Using the Variational Quantum Eigensolver to go from Hartree-Fock to Full Configuration Interaction using a Hartree-Fock Superposition method on the H2O molecule

Persoons, Robin D.L.

Award date:
2021

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Master Thesis

CQT-2021-12

Using the Variational Quantum Eigensolver to go from Hartree-Fock to Full Configuration Interaction using a Hartree-Fock Superposition method on the H_2O molecule

---

Robin Persoons
0903095

Supervisors
dr. ir. S.J.J.M.F. Kokkelmans
Department of Applied Physics
Eindhoven University of Technology

dr. B. Baumeier
Department of Mathematics and Computer Science
Eindhoven University of Technology

24th August 2021

Abstract

When quantum computers enter the Noisy Intermediate-Scale Quantum (NISQ) era, the Variational Quantum Eigensolver (VQE) algorithm will be able to tackle problems in computational chemistry that are unreachable by classical computers. The approximate and fast Hartree-Fock (HF) method, is not accurate enough and the exact Full Configuration Interaction (FCI) method is too slow for these system sizes.

In current research VQE is often applied to simplified or small systems, using algorithms that are not guaranteed to work on bigger systems, some even known not to. These limitations are worked around by precalculating good estimates and improving those instead of starting from scratch. There is no guarantee that these estimations are good enough (or indeed fast enough) to be a valid solution for large systems. In this report the goal is to go directly from HF to FCI using VQE.

The usual optimization algorithms failed in these conditions so other alternatives were explored. Orbital selection was looked into to find a balance between approximation on one side and system size on the other. An exhaustive set of exact results were gotten to help determine which orbital sets should be tested on VQE. Results from the literature were successfully reproduced in this area.

A Hartree-Fock Superposition (HFS) method was thought off and tested thoroughly. This method looks at superpositions of the VQE trial state with the HF solution, since the latter was a big component in the exact solutions of the geometries of interest. It did not only improve convergence speed, it also improved the solution itself, most likely by making the optimization easier due to a reduction in search space. Additionally the scheme can be implemented on a quantum circuit that is only slightly different than the normal VQE quantum circuit. This circuit has a running time increase of only a constant factor of approximately 3. While chemical accuracy of the FCI solution was not reached, some results were very close to that threshold. Using the best performing HFS parameter to analyse the asymptotic behavior of H_2O resulted in incorrect asymptotic behavior however, as the HF component loses value and the method requires different parameters.

Contents

Chapter 1

Introduction

Quantum computers speak to the imagination and their potential is often discussed in the media, but in reality they can not yet deliver to this ideal. It is known that they would be able to jeopardize current cryptologic protocols with Shors algorithm [1] or find specific entries in databases incredibly fast with Grovers search algorithm [2, 3]. However the current state of quantum computers is quite far from being able to use all the theorized superior algorithms. At the moment of writing quantum processors of ~ 30 qubits are cutting edge [4]. The next step in quantum computing is not an universal quantum computer with thousands of qubits but rather $50 - 1000$ qubit processors limited by their noise, gate errors and measurement errors. This era is called the Noisy Intermediate-Scale Quantum (NISQ) era, coined by John Preskill [5]. Multiple big companies (including Google and IBM) have announced the development of quantum processors that would mark the start of the NISQ era [4]. During this era, before big fault tolerant quantum computers, quantum computers can already start to outperform classical ones. Much work is being done to get these systems as faultless as possible and fix other hardware/implementation issues. In this thesis however, a perfect quantum system is considered and these issues are completely disregarded.

When quantum computers were first suggested by Richard Feynman in the 1980s, the main idea was to simulate quantum systems with a quantum system [6]. It turns out that this problem is one that can already be tackled handily with NISQ era quantum processors [5]. In this project the main interest is the calculation of the electronic ground state of a molecule. Classically this problem does not scale well. Oversimplifying, one has to limit themselves to a small function space and then either use the approximating Hartree-Fock (HF) method, which has a reasonable running time, but weak results, the Full Configuration Interaction (FCI) method, which gives exact results within the limited space, but has an exponential running time or something in between. For this problem the Variational Quantum Eigensolver (VQE) was proposed by Peruzzo et al. in 2014 [7]. VQE is an algorithm designed for NISQ systems to solve the ground state problem. The algorithm uses the variational principle and is a so called hybrid algorithm, where a specialized quantum processor does part of the work while a classical computer does the rest. The quantum processor determines the energies of states queried by the classical computer, while the classical computer uses standard optimization techniques to determine what to query next.

Many current quantum ground state energy experiments are being done on very small or very simple molecules such as H2, LiH and H-chains [8, 9], or simplified systems [10], which can be calculated to very high accuracy with classical methods due to their small size. NISQ era processors will be able to run VQE to analyse systems currently out of reach for classical methods. One of the challenges in scaling up the problem (next to creating quantum processors of the required size), is improving the methods as naive search algorithms that work very well for low dimensional problems start to fail when the dimension is increased. Most literature used local optimizers and either small systems or precalculation to start already close to the solution in order for them to work. It is not clear that this will work for larger systems. The simulations in this report are different in the sense that most of the time no information that was not from HF was used to help the VQE algorithm. The idea being to go from HF directly to FCI using VQE. Additionally a step further than the very small molecules was taken and the focus was placed on the water molecule, which in terms of VQE simulations, can be considered on the large side at the time of writing. In fact just after the start of this project a paper on the VQE simulation of H_2O was published in Nature [11].

This project is part of the exploratory look into hybrid algorithms that could run on the quantum platform currently in development by the center for quantum materials and technology (QT/e) at the TU/e. The QT/e

group is developing a scalable quantum platform based on ultracold Rydberg atoms in optical tweezers. This platform will be accessible 24/7 via the Quantum Inspire facility [1]. QT/e has strong ties with Quantum Delta NL [2], who have recently received €615 million funding from the Netherlands' National Growth Fund.

