Ruedenberg. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.*, **35**, 457–464 (1963).

[158] R. C. Raffenetti, K. Ruedenberg, C. L. Janssen & H. F. Schaefer. Efficient use of Jacobi rotations for orbital optimization and localization. *Theoretica chimica acta*, **86**, 149–165 (1993).

[159] R. Barr & H. Basch. Improved convergence in orbital localization methods. *Chemical Physics Letters*, **32**, 537–540 (1975).

[160] S. Lehtola & H. Jónsson. Unitary Optimization of Localized Molecular Orbitals. *Journal of Chemical Theory and Computation*, **9**, 5365–5372 (2013).

[161] J. M. Leonard & W. L. Luken. Quadratically convergent calculation of localized molecular orbitals. *Theoretica chimica acta*, **62**, 107–132 (1982).

[162] I.-M. Høyvik, B. Jansik & P. Jørgensen. Trust region minimization of orbital localization functions. *J. Chem. Theory Comput.*, **8**, 3137–3146 (2012).

[163] Q. Sun. Co-iterative augmented Hessian method for orbital optimization. *arXiv:1610.08423* (2016).

[164] F. Jensen. *Introduction to Computational Chemistry*. John wiley & sons (2017).

[165] B. Jansík, S. Høst, K. Kristensen & P. Jørgensen. Local orbitals by minimizing powers of the orbital variance. *The Journal of Chemical Physics*, **134**, 194104 (2011).

[166] I.-M. Høyvik, B. Jansik & P. Jørgensen. Orbital localization using fourth central moment minimization. *J. Chem. Phys.*, **137**, 224114 (2012).

[167] S. Lehtola & H. Jónsson. Pipek–Mezey orbital localization using various partial charge estimates. *J. Chem. Theory Comput.*, **10**, 642–649 (2014).

[168] G. Knizia. Intrinsic atomic orbitals: An unbiased bridge between quantum theory and chemical concepts. *Journal of Chemical Theory and Computation*, **9**, 4834–4843 (2013).

[169] B. Senjean, S. Sen, M. Repisky, G. Knizia & L. Visscher. Generalization of Intrinsic Orbitals to Kramers-Paired Quaternion Spinors, Molecular Fragments, and Valence Virtual Spinors. *Journal of Chemical Theory and Computation*, **17**, 1337–1354 (2021).

[170] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders & T. Ziegler. Chemistry with ADF. *J. Comput. Chem.*, **22**, 931–967 (2001).

[171] E. Koridon, S. Yalouz, B. Senjean, F. Buda, T. E. O'Brien & L. Visscher. Orbital transformations to reduce the 1-norm of the electronic structure Hamiltonian for quantum computing applications. *Physical Review Research*, **3**, 033127 (2021).

[172] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li et al. PySCF: The Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **8**, e1340 (2018).

[173] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, **17**, 261–272 (2020).

[174] F. Gerber & R. Furrer. optimParallel: An R Package Providing a Parallel Version of the L-BFGS-B Optimization Method. *The R Journal*, **11**, 352–358 (2019).

[175] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao et al. OpenFermion: The electronic structure package for quantum computers. *Quantum Science and Technology*, **5**, 034014 (2020).

[176] F. A. Evangelista, G. K.-L. Chan & G. E. Scuseria. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. *The Journal of Chemical Physics*, **151**, 244112 (2019).

[177] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen & A. K. Wilson. Basis-set convergence in correlated calculations on Ne, N2 , and H2O. *Chem. Phys. Lett.*, **286**, 243–252 (1998).

[178] M. Yoshimine. Construction of the hamiltonian matrix in large configuration interaction calculations. *Journal of Computational Physics*, **11**, 449–454 (1973).

[179] M. Feyereisen, G. Fitzgerald & A. Komornicki. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chemical Physics Letters*, **208**, 359–363 (1993).

[180] V. Havlíček, M. Troyer & J. D. Whitfield. Operator locality in the quantum simulation of fermionic models. *Physical Review A*, **95**, 032332 (2017).

[181] J. W. Park, R. Al-Saadon, M. K. MacLeod, T. Shiozaki & B. Vlaisavljevich. Multireference electron correlation methods: Journeys along potential energy surfaces. *Chemical Reviews*, **120**, 5878–5909 (2020).

[182] P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka & R. Shepard. Multiconfiguration Self-Consistent Field and Multireference Configuration Interaction Methods and Applications. *Chemical Reviews*, **112**, 108–181 (2012).

[183] J. Olsen, B. O. Roos, P. Jo/rgensen & H. J. A. Jensen. Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces. *The Journal of chemical physics*, **89**, 2185–2192 (1988).

[184] P. Å. Malmqvist, A. Rendell & B. O. Roos. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *Journal of Physical Chemistry*, **94**, 5477–5482 (1990).

[185] T. Fleig, J. Olsen & L. Visscher. The generalized active space concept for the relativistic treatment of electron correlation. II. Large-scale configuration interaction implementation based on relativistic 2-and 4-spinors and its application. *The Journal of chemical physics*, **119**, 2963–2971 (2003).

[186] D. Ma, G. Li Manni & L. Gagliardi. The generalized active space concept in multiconfigurational self-consistent field methods. *The Journal of chemical physics*, **135** (2011).

[187] G. Li Manni, F. Aquilante & L. Gagliardi. Strong correlation treated via effective hamiltonians and perturbation theory. *The Journal of chemical physics*, **134** (2011).

[188] G. Li Manni, D. Ma, F. Aquilante, J. Olsen & L. Gagliardi. SplitGAS method for strong correlation and the challenging case of Cr2. *Journal of chemical theory and computation*, **9**, 3375–3384 (2013).

[189] J. E. Smith, B. Mussard, A. A. Holmes & S. Sharma. Cheap and near exact CASSCF with large active spaces. *Journal of chemical theory and computation*, **13**, 5468–5478 (2017).

[190] M. S. Gordon, D. G. Fedorov, S. R. Pruitt & L. V. Slipchenko. Fragmentation methods: A route to accurate calculations on large systems. *Chemical reviews*, **112**, 632–672 (2012).

[191] M. A. Collins & R. P. Bettens. Energy-based molecular fragmentation methods. *Chemical reviews*, **115**, 5607–5642 (2015).

[192] K. Raghavachari & A. Saha. Accurate composite and fragment-based quantum chemical models for large molecules. *Chemical reviews*, **115**, 5643–5677 (2015).

[193] D. G. Fedorov, Y. Alexeev & K. Kitaura. Geometry optimization of the active site of a large system with the fragment molecular orbital method. *The Journal of Physical Chemistry Letters*, **2**, 282–288 (2011).

[194] I. W. Bulik, G. E. Scuseria & J. Dukelsky. Density matrix embedding from broken symmetry lattice mean fields. *Physical Review B*, **89**, 035140 (2014).

[195] S. Yalouz, S. Sekaran, E. Fromager & M. Saubanère. Quantum embedding of multi-orbital fragments using the block-Householder transformation. *The Journal of Chemical Physics*, **157**, 214112 (2022).

[196] I. W. Bulik, W. Chen & G. E. Scuseria. Electron correlation in solids via density embedding theory. *The Journal of chemical physics*, **141** (2014).

[197] H. Q. Pham, V. Bernales & L. Gagliardi. Can density matrix embedding theory with the complete activate space self-consistent field solver describe single and double bond breaking in molecular systems? *Journal of chemical theory and computation*, **14**, 1960–1968 (2018).

[198] B.-X. Zheng & G. K.-L. Chan. Ground-state phase diagram of the square lattice Hubbard model from density matrix embedding theory. *Physical Review B*, **93**, 035126 (2016).

[199] Q. Chen, G. H. Booth, S. Sharma, G. Knizia & G. K.-L. Chan. Intermediate and spin-liquid phase of the half-filled honeycomb Hubbard model. *Physical Review B*, **89**, 165134 (2014).