Because the background of the reader could be in Quantum Physics, Computational Chemistry or neither, all required theory is explained from a basis that should be appropriate in the first few sections. It is indicated as much as possible which sections relate to which subject such that the reader can skip the subjects they do not need explained. First some basic concepts and prerequisites are treated from both Quantum Physics (Chapter 2) and Computational Chemistry (Chapter 3). Then more classical Computational Chemistry and it's methods are discussed (Chapter 4). The theory behind the Variational Quantum Eigensolver itself is discussed afterwards (Chapter 5) and the last chapter before the results (Chapter 6) explains the VQE algorithm and simulations used in the project. The results are discussed in the three chapters afterwards (Chapters 7,8 and 9). Finally the conclusions and future work are discussed (Chapter 10).

If the reader has a background in quantum physics and quantum computation Chapter 2 can be skipped. Derivations and proofs shown in Chapters 3 and 4 can also be skipped. Specifically this means that sections 3.3, 4.1 and 4.2 are mostly optional. Sections 3.1, 3.2, 4.2.6, 4.3 and 4.4 are recommended however. Sections 5.1 and 5.2 can be skipped by those familiar with the second quantization and Jordan-Wigner transform. The remaining Chapters are essential and should not be skipped with the possible exception of section 6.4, which can be skipped if the optimizers are taken as a black box by the reader.

[1] `https://www.quantum-inspire.com/`
[2] `https://quantumdelta.nl/`

Chapter 2

Introductory Physics Concepts

In this section some basic quantum physics concepts are explained. The subjects that will be discussed are in order: The Schrödinger equation, Bra-Ket Notation, Wavefunctions, The Antisymmetry principle, The Variational Principle and Qubits. The reader is welcomed to skip the sections that they do not require.

2.1 The Schrödinger equation

In quantum physics, a system can largely be understood by knowing its wavefunction. It is no surprise that the wavefunction is the variable in the most central equation of quantum physics: The Schrödinger equation. The Schrödinger equation has two general forms: the time dependent and time independent Schrödinger equations. If the Hamiltonian, the operator that descrbies the total energy of a system, is time independent, it is fairly simple to create a solution to the time dependent equation using a solution from the time independent one (see for example [12]). This is the case in this report and thus the relevant form is the time independent Schrödinger equation:

$$\hat{H}|\Phi\rangle = E|\Phi\rangle. \tag{2.1}$$

Where $\hat{H}$ is the (time independent) Hamiltonian. It works on the wavefunction describing the system $|\Phi\rangle$. Wavefunctions and the used notation will be discussed in more depth shortly. Lastly E is a real scalar: the energy of the system (Note that E being real means that H is Hermitian). It is most likely clear that this equation is in fact an eigenvalue problem and indeed it will be explained later that the main interest of this report is the lowest eigenvalue of the Hamiltonian. The wavefunction that is the eigenvector/eigenfunction of that lowest eigenvalue is in fact the one describing the ground state of the system.

2.2 Bra-Ket Notation

The Bra-Ket notation is a notation very commonly used in quantum physics. The basis of the notation is representing an inner product, anti-linear in the first argument (not the second as is more standard) $(\cdot,\cdot)$ on some vector space, as $\langle\cdot|\cdot\rangle$. Note that the requirements on the inner product already make the vector space an Hilbert space. The right part, for example $|\phi\rangle$, corresponds to a vector and is called a **ket**. The left part, for example $\langle\psi|$, corresponds to the functional $(\psi,\cdot)$ and is called a **bra**. The inner product (ψ,ϕ) is noted as $\langle\psi|\phi\rangle$. Very often, and almost always in this report, the Hilbert space is complex and the scalar product is the Hermitian inner product $(\psi,\phi) = \psi^\dagger\phi$, where $\psi^\dagger$ is the Hermitian adjoint of ψ. Importantly using the operator O of an observable (such as the Hamiltonian for energy), the expected value of that observable is the same as $\langle O\rangle = \langle\phi|O|\phi\rangle$. It is important to note that the bra and ket have to be the same when calculating an expected value. Understanding the notation in more detail is not a requirement for understanding the report, with possible exceptions in some of the derivations.

2.3 Wavefunctions

A quantum state of a particle can be described with a set of basis states. If this set is countable the state can be written as:

$$|\phi\rangle = \sum_i c_i |\psi_i\rangle. \tag{2.2}$$

A physical state is normalized. This means that the basis states are normalized ($\langle\psi_i|\psi_i\rangle = 1$) and the absolute squares of the coefficients sum to 1 ($\sum_i |c_i|^2 = 1$). When the basis set represents an observable quantity, the probability of measuring the value corresponding to basis state $|\psi_i\rangle$ is $|c_i|^2$. This works out because of the normalization mentioned before. Choosing the basis states to be the position states (the infinitely many states corresponding to the particle being at some specific point in space), the probability distribution is no longer discrete but continuous. The c_i's from before become a probability density function and this function is the absolute square of the wavefunction. The state can now be described with the wavefunction:

$$|\phi\rangle = \int d^3x \Phi(x)|x\rangle. \tag{2.3}$$

For multiple particles the wavefunction will take multiple arguments ($\Phi(x_1, \ldots, x_n)$). Now $\int d^3x_i |\Phi|^2$ gives the probability that particle i is at x_i. Often the notation $|\Phi(x_1, \ldots, x_n)\rangle$ or similar is used to denote the state described by the wavefunction $\Phi(x_1, \ldots, x_n)$. While the position/momentum wavefunction and qubit wavefunctions are the only wavefunctions of interest in this report, there are also wavefunctions that describe spin or other observables of a system. Depending on this the wavefunctions are vectors in different Hilbert spaces. In this report non-qubit wavefunctions will always be position/momentum wavefunctions and therefore will always be vectors from the space of complex square-integrable functions $\mathcal{L}^2(\mathbb{C})$.

2.4 Antisymmetry principle

The Antisymmetry principle is one of the postulates (axioms) of Quantum Physics. It is a stronger requirement that includes Pauli's exclusion principle. Pauli's principle says that any two fermions (electrons) can not be in the same quantum state in the same system at the same time. The Antisymmetry principle says that any multiple fermion (electron) wavefunction has to be antisymmetric with respect to a pair exchange of particles. This means that for any i, j it holds that:

$$\Phi(\ldots, x_i, \ldots, x_j, \ldots) = -\Phi(\ldots, x_j, \ldots, x_i, \ldots). \tag{2.4}$$

Pauli's exclusion principle follows directly from looking at the case where $i = j$.

$$\Phi(\ldots, x_i, \ldots, x_i, \ldots) = -\Phi(\ldots, x_i, \ldots, x_i, \ldots) = 0 \tag{2.5}$$

Since the absolute value of the wavefunction squared is a probability density function it being zero means this state does not occur thereby enforcing Pauli's exclusion principle.

2.5 Variational Principle

The Variational Principle states that for any normalized wave-function ψ restricted to some (restricted) Hilbert space for which the Hamiltonian H is defined, the energy $E_\psi = \langle\psi|H|\psi\rangle$ is bound from below by the unique ground state energy E_0, which is the lowest eigenvalue of H in this space. Now follows a quick proof from Griffiths [12]: Since the Hamiltonian H has a complete set of eigenfunctions it holds for all normalized wave-functions Ψ that:

$$\Psi = \sum_n c_n \psi_n \text{ with } H\psi_n = E_n\psi_n \text{ and } \sum_n |c_n|^2 = 1. \tag{2.6}$$

It now follows quite simply for $\langle\Psi|H|\Psi\rangle$:

$$\langle\Psi|H|\Psi\rangle = \langle\sum_m c_m\psi_m|H|\sum_n c_n\psi_n\rangle = \sum_m\sum_n c_m^* E_n c_n \langle\psi_m|\psi_n\rangle = \sum_n E_n|c_n|^2 \tag{2.7}$$

and since the ground state energy is the smallest of the eigenvalues of H we have $E_{gs} \leq E_n$ for all n:

$$\langle H \rangle = \sum_n E_n |c_n|^2 \geq E_{gs} \sum_n |c_n|^2 = E_{gs} \tag{2.8}$$

The Variational Principle allows one to sample wavefunctions by evaluating its energy knowing that they are looking for the lowest value. Most methods discussed in this report use iterative sampling to converge to a minimum value, which is the (approximate) ground state energy or solve for the lowest eigenvalue of the Hamtilonian directly.

2.6 Qubits

Qubits are the quantum physical extension of regular bits. Instead of being binary valued, they always in a quantum superposition of the states $|0\rangle$ and $|1\rangle$, corresponding to the classical 0 and 1. A qubit $|\psi\rangle$ can be fully described by:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \tag{2.9}$$

where $|\alpha|^2 + |\beta|^2 = 1$, since $|\alpha|^2$ and $|\beta|^2$ are the probabilities of measuring a 0 or 1 respectively. Additionally since α and β are complex numbers different numbers can be taken with the same absolute square. Indeed multiplication of α or β with e^{ia} for real a will result in the same absolute squares. This term is called the phase of the state. Phases will mostly be ignored in this report, which leads to no problems in general, but where important this will be mentioned. It will be useful later to be familiar with the Bloch-sphere representation of the qubit, which is shown in Figure 2.1. This representation is useful because it is a nice visualization when considering single qubit manipulations, which can be interpreted as rotations along the Bloch sphere.

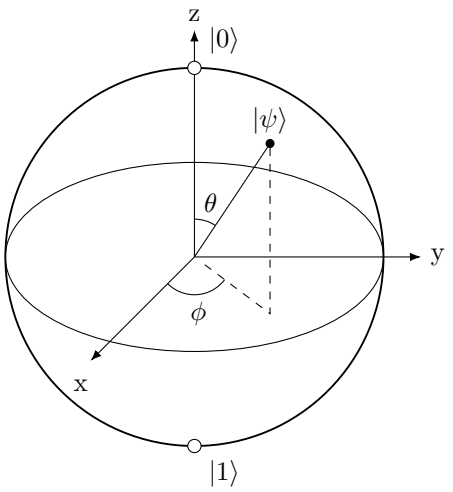


Figure 2.1: The Bloch-sphere representation of a qubit in state $|\psi\rangle$. ϕ is the phase angle determining the phase and θ is the azimuthal angle determining α and β. This image was made based on an illustration in [13].

This can naturally be extended to systems of multiple qubits. The two qubit state $|\chi\rangle$ can be fully described by:

$$|\chi\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle, \tag{2.10}$$

where $|\alpha|^2 + |\beta|^2 + |\gamma|^2 + |\delta|^2 = 1$, since $|\alpha|^2$, $|\beta|^2$, $|\gamma|^2$ and $|\delta|^2$ are the probabilities of measuring qubit one and two 0 , qubit one 0 and qubit two 1, qubit one 1 and qubit two 0 and both qubits 1 respectively. In general a N qubit system can be described with a vector of 2^N elements representing the probabilities for all measurable outcomes. Of course the normalization of the absolute squares of the coefficients is required in general as well. This vector representation is what is used in many of the simulations in this report to represent qubit states. An observant reader may have noticed that the normalization requirement fixes the value of the last vector element, thus reducing the amount of required elements to $2^N - 1$. In the simulations this is not used because it is preferable to work with powers of 2. The Bloch-sphere representation of a N qubit system is simply a collection of N (possibly coupled) Bloch-spheres. Multiple qubit operations can be seen as manipulating multiple spheres at the same time. They have to be unitary (norm-preserving) to keep the points on the sphere.