[200] B.-X. Zheng, J. S. Kretchmer, H. Shi, S. Zhang & G. K.-L. Chan. Cluster size convergence of the density matrix embedding theory and its dynamical cluster formulation: A study with an auxiliary-field quantum Monte Carlo solver. *Physical Review B*, **95**, 045103 (2017).

[201] L. O. Jones, M. A. Mosquera, G. C. Schatz & M. A. Ratner. Embedding methods for quantum chemistry: Applications from materials to life sciences. *Journal of the American Chemical Society*, **142**, 3281–3295 (2020).

[202] S. Sen. *Computational Studies of Excited States of Chlorophyll Dimers in Light-Harvesting Complexes.* PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam (2023).

[203] S. M. Parker, T. Seideman, M. A. Ratner & T. Shiozaki. Communication: Active-space decomposition for molecular dimers. *The Journal of Chemical Physics*, **139**, 021108 (2013).

[204] C. A. Jiménez-Hoyos & G. E. Scuseria. Cluster-based mean-field and perturbative description of strongly correlated fermion systems: Application to the one-and two-dimensional Hubbard model. *Physical Review B*, **92**, 085101 (2015).

[205] L. Rummel & P. R. Schreiner. Advances and Prospects in Understanding London Dispersion Interactions in Molecular Chemistry. *Angewandte Chemie International Edition*, **63**, e202316364 (2024).

[206] E. Schrödinger. Quantisierung als eigenwertproblem. *Annalen der physik*, **385**, 437–490 (1926).

[207] V. Abraham & N. J. Mayhall. Selected configuration interaction in a basis of cluster state tensor products. *Journal of Chemical Theory and Computation*, **16**, 6098–6113 (2020).

[208] M. Hapka, M. Przybytek & K. Pernal. Symmetry-adapted perturbation theory based on multiconfigurational wave function description of monomers. *Journal of Chemical Theory and Computation*, **17**, 5538–5555 (2021).

[209] A. Papastathopoulos-Katsaros, C. A. Jiménez-Hoyos, T. M. Henderson & G. E. Scuseria. Coupled cluster and perturbation theories based on a cluster mean-field reference applied to strongly correlated spin systems. *Journal of Chemical Theory and Computation*, **18**, 4293–4303 (2022).

[210] E. Xu & S. Li. Block correlated second order perturbation theory with a generalized valence bond reference function. *The Journal of Chemical Physics*, **139** (2013).

[211] Q. Wang, M. Duan, E. Xu, J. Zou & S. Li. Describing strong correlation with block-correlated coupled cluster theory. *The Journal of Physical Chemistry Letters*, **11**, 7536–7543 (2020).

[212] N. He & F. A. Evangelista. A zeroth-order active-space frozen-orbital embedding scheme for multireference calculations. *The Journal of Chemical Physics*, **152** (2020).

[213] D. J. Coughtrie, R. Giereth, D. Kats, H. J. Werner & A. Köhn. Embedded Multireference Coupled Cluster Theory. *Journal of Chemical Theory and Computation*, **14**, 693–709 (2018).

[214] M. Otten, M. R. Hermes, R. Pandharkar, Y. Alexeev, S. K. Gray & L. Gagliardi. Localized Quantum Chemistry on Quantum Computers. *Journal of Chemical Theory and Computation*, **18**, 7205–7217 (2022).

[215] S. Sen, B. Senjean & L. Visscher. Characterization of excited states in time-dependent density functional theory using localized molecular orbitals. *The Journal of Chemical Physics*, **158**, 054115 (2023).

[216] DR. Yarkony & R. Silbey. The band gap in linear polyenes. *Chemical Physics*, **20**, 183–195 (1977).

[217] D. Ghosh, J. Hachmann, T. Yanai & G. K. Chan. Orbital optimization in the density matrix renormalization group, with applications to polyenes and β-carotene. *The Journal of chemical physics*, **128** (2008).

[218] S. R. Marder, W. E. Torruellas, M. Blanchard-Desce, V. Ricci, G. I. Stegeman, S. Gilmour, J.-L. Bredas, J. Li, G. U. Bublitz & S. G. Boxer. Large molecular third-order optical nonlinearities in polarized carotenoids. *Science*, **276**, 1233–1236 (1997).

[219] R. L. Christensen, E. A. Barney, R. D. Broene, M. G. I. Galinato & H. A. Frank. Linear polyenes: Models for the spectroscopy and photophysics of carotenoids. *Archives of biochemistry and biophysics*, **430**, 30–36 (2004).

[220] A. G. Barrett, T.-K. Ma & T. Mies. Recent developments in polyene cyclizations and their applications in natural product synthesis. *Synthesis*, **51**, 67–82 (2019).

[221] B. Senjean, S. Sen, M. Repisky, G. Knizia & L. Visscher (2020).

[222] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt et al. Recent Developments in the PySCF Program Package. *The Journal of Chemical Physics*, **153**, 024109 (2020).

[223] S. M. G. Sanfeliciano & J. M. Schaus. Rapid assessment of conformational preferences in biaryl and aryl carbonyl fragments. *PLOS ONE*, **13**, e0192974 (2018).

[224] O. A. Mosher, W. M. Flicker & A. Kuppermann. Triplet states in 1, 3-butadiene. *Chemical Physics Letters*, **19**, 332–333 (1973).

[225] Y. Guo, K. Sivalingam, C. Kollmar & F. Neese. Approximations of density matrices in N-electron valence state second-order perturbation theory (NEVPT2). II. The full rank NEVPT2 (FR-NEVPT2) formulation. *The Journal of Chemical Physics*, **154** (2021).

[226] Y. Guo, K. Sivalingam & F. Neese. Approximations of density matrices in N-electron valence state second-order perturbation theory (NEVPT2). I. Revisiting the NEVPT2 construction. *The Journal of Chemical Physics*, **154** (2021).

[227] K. Eichkorn, O. Treutler, H. Öhm, M. Häser & R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chemical physics letters*, **240**, 283–290 (1995).

[228] D. Zgid, D. Ghosh, E. Neuscamman & G. K. Chan. A study of cumulant approximations to n-electron valence multireference perturbation theory. *The Journal of chemical physics*, **130** (2009).

[229] C. Kollmar, K. Sivalingam, Y. Guo & F. Neese. An efficient implementation of the NEVPT2 and CASPT2 methods avoiding higher-order density matrices. *The Journal of Chemical Physics*, **155** (2021).

[230] K. Chatterjee & A. Y. Sokolov. Extended Second-Order Multireference Algebraic Diagrammatic Construction Theory for Charged Excitations. *Journal of Chemical Theory and Computation*, **16**, 6343–6357 (2020).

[231] J. Cioslowski. Connected moments expansion: A new tool for quantum many-body theory. *Phys. Rev. Lett.*, **58**, 83–85 (1987).

[232] A. Mahajan, N. S. Blunt, I. Sabzevari & S. Sharma. Multireference configuration interaction and perturbation theory without reduced density matrices. *The Journal of Chemical Physics*, **151** (2019).

[233] N. S. Blunt, A. Mahajan & S. Sharma. Efficient multireference perturbation theory without high-order reduced density matrices. *The Journal of Chemical Physics*, **153** (2020).

[234] R. J. Anderson, T. Shiozaki & G. H. Booth. Efficient and stochastic multireference perturbation theory for large active spaces within a full configuration interaction quantum Monte Carlo framework. *The Journal of chemical physics*, **152** (2020).

[235] F. D. Malone, R. M. Parrish, A. R. Welden, T. Fox, M. Degroote, E. Kyoseva, N. Moll, R. Santagati & M. Streif. Towards the simulation of large scale protein–ligand interactions on NISQ-era quantum computers. *Chemical Science*, **13**, 3094–3108 (2022).

[236] M. Loipersberger, F. D. Malone, A. R. Welden, R. M. Parrish, T. Fox, M. Degroote, E. Kyoseva, N. Moll, R. Santagati & M. Streif. Interaction Energies on Noisy Intermediate-Scale Quantum Computers (2022).

[237] C. L. Cortes, M. Loipersberger, R. M. Parrish, S. Morley-Short, W. Pol, S. Sim et al. Fault-tolerant quantum algorithm for symmetry-adapted perturbation theory (2023).

[238] A. Mitra, R. D'Cunha, Q. Wang, M. R. Hermes, Y. Alexeev, S. K. Gray, M. Otten & L. Gagliardi. The Localized Active Space Method with Unitary Selective Coupled Cluster. *Journal of Chemical Theory and Computation*, **20**, 7865–7875 (2024).

[239] R. J. Bartlett, S. A. Kucharski & J. Noga. Alternative Coupled-Cluster Ansätze II. The Unitary Coupled-Cluster Method. *Chemical Physics Letters*, **155**, 133–140 (1989).

[240] H. R. Grimsley, S. E. Economou, E. Barnes & N. J. Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications*, **10**, 3007 (2019).

[241] A. Tammaro, D. E. Galli, J. E. Rice & M. Motta. N-Electron Valence Perturbation Theory with Reference Wave Functions from Quantum Computing: Application to the Relative Stability of Hydroxide Anion and Hydroxyl Radical. *The Journal of Physical Chemistry A*, **127**, 817–827 (2023).

[242] M. Krompiec & D. M. Ramo. Strongly Contracted N-Electron Valence State Perturbation Theory Using Reduced Density Matrices from a Quantum Computer (2022).

[243] H.-Y. Huang, R. Kueng & J. Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, **16**, 1050–1057 (2020).

[244] K. Wan, W. J. Huggins, J. Lee & R. Babbush. Matchgate Shadows for Fermionic Quantum Simulation. *Communications in Mathematical Physics*, **404**, 629–700 (2023).

[245] A. Szabo & N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., Mineola, first edition (1996).

[246] M. J. Phipps, T. Fox, C. S. Tautermann & C.-K. Skylaris. Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. *Chemical society reviews*, **44**, 3177–3211 (2015).

[247] M. T. Albelda, J. C. Frías, E. García-España & H.-J. Schneider. Supramolecular complexation for environmental control. *Chemical Society Reviews*, **41**, 3859–3877 (2012).

[248] S. Sculfort & P. Braunstein. Intramolecular d10–d10 interactions in heterometallic clusters of the transition metals. *Chemical Society Reviews*, **40**, 2741–2760 (2011).

[249] P. Pyykkö, J. Li & N. Runeberg. Predicted ligand dependence of the Au (I)… Au (I) attraction in (XAuPH3) 2. *Chemical physics letters*, **218**, 133–138 (1994).

[250] A. Otero-de-la-Roza, J. D. Mallory & E. R. Johnson. Metallophilic interactions from dispersion-corrected density-functional theory. *The Journal of Chemical Physics*, **140** (2014).

[251] P. Verma & D. G. Truhlar. Status and challenges of density functional theory. *Trends Chem.*, **2**, 302–318 (2020).

[252] C. Adamo & D. Jacquemin. The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chem. Soc. Rev.*, **42**, 845–856 (2013).

[253] Z. Wang, C. Wu & W. Liu. NAC-TDDFT: Time-Dependent Density Functional Theory for Nonadiabatic Couplings. *Acc. Chem. Res.*, **54**, 3288–3297 (2021).

[254] C. A. Rozzi, F. Troiani & I. Tavernelli. Quantum modeling of ultrafast photoinduced charge separation. *J. Condens. Matter Phys.*, **30**, 013002 (2017).

[255] N. Minezawa & T. Nakajima. Trajectory surface hopping molecular dynamics simulation by spin-flip time-dependent density functional theory. *Journal of Chemical Physics*, **150**, 204120 (2019).

[256] S. Gozem, H. L. Luk, I. Schapiro & M. Olivucci. Theory and simulation of the ultrafast double-bond isomerization of biological chromophores. *Chem. Rev.*, **117**, 13502–13565 (2017).

[257] F. Agostini, S. K. Min, A. Abedi & E. K. U. Gross. Quantum-classical nonadiabatic dynamics: Coupled-vs independent-trajectory methods. *J. Chem. Theory Comput.*, **12**, 2127–2143 (2016).

[258] S. Faraji, S. Matsika & A. I. Krylov. Calculations of non-adiabatic couplings within equation-of-motion coupled-cluster framework: Theory, implementation, and validation against multi-reference methods. *J. Chem. Phys.*, **148**, 044103 (2018).

[259] S. Lee, Y. Horbatenko, M. Filatov & C. H. Choi. Fast and Accurate Computation of Nonadiabatic Coupling Matrix Elements Using the Truncated Leibniz Formula and Mixed-Reference Spin-Flip Time-Dependent Density Functional Theory. *J. Phys. Chem. Lett.*, **12**, 4722–4728 (2021).

[260] T. Shiozaki, W. Gy\Horffy, P. Celani & H.-J. Werner. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *J. Chem. Phys.*, **135**, 081106 (2011).

[261] A. A. Granovsky. Extended multi-configuration quasi-degenerate perturbation theory: The new approach to multi-state multi-reference perturbation theory. *J. Chem. Phys.*, **134**, 214113 (2011).

[262] J. W. Park. Analytical gradient theory for quasidegenerate N-electron valence state perturbation theory (QD-NEVPT2). *J. Chem. Theory Comput.*, **16**, 326–339 (2019).

[263] C. Angeli, S. Borini, M. Cestari & R. Cimiraglia. A quasidegenerate formulation of the second order n-electron valence state perturbation theory approach. *J. Chem. Phys.*, **121**, 4043–4049 (2004).

[264] S. Gozem, F. Melaccio, A. Valentini, M. Filatov, M. Huix-Rotllant, N. Ferré et al. Shape of multireference, equation-of-motion coupled-cluster, and density functional theory potential energy surfaces at a conical intersection. *J. Chem. Theory Comput.*, **10**, 3074–3084 (2014).

[265] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush & J. R. McClean. Increasing the Representation Accuracy of Quantum Simulations of Chemistry without Extra Quantum Resources. *Physical Review X*, **10**, 011004 (2020).

[266] B. H. Lengsfield, P. Saxe & D. R. Yarkony. On the evaluation of nonadiabatic coupling matrix elements using SA-MCSCF/CI wave functions and analytic gradient methods. I. *The Journal of Chemical Physics*, **81**, 4549–4553 (1984).

[267] J. Stålring, A. Bernhardsson & R. Lindh. Analytical gradients of a state average MCSCF state and a state average diagnostic. *MP*, **99**, 103–114 (2001).

[268] B. H. Lengsfield & D. R. Yarkony. Nonadiabatic interactions between potential energy surfaces: Theory and applications. *Adv. Chem. Phys.*, **82**, 1–71 (1992).

[269] D. Yarkony. *Modern Electronic Structure Theory*, volume 2. World Scientific (1995).

[270] J. W. Snyder Jr, E. G. Hohenstein, N. Luehr & T. J. Martínez. An atomic orbital-based formulation of analytical gradients and nonadiabatic coupling vector elements for the state-averaged complete active space self-consistent field method on graphical processing units. *J. Chem. Phys.*, **143**, 154107 (2015).

[271] J. W. Snyder Jr, B. S. Fales, E. G. Hohenstein, B. G. Levine & T. J. Martínez. A direct-compatible formulation of the coupled perturbed complete active space self-consistent field equations on graphical processing units. *J. Chem. Phys.*, **146**, 174113 (2017).