Chapter 3

Computational Chemistry

This chapter and the following one discuss the background of computational chemistry that has been utilized. Both chapters are strongly based on a text book on the topic by Szabo and Ostlund [14] and a video series most likely based on the same book [15]. The derivations in sections 3.2.2 and 3.3 are not based on any source as the aforementioned sources did not include them and no other derivations in the same notation were found.

3.1 Electronic Hamiltonian

The main interest in this report is calculating the ground state energy of molecules, ignoring relativistic effects. This is the energy described by the non relativistic time-independent N-electron, M-nuclei Hamiltonian. However, as it turns out, finding the ground state of the electronic Hamiltonian, which we derive in this section, is the main challenge. The start, of course, is the mentioned N-electron, M-nuclei Hamiltonian:

$$\mathcal{H} = -\sum_{I=1}^{M} \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_{i=1}^{N} \frac{\hbar^2}{2m}\nabla_i^2 + \frac{e^2}{2}\sum_{I=1}^{M}\sum_{J\neq 1}^{M} \frac{Z_I Z_J}{R_{IJ}} + \frac{e^2}{2}\sum_{i=1}^{N}\sum_{j\neq 1}^{N} \frac{1}{R_{ij}} - e^2 \sum_{i=1}^{N}\sum_{I=1}^{M} \frac{Z_I}{R_{iI}}. \tag{3.1}$$

Here $i = 1, ..., N$ are indices for the electrons, $I = 1, ..., M$ are indices for the nuclei, $\hbar$ is the reduced Planck constant, M_I is the mass of nucleus I and Z_I is its charge, m is the electron mass, e is the elementary charge and R_{ab} is the distance between a and b. The Laplacians ∇_I^2 and ∇_i^2 act on individual nuclei and electrons respectively. The terms are can be identified: the first two are sums over the nucleus and electron kinetic energies respectively, the other three are potential energies due to nuclear repulsion, electron repulsion and electron-nuclear attraction respectively. The first step is to get rid of the constants in this expression. This is done by a change of units, as the constants have units themselves. The units are transformed to atomic units which are defined by the scalings below.

1. Mass: $m_e = 1$
2. Charge: $e = 1$
3. Angular Momentum: $\hbar = 1$
4. Permittivity: $4\pi\epsilon_0 = 1$

Where the last one is (in this case) equivalent to changing the unit of length: $a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2} = 1$. This means mass will be in units of electron mass, charge will be in units of elementary charge and length will be in units of Bohr radii. Additionally the double counting for the nuclear repulsion and electron repulsion terms is avoided to get rid of the factors $\frac{1}{2}$ by changing the summation and the result is:

$$\mathcal{H} = -\sum_{I=1}^{M} \frac{1}{2M_I}\nabla_I^2 - \frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 + \sum_{I<J}^{M} \frac{Z_I Z_J}{R_{IJ}} + \sum_{i<j}^{N} \frac{1}{R_{ij}} - \sum_{i=1}^{N}\sum_{I=1}^{M} \frac{Z_I}{R_{iI}} \tag{3.2}$$

Where $\sum_{i<j}^{N}$ is shorthand for $\sum_{i=1}^{N}\sum_{j=i+1}^{N}$. The next step is applying the Born-Oppenheimer Approximation [16]. Nuclei are considered as motionless in comparison to the electrons such that the kinetic energy term of the

nuclei will vanish. Additionally the nuclear repulsion is reduced to a constant in the Hamiltonian, which has no influence on the eigenfunctions and only adds a constant to the eigenvalue. This can be shown easily using some operator Q with eigenfunctions f_n, eigenvalues q_n and another operator $\bar{Q} = Q + c$:

$$\bar{Q}f_n = (Q + c)f_n = Qf_n + cf_n = (q + c)f_n. \tag{3.3}$$

In this report the nuclear repulsion term is calculated separately once and added at the end to find the total energy. Removing it together with the first term results in the electronic Hamiltonian:

$$H_e = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i<j}^{N}\frac{1}{R_{ij}} - \sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_I}{R_{iI}} \tag{3.4}$$

3.2 Spaces, (Basis Sets) and Orbitals

The electronic Hamiltonian is an operator in some Hilbert space of N-electron wavefunctions. Since this space is infinite dimensional it can not be searched completely. Because none of the systems of interest can be solved exactly, sub-spaces of this Hilbert space are used.

Firstly, because symmetric wavefunctions are non-physical for fermions, the subspace of antisymmetric N-electron wavefunctions is chosen as starting point. This subspace is also infinite dimensional and therefore needs to be reduced. This is done by looking at antisymmetric wavefunctions that are built using orbitals. Orbitals are single particle wave-functions, describing electrons. They have two parts. The first part is called a spatial orbital, which describes the probability of the position of the electron. Sets of spatial orbitals are often assumed/chosen to be orthonormal. The second part is a spin-function which is either $\alpha(\omega)$ or $\beta(\omega)$ (spin up $\uparrow$ or spin down $\downarrow$). The spin-functions are defined to be orthonormal to enforce Pauli's exclusion principle. The combination of a spatial orbital with a spin-function creates a spin orbital. Note that a set of K spatial orbitals creates $2K$ unique spin orbitals and additionally that if the set of spatial orbitals is orthonormal then so is the set of spin orbitals.