[272] I. Fdez. Galván, M. G. Delcey, T. B. Pedersen, F. Aquilante & R. Lindh. Analytical state-average complete-active-space self-consistent field nonadiabatic coupling vectors: Implementation with density-fitted two-electron integrals and application to conical intersections. *J. Chem. Theory Comput.*, **12**, 3636–3653 (2016).

[273] F. Gatti, B. Lasorne, H.-D. Meyer & A. Nauts. Vibronic Couplings. In *Applications of Quantum Dynamics in Chemistry*, pages 81–109. Springer (2017).

[274] M. Baer. *Beyond Born-Oppenheimer: Conical Intersections and Electronic Nonadiabatic Coupling Terms.* Wiley Online Library (2006).

[275] E. K. Gross, L. N. Oliveira & W. Kohn. Rayleigh-Ritz variational principle for ensembles of fractionally occupied states. *Phys. Rev. A*, **37**, 2805 (1988).

[276] J. Wen, D. Lv, M.-H. Yung & G.-L. Long. Variational Quantum Packaged Deflation for Arbitrary Excited States. *Quantum Eng.*, page e80 (2021).

[277] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong & I. Siddiqi. Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm. *Physical Review X*, **8**, 011021 (2018).

[278] M. Motta, C. Sun, A. T. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. Brandão & G. K.-L. Chan. Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat. Phys.*, **16**, 205–210 (2020).

[279] R. M. Parrish & P. L. McMahon. Quantum filter diagonalization: Quantum eigendecomposition without full quantum phase estimation. *arXiv:1909.08925* (2019).

[280] D.-B. Zhang, Z.-H. Yuan & T. Yin. Variational quantum eigensolvers by variance minimization. *arXiv:2006.15781* (2020).

[281] T. A. Bespalova & O. Kyriienko. Hamiltonian operator approximation for energy measurement and ground state preparation. *arXiv:2009.03351* (2020).

[282] U. Azad & H. Singh. Quantum Chemistry Calculations using Energy Derivatives on Quantum Computers. *arXiv:2106.06463* (2021).

[283] W. Mizukami, K. Mitarai, Y. O. Nakagawa, T. Yamamoto, T. Yan & Y.-y. Ohnishi. Orbital optimized unitary coupled cluster theory for quantum computer. *Physical Review Research*, **2**, 33421 (2020).

[284] T. U. Helgaker & J. Almlöf. A second-quantization approach to the analytical evaluation of response properties for perturbation-dependent basis sets. *Int. J. Quantum Chem.*, **26**, 275–291 (1984).

[285] K. Mitarai, M. Negoro, M. Kitagawa & K. Fujii. Quantum circuit learning. *Phys. Rev. A*, **98**, 032309 (2018).

[286] S. Tamiya, S. Koh & Y. O. Nakagawa. Calculating nonadiabatic couplings and Berry's phase by variational quantum eigensolvers. *Phys. Rev. Research*, **3**, 023244 (2021).

[287] M. Chahre. Trigger and Amplification Mechanisms in Visual Phototransduction. *Annual Review of Biophysics and Biophysical Chemistry*, **14**, 331–360 (1985).

[288] R. R. Birge. Nature of the Primary Photochemical Events in Rhodopsin and Bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1016**, 293–327 (1990).

[289] C. Schnedermann, X. Yang, M. Liebel, KM. Spillane, J. Lugtenburg, I. Fernández et al. Evidence for a vibrational phase-dependent isotope effect on the photochemistry of vision. *Nature chemistry*, **10**, 449–455 (2018).

[290] I. Fdez. Galván, M. Vacher, A. Alavi, C. Angeli, F. Aquilante, J. Autschbach et al. OpenMolcas: From Source Code to Insight. *Journal of Chemical Theory and Computation*, **15**, 5925–5964 (2019).

[291] D. G. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.*, **152**, 184108 (2020).

[292] C. Developers. Cirq (2021).

[293] D. R. Yarkony. Systematic determination of intersections of potential energy surfaces using a Lagrange multiplier constrained procedure. *J. Phys. Chem.*, **97**, 4407–4412 (1993).

[294] D. R. Yarkony. Theoretical studies of spin-forbidden radiationless decay in polyatomic systems: Insights from recently developed computational methods. *JACS*, **114**, 5406–5411 (1992).

[295] M. J. Bearpark, M. A. Robb & H. B. Schlegel. A direct method for the location of the lowest energy point on a potential surface crossing. *Chem. Phys. Lett.*, **223**, 269–274 (1994).

[296] T. W. Keal, A. Koslowski & W. Thiel. Comparison of algorithms for conical intersection optimisation using semiempirical methods. *Theoretical Chemistry Accounts*, **118**, 837–844 (2007).

[297] F. Sicilia, L. Blancafort, M. J. Bearpark & M. A. Robb. New Algorithms for Optimizing and Linking Conical Intersection Points. *J. Chem. Theory Comput.*, **4**, 257–266 (2008).

[298] LS. Cederbaum, J. Schirmer & H.-D. Meyer. Block diagonalisation of Hermitian matrices. *J. Phys. A Math. Theor.*, **22**, 2427 (1989).

[299] M. Richter, P. Marquetand, J. González-Vázquez, I. Sola & L. González. SHARC: Ab initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary Couplings. *J. Chem. Theory Comput.*, **7**, 1253–1258 (2011).

[300] J. C. Tully. Molecular dynamics with electronic transitions. *J. Chem. Phys.*, **93**, 1061–1071 (1998).

[301] X. Andrade, A. Castro, D. Zueco & J. L. Alonso. Modified Ehrenfest Formalism for Efficient Large-Scale ab initio Molecular Dynamics. *J. Chem. Theory Comput.*, **5**, 728—742 (2009).

[302] Y. Li & S. C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Phys. Rev. X*, **7**, 021050 (2017).

[303] A. Mari, T. R. Bromley & N. Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *arXiv:2008.06517* (2020).

[304] J. J. Meyer. Gradients just got more flexible. *Quantum Views*, **5**, 50 (2021).

[305] T. Hubregtsen, F. Wilde, S. Qasim & J. Eisert. Single-component gradient rules for variational quantum algorithms. *arXiv:2106.01388* (2021).

[306] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love & A. Aspuru-Guzik. Strategies for Quantum Computing Molecular Energies Using the Unitary Coupled Cluster Ansatz. *Quantum Science and Technology*, **4**, 014008 (2018).

[307] J. S. Kottmann, A. Anand & A. Aspuru-Guzik. A feasible approach for automatically differentiable unitary coupled-cluster on quantum computers. *Chemical Science*, **12**, 3497–3508 (2021).

[308] L. Banchi & G. E. Crooks. Measuring Analytic Gradients of General Quantum Evolution with the Stochastic Parameter Shift Rule. *arXiv:2005.10299* (2020).

[309] D. Wierichs, J. Izaac, C. Wang & C. Y.-Y. Lin. General parameter-shift rules for quantum gradients. *Quantum*, **6**, 677 (2022).

[310] A. F. Izmaylov, R. A. Lang & T.-C. Yen. Analytic gradients in variational quantum algorithms: Algebraic extensions of the parameter-shift rule to general unitary transformations. *arXiv:2107.08131* (2021).

[311] T. Helgaker & P. Jørgensen. Analytical calculation of geometrical derivatives in molecular electronic structure theory. *Adv. Quantum Chem.*, **19**, 183–245 (1988).

[312] J. Simons, P. Jørgensen & T. U. Helgaker. Higher molecular-deformation derivatives of the configuration-interaction energy. *Chem. Phys.*, **86**, 413–432 (1984).

[313] A. K. Geim & K. S. Novoselov. The Rise of Graphene. *Nature Materials*, **6**, 183–191 (2007).

[314] M. V. Berry. Quantal Phase Factors Accompanying Adiabatic Changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, **392**, 45–57 (1984).

[315] W. Domcke, D. Yarkony & H. Köppel, editors. *Conical Intersections: Theory, Computation and Experiment*. Number v. 17 in Advanced Series in Physical Chemistry. World Scientific, Singapore ; Hackensack, NJ (2011). ISBN 978-981-4313-44-5.

[316] D. R. Yarkony. Nonadiabatic Quantum Chemistry—Past, Present, and Future. *Chemical Reviews*, **112**, 481–498 (2012).

[317] D. Polli, P. Altoè, O. Weingart, K. M. Spillane, C. Manzoni, D. Brida et al. Conical Intersection Dynamics of the Primary Photoisomerization Event in Vision. *Nature*, **467**, 440–443 (2010).

[318] G. Olaso-González, M. Merchán & L. Serrano-Andrés. Ultrafast Electron Transfer in Photosynthesis: Reduced Pheophytin and Quinone Interaction Mediated by Conical Intersections. *The Journal of Physical Chemistry B*, **110**, 24734–24739 (2006).

[319] H. E. Zimmerman. Molecular Orbital Correlation Diagrams, Mobius Systems, and Factors Controlling Ground- and Excited-State Reactions. II. *Journal of the American Chemical Society*, **88**, 1566–1567 (1966).

[320] F. Bernardi, M. Olivucci & M. A. Robb. Potential energy surface crossings in organic photochemistry. *Chemical Society Reviews*, **25**, 321–328 (1996).

[321] L. González, D. Escudero & L. Serrano-Andrés. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem*, **13**, 28–51 (2012).

[322] R. P. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, **21**, 467–488 (1982).

[323] D. Wecker, M. B. Hastings, N. Wiebe, B. K. Clark, C. Nayak & M. Troyer. Solving strongly correlated electron models on a quantum computer. *Physical Review A*, **92**, 062318 (2015).

[324] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush & H. Neven. Barren Plateaus in Quantum Neural Network Training Landscapes. *Nature Communications*, **9**, 4812 (2018).

[325] X. Xiao, J. K. Freericks & A. F. Kemper. Robust Measurement of Wave Function Topology on NISQ Quantum Computers (2022).

[326] B. Murta, G. Catarina & J. Fernández-Rossier. Berry phase estimation in gate-based adiabatic quantum simulation. *Phys. Rev. A*, **101**, 020302 (2020).

[327] H. C. Longuet-Higgins, U. Öpik, M. H. L. Pryce & R. A. Sack. Studies of the Jahn-Teller Effect .II. The Dynamical Problem. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **244**, 1–16 (1958).

[328] C. A. Mead & D. G. Truhlar. On the Determination of Born–Oppenheimer Nuclear Motion Wave Functions Including Complications Due to Conical Intersections and Identical Nuclei. *The Journal of Chemical Physics*, **70**, 2284–2296 (1979).

[329] I. G. Ryabinkin, L. Joubert-Doriol & A. F. Izmaylov. Geometric Phase Effects in Nonadiabatic Dynamics near Conical Intersections. *Accounts of Chemical Research*, **50**, 1785–1793 (2017).

[330] J. Whitlow, Z. Jia, Y. Wang, C. Fang, J. Kim & K. R. Brown. Simulating Conical Intersections with Trapped Ions (2023).

[331] C. H. Valahu, V. C. Olaya-Agudelo, R. J. MacDonell, T. Navickas, A. D. Rao, M. J. Millican et al. Direct Observation of Geometric Phase in Dynamics around a Conical Intersection. *Nature Chemistry*, **15**, 1503–1508 (2023).

[332] C. S. Wang, N. E. Frattini, B. J. Chapman, S. Puri, S. M. Girvin, M. H. Devoret & R. J. Schoelkopf. Observation of Wave-Packet Branching through an Engineered Conical Intersection. *Physical Review X*, **13**, 011008 (2023).

[333] E. Koridon & S. Polla. Auto_oo: An autodifferentiable framework for molecular orbital-optimized variational quantum algorithms (2024).

[334] E. Teller. The Crossing of Potential Surfaces. *The Journal of Physical Chemistry*, **41**, 109–116 (1937).

[335] G. Herzberg & H. C. Longuet-Higgins. Intersection of Potential Energy Surfaces in Polyatomic Molecules. *Discussions of the Faraday Society*, **35**, 77–82 (1963).

[336] R. Broer, L. Hozoi & W. C. Nieuwpoort. Non-Orthogonal Approaches to the Study of Magnetic Interactions. *Molecular Physics*, **101**, 233–240 (2003).

[337] V. Veryazov, P. Å. Malmqvist & B. O. Roos. How to select active space for multiconfigurational quantum chemistry? *International Journal of Quantum Chemistry*, **111**, 3329–3338 (2011).

[338] D. R. Yarkony. Diabolical Conical Intersections. *Reviews of Modern Physics*, **68**, 985–1013 (1996).

[339] C. Alden Mead. The Molecular Aharonov—Bohm Effect in Bound States. *Chemical Physics*, **49**, 23–32 (1980).

[340] S. M. Harwood, D. Trenev, S. T. Stober, P. Barkoutsos, T. P. Gujarati, S. Mostame & D. Greenberg. Improving the Variational Quantum Eigensolver Using Variational Adiabatic Quantum Computing. *ACM Transactions on Quantum Computing*, **3**, 1:1–1:20 (2022).

[341] C. A. Mead. The ”noncrossing” Rule for Electronic Potential Energy Surfaces: The Role of Time-reversal Invariance. *The Journal of Chemical Physics*, **70**, 2276–2283 (1979).

[342] H. J. A. Jensen & P. Jorgensen. A Direct Approach to Second-order MCSCF Calculations Using a Norm Extended Optimization Scheme. *The Journal of Chemical Physics*, **80**, 1204–1214 (1984).

[343] B. Helmich-Paris. A Trust-Region Augmented Hessian Implementation for Restricted and Unrestricted Hartree–Fock and Kohn–Sham Methods. *The Journal of Chemical Physics*, **154**, 164104 (2021).

[344] T. E. O’Brien, S. Polla, N. C. Rubin, W. J. Huggins, S. McArdle, S. Boixo, J. R. McClean & R. Babbush. Error Mitigation via Verified Phase Estimation. *PRX Quantum*, **2** (2021).

[345] S. Polla, G.-L. R. Anselmetti & T. E. O'Brien. Optimizing the Information Extracted by a Single Qubit Measurement. *Physical Review A*, **108**, 012403 (2023).

[346] J. Nocedal & S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, 2nd ed edition (2006). ISBN 978-0-387-30303-1.

[347] E. P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, **62**, 548–564 (1955).

[348] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda & L. Visscher. Analytical Nonadiabatic Couplings and Gradients within the State-Averaged Orbital-Optimized Variational Quantum Eigensolver. *Journal of Chemical Theory and Computation*, **18**, 776–794 (2022).

[349] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends et al. Hartree-Fock on a Superconducting Qubit Quantum Computer. *Science*, **369**, 1084–1089 (2020).

[350] P. Huembeli & A. Dauphin. Characterizing the Loss Landscape of Variational Quantum Circuits. *Quantum Science and Technology*, **6**, 025011 (2021).

[351] H. Hirai. Excited-state molecular dynamics simulation based on variational quantum algorithms. *Chemical Physics Letters*, **816**, 140404 (2023).

[352] V. Bonačić-Koutecký & J. Michl. Photochemicalsyn-Anti Isomerization of a Schiff Base: A Two-Dimensional Description of a Conical Intersection in Formaldimine. *Theoretica chimica acta*, **68**, 45–55 (1985).

[353] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, S. Ahmed, V. Ajith et al. PennyLane: Automatic Differentiation of Hybrid Quantum-Classical Computations (2022).