These orbitals are often written as $\chi_1, \dots \chi_{2K}$, where the orbitals are sorted in pairs of identical spatial orbitals and opposite spin. This means χ_1 and χ_2 have the same spatial orbital but opposite spin functions, just like χ_{2i+1} and χ_{2i+2}. All odd numbered orbitals have opposite spin to all even orbitals. In all methods discussed in this report a finite set of orbitals is chosen as basis set to span a search space in the antisymmetric subspace. The next section will discuss how N-electron wavefunctions are built from the orbitals. Only finitely many wavefunctions from the antisymmetric N-electron wavefunction subspace are created. This subspace spanned by the orbitals is not complete in that subspace and therefore not every valid wave-function can be analysed. This means that in general the exact solution can not be found and that results are strongly dependent on the quality of basis choice. The exact solution within the space spanned by the chosen orbital basis set however, can be found by some of the methods discussed later.

3.2.1 Hartree-Products and Slater Determinants

To make N-electron wavefunctions from spin orbitals we can not use linear combinations. A multiple particle state is described by a product of single particle states, not a sum. The basic building blocks for these structures are Hartree-products. Hartree-products are N-electron wave-functions that are the product of N single electron wave-functions (spin orbitals):

$$\Phi(x_1, x_2, \dots x_n) = \chi_1(x_1)\chi_2(x_2)\dots\chi_n(x_n) = \prod_{i=1}^{n}\chi_i(x_i). \tag{3.5}$$

There are some problems with assuming that your solution wavefunction is a Hartree-product. Firstly, per definition, it assumes electrons to be independent, which they are not. Secondly, these wavefunctions are not antisymmetric. This is a problem because the antisymmetry principle needs to be obeyed, since it is one of the axioms of quantum physics. Instead antisymmetric combinations of Hartree-products are used. The Slater determinant is one such a combination. The name comes from the fact that these combinations can be rewritten as the determinant of a N by N matrix of orbitals. The notation $|\chi_i\chi_j\cdots\chi_k\rangle$ is used for the normalized Slater

determinant:

$$|\chi_i\chi_j\cdots\chi_k\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix}\chi_i(x_1) & \chi_j(x_1) & \cdots & \chi_k(x_1)\\ \chi_i(x_2) & \chi_j(x_2) & \cdots & \chi_k(x_2)\\ \vdots & \vdots & & \cdots\\ \chi_i(x_N) & \chi_j(x_N) & \cdots & \chi_k(x_N)\end{vmatrix} \tag{3.6}$$

Here the columns represent the orbitals and the rows the electrons. The determinant fixes the problem of independent electrons: it describes N electrons in N spin orbitals but it is not clear which one is where. Additionally having two electrons in the same orbital (equal rows) or two identical orbitals (equal columns) results in the determinant being zero (Pauli exclusion principle). Finally the anti-symmetry property follows immediately as well, because the determinant switches sign when rows are switched. Note that there is still an issue with the Slater determinant: there is no correlation between electrons of different spin. The single Slater determinant wavefunctions created by the basis set of orbitals $\chi_1,\ldots\chi_{2K}$ is the first concrete subspace that will be searched for a solution. This is done in the Hartree-Fock method that is discussed later. It finds the best single determinant approximation of the ground state. This approximate ground state is called the Hartree-Fock ground state [14]

$$|\Phi_0\rangle = |\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\rangle \tag{3.7}$$

and is used as the starting point of defining excited determinants, both conceptually and in a notational sense. Exchanging spin orbitals from the Hartree-Fock ground state with others from the $2K-N$ unused spin orbitals creates excited Slater determinants. In this way it is analogue to actual ground states and excited states. A singly excited determinant has a single electron that was in χ_a in the Hartree-Fock groundstate put into χ_r, one of the virtual orbitals

$$|\Phi_a^r\rangle = |\chi_1\chi_2\cdots\chi_r\chi_b\cdots\chi_N\rangle. \tag{3.8}$$

Similarly a doubly excited determinant where χ_a and χ_b were moved to χ_r and χ_s respectively is written as

$$|\Phi_{ab}^{rs}\rangle = |\chi_1\chi_2\cdots\chi_r\chi_s\cdots\chi_N\rangle. \tag{3.9}$$

This extends naturally up to N-tuply excited determinants making it such that all $\binom{2K}{N}$ determinants are either the Hartree-Fock ground state or singly, doubly, ..., N-tuply excited determinants. These excited determinants can be used to create more intricate wavefunctions. Most notably the following one:

$$|\Phi\rangle = c_0|\Phi_0\rangle + \sum_{ar} c_a^r|\Phi_a^r\rangle + \sum_{\substack{a<b\\r<s}} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c\\r<s<t}} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d\\r<s<t<u}} c_{abcd}^{rstu}|\Phi_{abcd}^{rstu}\rangle + \ldots, \tag{3.10}$$

which includes all $\binom{2K}{N}$ Slater determinants exactly once, with different coefficients.

3.2.2 Relevant spaces

The linear combinations of all $\binom{2K}{N}$ determinants in fact span the entire space of antisymmetric N electron wavefunctions created from the chosen orbitals. To show this, first the space has to be defined. Since products are the only way to create N electron wave functions out of N one electron wavefunctions, it is easiest to build the space from Hartree-products. Define I as a subset of exactly N different spin orbitals out of the total $2K$ spin orbitals. Taking the same orbital twice gives an impossible state due to Pauli's exclusion principle and thus there are exactly $\binom{2K}{N}$ subsets I with non-zero Hartree-products. Putting a constant in front of every Hartree-product spans the biggest space. This is O_{full}, the entire space of valid N-electron wavefunctions created from the orbitals.

$$O_{\text{full}} = \left\{ \Phi(x_1,\ldots x_n) = \sum_{I}^{\binom{2K}{N}} \sum_{i\neq j\neq\ldots\neq k\in I} c_{ij\ldots k}\ \chi_i(x_1)\chi_j(x_2)\cdots\chi_k(x_N) \,\middle|\, c_{ij\ldots k}\in\mathbb{C} \text{ s.t. } \Phi \text{ antisymmetric} \right\} \tag{3.11}$$