[354] J. Stokes, J. Izaac, N. Killoran & G. Carleo. Quantum Natural Gradient. *Quantum*, **4**, 269 (2020).

[355] J. J. Meyer. Fisher Information in Noisy Intermediate-Scale Quantum Applications. *Quantum*, **5**, 539 (2021).

[356] S.-i. Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, **10**, 251–276 (1998).

[357] T. Liang, T. Poggio, A. Rakhlin & J. Stokes. Fisher-Rao Metric, Geometry, and Complexity of Neural Networks (2019).

[358] E. G. Hohenstein, O. Oumarou, R. Al-Saadon, G.-L. R. Anselmetti, M. Scheurer, C. Gogolin & R. M. Parrish. Efficient Quantum Analytic Nuclear Gradients with Double Factorization (2022).

[359] J. K. Asóth, L. Oroszlány & A. Pályi. *A Short Course on Topological Insulators: Band Structure and Edge States in One and Two Dimensions*. Springer (2016). ISBN 978-3-319-25607-8 978-3-319-25605-4.

[360] J. Zak. Berry's phase for energy bands in solids. *Phys. Rev. Lett.*, **62**, 2747–2750 (1989).

[361] Y. Hatsugai. Quantized Berry Phases as a Local Order Parameter of a Quantum Liquid. *Journal of the Physical Society of Japan*, **75**, 123601 (2006).

[362] T. Fukui, Y. Hatsugai & H. Suzuki. Chern Numbers in Discretized Brillouin Zone: Efficient Method of Computing (Spin) Hall Conductances. *Journal of the Physical Society of Japan*, **74**, 1674–1677 (2005).

[363] S.-s. Chern. Characteristic Classes of Hermitian Manifolds. *Annals of Mathematics*, **47**, 85–121 (1946).

[364] R. Citro & M. Aidelsburger. Thouless Pumping and Topology. *Nature Reviews Physics*, **5**, 87–101 (2023).

[365] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean & T. E. O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, **95**, 045005 (2023).

[366] K. Burke & L. O. Wagner. DFT in a nutshell. *International Journal of Quantum Chemistry*, **113**, 96–101 (2013).

[367] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller & K. Burke. Finding Density Functionals with Machine Learning. *Physical Review Letters*, **108**, 253002 (2012).

[368] L. Li, J. C. Snyder, I. M. Pelaschier, J. Huang, U.-N. Niranjan, P. Duncan, M. Rupp, K.-R. Müller & K. Burke. Understanding machine-learned density functionals. *International Journal of Quantum Chemistry*, **116**, 819–833 (2016).

[369] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf & M. Ceriotti. Transferable Machine-Learning Model of the Electron Density. *ACS Central Science*, **5**, 57–64 (2019).

[370] K. Ryczko, D. A. Strubbe & I. Tamblyn. Deep Learning and Density-Functional Theory. *Physical Review A*, **100**, 022512 (2019).

[371] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, **374**, 1385–1389 (2021).

[372] A. J. Cohen, P. Mori-Sánchez & W. Yang. Challenges for Density Functional Theory. *Chemical Reviews*, **112**, 289–320 (2012).

[373] A. S. Christensen, T. Kubař, Q. Cui & M. Elstner. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chemical Reviews*, **116**, 5301–5337 (2016).

[374] T. E. Baker & D. Poulin. Density functionals and Kohn-Sham potentials with minimal wavefunction preparations on a quantum computer. *Physical Review Research*, **2**, 043238 (2020).

[375] E. Sheridan, L. Mineh, R. A. Santos & T. Cubitt. Enhancing density functional theory using the variational quantum eigensolver (2024).

[376] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven & J. R. McClean. Power of data in quantum machine learning. *Nature Communications*, **12**, 2631 (2021).

[377] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert & J. Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, **377**, eabk3333 (2022).

[378] L. Lewis, H.-Y. Huang, V. T. Tran, S. Lehner, R. Kueng & J. Preskill. Improved machine learning algorithm for predicting ground state properties. *Nature Communications*, **15**, 895 (2024).

[379] A. Dawid, J. Arnold, B. Requena, A. Gresch, M. Płodzień, K. Donatella et al. Modern applications of machine learning in quantum sciences (2023).

[380] S. Jerbi, C. Gyurik, S. C. Marshall, R. Molteni & V. Dunjko. Shadows of quantum machine learning. *Nature Communications*, **15**, 5676 (2024).

[381] H. Liao, D. S. Wang, I. Sitdikov, C. Salcedo, A. Seif & Z. K. Minev. Machine Learning for Practical Quantum Error Mitigation (2023).

[382] J. P. F. LeBlanc, A. E. Antipov, F. Becca, I. W. Bulik, G. K.-L. Chan, C.-M. Chung et al. Solutions of the Two-Dimensional Hubbard Model: Benchmarks and Results from a Wide Range of Numerical Algorithms. *Physical Review X*, **5**, 041041 (2015).

[383] R. Babbush, N. Wiebe, J. McClean, J. McClain, H. Neven & G. K.-L. Chan. Low-Depth Quantum Simulation of Materials. *Physical Review X*, **8**, 011044 (2018).

[384] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li et al. The Variational Quantum Eigensolver: A review of methods and best practices. *Physics Reports*, **986**, 1–128 (2022).

[385] C. Cade, L. Mineh, A. Montanaro & S. Stanisic. Strategies for solving the Fermi-Hubbard model on near-Term quantum computers. *Physical Review B*, **102**, 235122 (2020).

[386] S. Stanisic, J. L. Bosse, F. M. Gambetta, R. A. Santos, W. Mruczkiewicz, T. E. O'Brien, E. Ostby & A. Montanaro. Observing ground-state properties of the Fermi-Hubbard model using a scalable algorithm on a quantum computer. *Nature Communications*, **13**, 5743 (2022).

[387] E. Farhi, J. Goldstone, S. Gutmann & M. Sipser. Quantum Computation by Adiabatic Evolution (2000).

[388] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren & D. Preda. A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem. *Science*, **292**, 472–475 (2001).

[389] E. Farhi, J. Goldstone & S. Gutmann. A Quantum Approximate Optimization Algorithm (2014).

[390] S. Hadfield, Z. Wang, B. O'Gorman, E. G. Rieffel, D. Venturelli & R. Biswas. From the Quantum Approximate Optimization Algorithm to a Quantum Alternating Operator Ansatz. *Algorithms*, **12**, 34 (2019).

[391] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'95, pages 1137–1143. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995). ISBN 978-1-55860-363-9.

[392] D. P. Kingma & J. Ba. Adam: A Method for Stochastic Optimization (2017).

[393] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems (2016).

[394] F. Chollet et al. Keras (2015).

[395] A. B. Owen. A Robust Hybrid of Lasso and Ridge Regression. In J. S. Verducci, X. Shen & J. Lafferty, editors, *Contemporary Mathematics*, volume 443, pages 59–71. American Mathematical Society, Providence, Rhode Island (2007). ISBN 978-0-8218-4195-2 978-0-8218-8122-4.

[396] X. Dang, H. Peng, X. Wang & H. Zhang. Theil-Sen Estimators in a Multiple Linear Regression Model. *Olemiss Edu* (2009).

[397] M. A. Fischler & R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of The Acm*, **24**, 381–395 (1981).

[398] H. Cantzler. Random sample consensus (ransac). *Institute for Perception, Action and Behaviour, Division of Informatics, University of Edinburgh*, **3** (1981).

[399] Y. Freund & R. E. Schapire. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence* (1999).

[400] L. Breiman. Random Forests. *Machine Learning*, **45**, 5–32 (2001).

[401] A. Natekin & A. Knoll. Gradient Boosting Machines, a Tutorial. *Frontiers in Neurorobotics*, **7** (2013).

[402] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola & V. Vapnik. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, pages 155–161. MIT Press, Cambridge, MA, USA (1996).

[403] N. S. Altman. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician*, **46**, 175–185 (1992).

[404] T. Chen & C. Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794 (2016).

[405] S. J. Pan & Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345–1359 (2010).

[406] K. Pernal & K. J. H. Giesbertz. Reduced Density Matrix Functional Theory (RDMFT) and Linear Response Time-Dependent RDMFT (TD-RDMFT). In N. Ferré, M. Filatov & M. Huix-Rotllant, editors, *Density-Functional Methods for Excited States*, pages 125–183. Springer International Publishing, Cham (2016). ISBN 978-3-319-22081-9.

[407] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke & K.-R. Müller. Bypassing the Kohn-Sham equations with machine learning. *Nature Communications*, **8**, 872 (2017).

Summary

The field of quantum chemistry models the behavior of atoms and molecules using quantum mechanical models. At its core is the electronic structure problem, that focuses on determining the energy and properties of electrons in the molecule. This problem is known to be exponentially hard to solve exactly for classical computers. To still simulate the electronic structure, effective models have to be constructed that contain layers of approximations. For a particular class of systems, called *strongly correlated systems*, these approximations can break down completely. This suggests us to look beyond classical computers to solve problems in quantum chemistry. Quantum computers present a solution: they can process quantum information natively, holding the potential to simulate complex systems beyond the reach of classical methods. These quantum simulations of complex molecular systems could significantly impact catalysis, materials design, and photochemistry.

Although quantum hardware has seen impressive advancements recently, building a quantum computer that can perform practical calculations remains a significant challenge. Given the limited quantum resources in near-term devices, optimizing every step of the pipeline in a quantum computational chemistry calculation is essential. This thesis advances quantum computation for chemistry by optimizing quantum resource usage, introducing novel embedding techniques, and exploring applications beyond ground-state energies.

The introductory chapter 1 establishes the basis of this thesis by covering the basics of quantum chemistry methods for the electronic structure problem. It outlines quantum chemistry methods for classical computers like Hartree-Fock and DFT, and goes into depth on active space methods that are used throughout the thesis. Additionally, it gives a concise introduction to quantum computing for chemistry, focused on near-term algorithms. Subsequent chapters introduce and detail novel methods for quantum simulation, all connected by the focus on

reducing quantum resource requirements and exploring targets beyond ground-state energies that are particularly challenging for classical methods.

Chapter 2 prescribes a method to reduce the λ-norm of the electronic structure Hamiltonian. This norm is a crucial measure of quantum resources required for simulations in both Noisy Intermediate-Scale Quantum (NISQ) devices and Fault-Tolerant Quantum Computing (FTQC). By transforming molecular orbitals, the λ-norm can be significantly reduced, optimizing both sampling efficiency in NISQ algorithms and runtime requirements in FTQC. Localized orbitals are demonstrated to outperform canonical molecular orbitals, and an orbital-optimization scheme tailored to minimize the λ-norm is proposed. Benchmark results show its effectiveness for diverse systems, including hydrogen chains, FeMoco, and Ruthenium complexes.

In chapter 3, the FragPT2 framework is developed to address the challenge of simulating large systems with multiple interacting fragments. By embedding each fragment within the mean field of the others and applying multi-reference perturbation theory, inter-fragment correlations such as dispersion and charge transfer are captured efficiently. This one-step embedding approach reduces computational overhead and enables scalable simulations. Applications to molecular systems like a N_2 dimer, aromatic dimers, and butadiene demonstrate its robustness and potential to extend wavefunction-based techniques to larger systems.

In chapter 4, the thesis explores photochemical systems, where multiple electronic states must be treated on equal footing. Building on the state-averaged orbital-optimized variational quantum eigensolver (SAOOVQE), analytical energy gradients and non-adiabatic couplings are derived. These extensions enable geometry optimizations and simulations of non-adiabatic dynamics on a quantum computer, as demonstrated by locating the conical intersection in the formaldimine molecule. These advancements make SAOOVQE a promising tool for studying photoisomerization and related processes in larger biomolecules.

Building on the subject of photochemistry, chapter 5 explores a potential application of near-term quantum computers: the detection of conical intersections in a molecular model. In these regions, where the Born-Oppenheimer approximation breaks down, it is crucial to describe these correctly for understanding photochemical processes such as non-radiative relaxation. Conical intersections are marked by a discrete Berry phase, either 0 or π, which encapsulates their topological nature. This chapter introduces a robust quantum algorithm to identify these intersections by extracting the Berry phase as a single bit of information. The algorithm's discrete outcome ensures resilience to noise, supported by analytical proof and numerical validation using a toy molecular model that mimics behaviors of light-perception biochemistry.

Finally, in chapter 6, the thesis explores machine learning density functionals using noisy quantum data. While Density Functional Theory (DFT) is computationally efficient, its accuracy depends on approximated exchange-correlation functionals. This work proposes a neural network-based approach to learn the exact functional from quantum data. Training on noisy outputs from NISQ devices demonstrates that sampling noise can be mitigated effectively. Applications to the Fermi-Hubbard model and other systems highlight the potential of this method to generalize to new potentials and deliver accurate results under noisy conditions.

Acknowledgments

This small section is devoted to thanking those that have been indispensable, both in completing this thesis and for any future path I may take. Already during my studies, it quickly became clear that working alone wasn't for me. While I've always loved learning, and find great satisfaction in solving puzzles, what really motivates me is working with others – sharing ideas, building something together, and forming connections. These bonds go far beyond the scientific sphere; the fun, interesting discussions, and the mental support from my closest friends have been just as vital to this process (and my well-being).

First of all, I want to thank my supervisors, Luuk and Franco, for their guidance and support throughout my PhD. Luuk, you adopted me in your group as a master student coming from theoretical physics that didn't have any clue about chemistry. I'm grateful that you let me switch topics halfway through my master's project – a turning point that really made things click. You also gave me the opportunity to join the collaboration with the Applied Quantum Algorithms (aQa) group in Leiden, which shaped my future research. I'm thankful that, despite your busy schedule, when you have time, we often have long discussions where we can dive deep into a topic and you offer sharp and insightful perspectives, teaching me a lot about quantum chemistry. Franco, thank you for your kind and thoughtful guidance, and for helping me better understand the application side of my work.

In the same collaboration I met Saad, Bruno and Tom. Saad and Bruno, thank you both for your guidance during my master project and the early stages of my PhD. You taught me a lot about how to write a paper – and your support helped me get off to a strong start. Saad, your mentorship, support and bad jokes made all the difference, and I'm happy to call you a good friend. I'm glad we stayed in touch, that I could visit your new group in Strasbourg, and that we're now collaborating again. Tom, thank you for the opportunities you gave me, and for pushing/teaching us to really finish a project.

Carlo, thank you for the opportunity to join the Lorentz Institute and for handling the details of my contract in Leiden. You made me feel welcome, and I always enjoyed being part of the daily coffee breaks.

Then I want to thank my paranymphs Stefano Polla and Alicja Dutkiewicz. It was in the middle of the COVID pandemic that I started my PhD. Stefano, whenever we were allowed to go to the university during this time, you made me feel welcome immediately. You offered me a spot in your office, where we had discussions about science and more – sometimes over a beer. Your insight and uncanny breadth of knowledge for research topics keeps impressing me, and collaborating with you has been the biggest source of motivation to keep going on my academic path. I'm grateful for the tight friendship that grew alongside it, and I'm sure we'll keep working together for many years – you might just be the "love of my science life".

Alicja, thank you for the endless support and fun times we've shared. From going to QIP together to the many train rides and "car therapy" sessions between Amsterdam and Leiden, your friendship means so much to me.