Where for each permutation of the orbital indices in I there is a Hartree-product. The indices indicate the order of the orbitals in the product, while the coordinates stay in the same order. The second sum has inequality signs to prevent the same orbital getting into the Hartree-product twice (which would make it zero). The space spanned by the linear combination of all Slater determinants will be suggestively called O_{FCI} and is given by

$$O_{\text{FCI}} = \left\{ \Phi(x_1,\ldots x_n) = \sum_{I}^{\binom{2K}{N}} c_I|\chi_i(x_1)\chi_j(x_2)\cdots\chi_k(x_N)\rangle_{i\neq j\neq\ldots\neq k\in I} \,\middle|\, c_I\in\mathbb{C} \right\}. \tag{3.12}$$

Since all orbitals are orthogonal and different no constructions can be made where the outer sum in Equation 3.11 sums over non-antisymmetric functions, picking the coefficients in such a way that the total is antisymmetric. Therefore it is sufficient to show that the elements within the sums over all I are the same. It has to be shown that for $I = i, j, \dots k$ the set $\{c|\chi_i\chi_j\cdots\chi_k\rangle\}_{c\in\mathbb{C}}$ is the same as the antisymmetric part of the set $\left\{\sum_{i\neq j\neq\ldots\neq k\in I} c_{ij\ldots k}\ \chi_i(x_1)\chi_j(x_2)\cdots\chi_k(x_N)\right\}_{c_{ij\ldots k}\in\mathbb{C}}$.

From here some arbitrary I is fixed. The first step is to show that the coefficients of the separated Hartree-products have the same absolute value. This will follow quite fast from the antisymmetry requirement. Given some pair of coordinates i, j, antisymmetry requires that a Hartree-product $\chi_a(x_1)\dots\chi_i(x_i)\dots\chi_j(x_j)\dots\chi_b(x_N)$ has the same coefficient and opposite sign to $\chi_a(x_1)\dots\chi_j(x_i)\dots\chi_i(x_j)\dots\chi_b(x_N)$. Now take any two coefficients $c_0 = c_{ij\ldots k}$ and $c_N = c_{ab\ldots c}$, corresponding to Hartree-products p_0 and p_N built with orbitals from I. Since their products are permutations of each other, it is clear that by switching coordinates at most $N-1$ times we can go from p_0 to p_N, by switching one orbital to its new place at a time. This means there is some path $c_0, c_1 \dots c_k, c_N$ of coefficients such the that the function p_i of coefficient c_i is one switch from function p_{i+1} of coefficient c_{i+1} for all $0 \leq i < k$ and such that p_k of coefficient c_k is one switch from p_N. This means $c_0 = -c_1 = c_2 = -c_3 = \cdots = \pm c_N$ due to antisymmetry and therefore $|c_0| = |c_N|$. Note that this implies that picking one coefficient fixes all others and thus fully describes the function. Since each Slater determinant has all permutations of its N orbital Hartree-Products we have that for a given c and $c_{ij\ldots k}$ the functions $c|\chi_i\chi_j\cdots\chi_k\rangle$ and $\sum_{i\neq j\neq\ldots\neq k}^{N} c_{ij\ldots k}\chi_i(x_1)\chi_j(x_2)\cdots\chi_k(x_N)$ are equal up to sign if $|c| = |c_{ij\ldots k}|$ and the permutation sum is antisymmetric. This means the spaces are equal, since each pair $c, -c$ defines the same two antisymmetric functions as the pair $c_{ij\ldots k}, -c_{ij\ldots k}$ if $|c| = |c_{ij\ldots k}|$. Now all relevant spaces are listed.

$$H_{\text{full}} \supset H_{\text{antisymmetric}} \supset O_{\text{full}} = O_{\text{FCI}} \supset O_{\text{HF}} \tag{3.13}$$

The spaces indicated with an H are infinite dimensional spaces. H_{full} is the full Hilbert space that the electronic Hamiltonian works on, including the symmetric functions. $H_{\text{antisymmetric}}$ is the subspace of antisymmetric functions in H_{full} and solving inside this space would result in the exact solution. The spaces indicated with an O are finite dimensional spaces determined by some set of $2K$ spin orbitals. O_{full} is the space of all possible N-electron wavefunctions built from the set of orbitals. It is equal to O_{FCI}, the space spanned by all $\binom{2K}{N}$ Slater determinants. If the orbital set is infinite these spaces can become complete in $H_{\text{antisymmetric}}$, allowing one to find exact results in them. Lastly O_{HF} is the space made out of all different single Slater determinants that can be created with the orbital set.

3.3 Electron Integrals

In this section the notation conventions for the integrals that result from H_e are introduced. The convention shortens the notation of integrals over Slater determinants. These integrals will show up naturally when looking at the energy of Slater determinants. In this section an expression for the energy of a single Slater determinant (without loss of generality the Hartree-Fock ground state $|\Phi_0\rangle$) will be given and the mentioned notation conventions for integrals will be defined as they appear. Consider the energy of the Hartree-Fock ground state $\langle\Phi_0|H_e|\Phi_0\rangle$. The the three operators in H_e are split into two categories. The operators $-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2$ and $\sum_{i=1}^{N}\sum_{I=1}^{M}\frac{Z_I}{R_{iI}}$ work on only one electron, while $\sum_{i<j}^{N}\frac{1}{R_{ij}}$ works on two electrons at a time. These and the one- and two-electron energies E_1 and E_2 will be considered separately, starting with the one-electron operators and energy. Because the derivation is not that insightful it is not included in [14] or [15]. The derivation below is a generalization of the small examples that are given instead. A different derivation which uses a completely different notation can be found in [17].