Thank you to Kshiti and Marshall for sharing an office with me – I hope the loud meetings with Stefano weren't too distracting. I'm happy I can consider you good friends.

I'm grateful to have been a part of the aQa group, of which I feel a full member despite my shared affiliation. Thank you, Jordi and Vedran, for including me and giving me the opportunity to give tutorials and lectures in the aQa course.

Thank you to all my other friends in aQa, without whom it would not have been the same. In this limited space I want to give a special mention to Xavi, Patrick, David, Nastya, Mahtab, Ilse, Nina, and Felix for their support both in and outside science.

I want to thank Sharath, Souloke, Arno, Ansgar and Masha, along with my other colleagues and friends in the theoretical chemistry group in Amsterdam. I'm grateful to have learned a lot about chemistry from you during all the discussions and group meetings.

Thank you, Kareljan, Seenivasan, Edison and Dyon, as well as others associated to the QC^2 collaboration. These meetings have proved fruitful for many new ideas, meanwhile learning from each other coming from different fields.

And finally, it would all not have been possible without the loving support of my friends and family. Unfortunately, it would take too much space to mention all of you by name, but I think you know who you are.

Thank you, Maryam, for supporting me and helping to put the final touch on the thesis cover.

I want to give a special thanks to Loïs, who supported me through much of the PhD.

Finally, I want to thank my father Erik and my sister Eva for always having my back. A special thanks to Eva for painting the cover of this thesis. I hope you're proud of your first printed art, I sure am. And to my mom, Ella, who lives in our fond memories. I think she would have been proud of us all.

List of publications

[1] E. Koridon, S. Yalouz, B. Senjean, F. Buda, T. E. O'Brien & L. Visscher. Orbital transformations to reduce the 1-norm of the electronic structure Hamiltonian for quantum computing applications. *Physical Review Research*, **3**, 033127 (2021).
Author contributions: All authors contributed to conceiving the project. E.K. developed the code. E.K., S.Y., B.S., T.O. and L.V. wrote the manuscript, which all authors contributed to reviewing. S.Y., B.S., T.O., F.B. and L.V. supervised the work.
[Chapter 2 is based on this publication.]

[2] S. Yalouz, E. Koridon, B. Senjean, B. Lasorne, F. Buda, & L. Visscher. Analytical Nonadiabatic Couplings and Gradients within the State-Averaged Orbital-Optimized Variational Quantum Eigensolver. *Journal of Chemical Theory and Computation*, **18**, 776-794 (2022).
Author contributions: S.Y, E.K., B.S., F.B. and L.V contributed to conceiving the project. E.K., S.Y., B.S. and L.V. developed the theory. E.K., S.Y. and B.S. developed the code. S.Y., E.K., B.S., B.L. and L.V. wrote the manuscript, which all authors contributed to reviewing. L.V. and F.B. supervised the work.
[Chapter 4 is based on this publication.]

[3] E. Koridon, J. Fraxanet, A. Dauphin, L. Visscher, T. E. O'Brien & S. Polla. A hybrid quantum algorithm to detect conical intersections. *Quantum*, **8**, 1259 (2024).
Author contributions: E.K. and S.P. conceived the project and developed the codebase. E.K., J.F., A.D. and S.P. worked on the theoretical framework. E.K., S.P., T.O. and L.V. wrote the manuscript, which all authors contributed to reviewing. S.P., L.V. and T.O. supervised the work.
[Chapter 5 is based on this publication.]

[4] E. Koridon, S. Sen, L. Visscher & S. Polla. FragPT2: Multi-fragment Wave Function Embedding with Perturbative Interactions. Journal of Chemical Theory and Computation (2025).
Author contributions: All authors contributed to conceiving the project and developing the theory. E.K. developed the method and the codebase. All authors contributed in writing the manuscript. L.V. and S.P. supervised the work.
[Chapter 3 is based on this publication.]

[5] E. Koridon, F. Frohnert, E. Prehn, E. van Nieuwenburg, J. Tura & S. Polla.. Learning Density Functionals from Noisy Quantum Data, arXiv:2409.02921 (2024).
Author contributions: E.P., E.K. and S.P. conceived the project and developed a first prototype of the code. S.P. developed the codebase for data generation and learning. E.K. developed and optimized VQE. F.F. performed the model selection studies. E.K., F.F. and S.P. wrote the manuscript, which all authors contributed to reviewing. J.T., E.vN. and S.P. supervised the work and provided funding.
[Chapter 6 is based on this preprint.]

List of Abbreviations

AO Atomic Orbital

BCH Baker-Campbell-Hausdorff

CAS Complete Active Space

CASCI Complete Active Space Configuration Interaction

CASSCF Complete Active Space Self-Consistent Field

CC Coupled Cluster

CI Configuration Interaction OR Conical Intersection

CISD Configuration Interaction Singles and Doubles

CMO Canonical Molecular Orbital (equivalent to Hartree-Fock orbitals)

CNN Convolutional Neural Network

CNOT Controlled-NOT

COBYLA Constrained Optimization By Linear Approximation

CP-MCSCF Coupled Perturbed Multi-Configurational Self-Consistent Field

CPU Central Processing Unit

CSF Configuration State Function

DFT Density Functional Theory

DMET Density Matrix Embedding Theory

DMRG Density Matrix Renormalization Group

ED Exact Diagonalization

ER Edmiston-Ruedenberg

EVE Expectation Value Estimation

FB Foster-Boys

FCI Full Configuration Interaction

FTQC Fault-Tolerant Quantum Computing

GTO Gaussian-Type Orbital

HF Hartree-Fock

HOMO Highest Occupied Molecular Orbital

KS Kohn-Sham

LASSCF Localized Active Space Self-Consistent Field

LCAO Linear Combination of Atomic Orbitals

LCU Linear Combination of Unitaries

LMO Localized Molecular Orbital

LUMO Lowest Unoccupied Molecular Orbital

MC Multi-Configurational

MCSCF Multi-Configurational Self-Consistent Field

MECI Minimal Energy Conical Intersection

MF Mean-Field

ML Machine Learning

MO Molecular Orbital

MRPT Multi-Reference Perturbation Theory

MSE Mean Squared Error

NAC Non-Adiabatic Coupling

NISQ Noisy Intermediate-Scale Quantum

NN Neural Network

NP Non-deterministic Polynomial time

NPF Number-Preserving Fabric

NR Newton-Raphson

OAO Orthogonal Atomic Orbitals

ONV Occupation Number Vector

OO Orbital-Optimized

PES Potential Energy Surface

PM Pipek-Mezey

PQC Parameterized Quantum Circuit

PT Perturbation Theory

QEC Quantum Error Correction

QMA Quantum Merlin-Arthur

QMC Quantum Monte-Carlo

QML Quantum Machine Learning

QNG Quantum Natural Gradient

QPE Quantum Phase Estimation

QPU Quantum Processing Unit

RDM Reduced Density Matrix

RHF Restricted Hartree-Fock

RI Resolution of Identity

RILMO Recanonicalized Intrinsic Localized Molecular Orbital

ROSE Reduction Of Spatial Extent

SA State-Averaged

SA-CASSCF State-Averaged Complete Active Space Self-Consistent Field

SA-OO-VQE State-Averaged Orbital-Optimized Variational Quantum Eigensolver

SAPT Symmetry Adapted Perturbation Theory

SCF Self-Consistent Field

SLSQP Sequential Least SQuares Programming algorithm

STO Slater-Type Orbital

TD-DFT Time-Dependent Density Functional Theory

THC Tensor Hyper-Contraction

UCC Unitary Coupled Cluster

UCCSD Unitary Coupled Cluster Singles and Doubles

VHA Variational Hamiltonian Ansatz

VQA Variational Quantum Algorithm

VQD Variational Quantum Deflation

VQE Variational Quantum Eigensolver