3.3.1 The One-electron Integral

The one-electron operators restricted to single electrons are called the core-Hamiltonian, which is given by:

$$h_i = -\frac{1}{2}\nabla_i^2 - \sum_{I=1}^{M}\frac{Z_I}{R_{iI}} \tag{3.14}$$

Since the one electron part of H_e is equal to $\sum_{i=1}^{N} h_i$ It is simple to see that $E_1 = \sum_{i=1}^{N} \langle \Phi_0 | h_i | \Phi_0 \rangle$. Consider this energy of one core-Hamiltonian.

$$\langle \Phi_0 | h_i | \Phi_0 \rangle = \int dx_1 \dots dx_N \left[\frac{1}{\sqrt{N!}} (\dots) \right]^* h_i \left[\frac{1}{\sqrt{N!}} (\dots) \right] \tag{3.15}$$

here the dots inside the brackets represent all Hartree-products that build the Slater determinant. Note that h_i only acts on orbitals with coordinate x_i. This means that the integrals over all $dx_j, j \neq i$ can be taken out (so can the constant). The next step is to multiply out all terms of the determinant and then take out everything independent of x_i. The result will be a large sum ($(N!)^2$ terms) with terms of the following shape:

$$\pm \frac{1}{N!} \int dx_1 \chi_a^*(x_1) \chi_b(x_1) \int dx_2 \chi_c^*(x_2) \chi_d(x_2) \cdots \int dx_i \chi_e^*(x_i) h_i \chi_f(x_i) \tag{3.16}$$

Now any pair χ_a, χ_b with the same coordinate will make the entire term 0 if $a \neq b$ due to the orthogonality of the spatial orbitals. The only entries in the sum that are non-zero are the ones that follow from multiplying the same terms in the adjoint and the regular Slater determinant. This is because any difference would result in two integrals having different orbitals and, while the integral with h_i would still be non-zero, the other integral would be zero and cancel out the term. For each permutation $\mathcal{P}$ of the indices $1, \dots, N$ there is a term in the Slater determinant and thus also in the sum. $\langle \Phi_0 | h_i | \Phi_0 \rangle$ is equal to:

$$\frac{1}{N!} \sum_{\mathcal{P}} \int dx_1 \chi_{\mathcal{P}(1)}^*(x_1) \chi_{\mathcal{P}(1)}(x_1) \int dx_2 \chi_{\mathcal{P}(2)}^*(x_2) \chi_{\mathcal{P}(2)}(x_2) \cdots \int dx_i \chi_{\mathcal{P}(i)}^*(x_i) h_i \chi_{\mathcal{P}(i)}(x_i) \tag{3.17}$$

Because of the previously mentioned orthogonality this simplifies easily to

$$\langle \Phi_0 | h_i | \Phi_0 \rangle = \frac{1}{N!} \sum_{\mathcal{P}} \int dx_i \chi_{\mathcal{P}(i)}^*(x_i) h_i \chi_{\mathcal{P}(i)}(x_i) \tag{3.18}$$

Since the sum is over all permutations each orbital gets to have the x_i coordinate the same amount of times. There are $N!$ permutations and therefore $(N-1)!$ times each orbital has coordinates x_i. With this the following expression for a single core-Hamiltonian is reached:

$$\langle \Phi_0 | h_i | \Phi_0 \rangle = \frac{1}{N!} \left(\sum_{j=1}^{N} (N-1)! \int dx_i \chi_j^*(x_i) h_i \chi_j(x_i) \right) \tag{3.19}$$

Before the sum over all core-Hamiltonians is taken to get an expression for E_1, the one-electron integral notation convention is introduced. For some operator $\mathcal{O}$ define

$$[i|\mathcal{O}|j] = \langle i|\mathcal{O}|j\rangle = \int dx_1 \chi_i^*(x_1) \mathcal{O} \chi_j(x_1) \tag{3.20}$$

and analogously without an operator the notation

$$[i|j] = \int dx_1 \chi_i^*(x_1) \chi_j(x_1). \tag{3.21}$$

Using this we can write the one-electron energy of the HF ground state as

$$E_1 = \sum_{i=1}^{N} \frac{1}{N!} \sum_{j=1}^{N} (N-1)! [j|h_i|j] \tag{3.22}$$

Now because $[j|h_i|j]$ is independent of i, as it is just the integration variable, the outer sum can be replaced with a factor N and the subscript of h_i can be dropped. Renaming j to i afterwards gives the final expression for E_1:

$$E_1 = \sum_{i=1}^{N} [i|h|i] \tag{3.23}$$

3.3.2 The Two-electron Integral

Next consider the two-electron energy E_2. Just like with the one-electron energy only single terms of the sum are considered. For E_2 this means each term of $\sum_{i<j}^{N} \langle \Phi_0 | r_{ij}^{-1} | \Phi_0 \rangle$. A similar thing will happen where for each

of these i, j a large sum of integrals follows from determinant multiplication. Again the terms independent of the operator are taken out of the main integral. The terms will have the shape

$$\pm \frac{1}{N!} \int dx_1 \chi_a^*(x_1)\chi_b(x_1) \int dx_2 \chi_c^*(x_2)\chi_d(x_2) \cdots \int dx_i dx_j \chi_e^*(x_i)\chi_f^*(x_j) r_{12}^{-1} \chi_g(x_i)\chi_h(x_j) \tag{3.24}$$

Once again the terms vanish if in one of the integrals without the operator the orbitals are not equal due to orthogonality. They only do not vanish if all terms without the operator are 1. Therefore the orbitals in the integral with the operator have to be a pair, otherwise at least one of the integrals without the operator would be zero. At this point the notation for this integral with the operator is introduced:

$$\langle ij|kl\rangle = [ik|jl] = \int dx_1 dx_2 \chi_i^*(x_1)\chi_j^*(x_2) r_{12}^{-1} \chi_k(x_1)\chi_l(x_2) \tag{3.25}$$

which is called the physicists' notation and

$$[ij|kl] = \langle ik|jl\rangle = \int dx_1 dx_2 \chi_i^*(x_1)\chi_j(x_1) r_{12}^{-1} \chi_k^*(x_2)\chi_l(x_2) \tag{3.26}$$

which is referred to as the chemists' notation. The current expression of E_2 looks like this

$$E_2 = \sum_{i<j}^{N} \frac{1}{N!}(N-2)! \sum_{ab}^{N} [\langle ab|ab\rangle - \langle ab|ba\rangle] \tag{3.27}$$

where the $(N-2)!$ is the amount of times the term occurs and is not zero (all combinations of coordinates and orbitals in the other integrals). The signs can be determined by looking the difference between the adjoint and the normal term in the Slater determinant. For all orbitals not in this integral the coordinates have to be the same so the sign starts out as positive. Then for $\langle ab|ab\rangle$ they are again the same so the sign is positive, while for $\langle ab|ba\rangle$ there is exactly one coordinate change between the adjoint and normal term resulting in a negative sign. Again since the integrals are independent of i, j, as they are just integration constants, the outer sum can be replaced with a factor $\binom{N}{2}$. This gives

$$E_2 = \binom{N}{2} \frac{1}{N!}(N-2)! \sum_{ab}^{N} [\langle ab|ab\rangle - \langle ab|ba\rangle] = \frac{1}{2} \sum_{ab}^{N} [\langle ab|ab\rangle - \langle ab|ba\rangle] \tag{3.28}$$

Combining Equation 3.23 and Equation 3.28 and switching to the chemists' notation the electronic Hamiltonian energy of a Slater determinant is found:

$$E_e = \sum_i [i|h|i] + \frac{1}{2} \sum_{ij} \left([ii|jj] - [ij|ji]\right). \tag{3.29}$$

Chapter 4

Hartree-Fock and FCI

In this section two classical methods for solving the electronic Hamiltonian are treated. They are called Hartree-Fock and Full Configuration Interaction (Full CI/FCI). While the methods are described in some detail the main take away from this section should be which approximations are used in which method and how accurate they are. Note that, since our starting point will be the electronic Hamiltonian (Equation 3.4), two approximations have already been made. Firstly, a non-relativistic Hamiltonian was taken as starting point and secondly the Born-Oppenheimer Approximation was applied. The third approximation that is made in both methods is that the solution is in a subspace built with a finite amount of basis-functions (orbitals). Because this subspace is not complete, although it is ideally taken to as complete as possible, the solution will not be exact.

Now we will discuss some of the main takeaways from this section, mainly the differences between both methods. The Hartree-Fock method looks at single Slater determinant solutions to reduce computing time and simplify the algorithm, while Full CI uses the full linear combination of all Slater determinants. The result is that even when the subspace approaches completeness the Hartree-Fock solution is not the exact solution of the electronic Hamiltonian. The limit of the Hartree-Fock solution when the orbital basis set reaches completeness is called the **Hartree-Fock limit**. Full CI however, approaches the exact solution of the electronic Hamiltonian when the basis set approaches completeness. This limit is called the **Full-CI limit**. The difference between both limits is called the **correlation energy**, as HF does not consider correlation, while FCI does. The drawback of FCI is long running times for bigger basis sets, which are required for both better results and bigger molecules. This chapter is based on [14] and [15], with some additional explanation and derivation steps.

4.1 Hartree-Fock Theory

4.1.1 Non-canonical Hartree-Fock equations

The starting point of this derivation is already the before mentioned assumption that the groundstate has a single Slater determinant wavefunction $|\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\rangle$. Consider the functional $E_0(\{\chi_i\}_{i=1}^N) = \langle\Psi_0|H_e|\Psi_0\rangle$. To start E_0 is written as:

$$E_0(\{\chi_i\}_{i=1}^N) = \sum_{a=1}^{N}[a|h|a] + \frac{1}{2}\sum_{ab}\left([aa|bb] - [ab|ba]\right) \tag{4.1}$$

The goal is to minimize E_0 with respect to $\{\chi_i\}_{i=1}^N$ (from here on $\{\chi_i\}$) under the constraint that the spin orbitals are orthonormal

$$\int dx_1 \chi_a^*(x_1)\chi_b(x_1) = [a|b] = \delta_{ab} \tag{4.2}$$

For this the method of Largrange multipliers is used. The multipliers are written as ϵ_{ab} and at the functional of interest is $\mathcal{L}$ given by:

$$\mathcal{L}(\{\chi_i\}) = E_0(\{\chi_i\}) - \sum_{ab}\epsilon_{ab}([a|b] - \delta_{ab}) \tag{4.3}$$

Note firstly that since $[a|b] = [b|a]^*$, $\epsilon_{ab} = \epsilon_{ba}^*$, which means the matrix ϵ_{ab} is Hermitian, a fact needed later. Then by the Lagrange multiplier method for any set $\{\hat{\chi}_i\}$ that minimizes E_0 and satisfies the orthonormality

constraint there is a matrix ϵ_{ab} such that $\mathcal{L}(\{\hat{\chi_i}\}, \epsilon_{ab})$ is a stationary point. The stationary points of $\mathcal{L}$ are found by introducing a small pertubation in the spin orbitals $\chi_a \to \chi_a + \delta\chi_a$ and setting $\delta\mathcal{L}$ (only the first order) to zero:

$$\delta\mathcal{L} = \delta E_0 - \sum_{ab} \epsilon_{ab}\delta[a|b] = 0. \tag{4.4}$$

Here $\delta[a|b] = [\delta a|b] + [a|\delta b]$. This follows directly since the perturbation in a constant is zero. For the first term (δE_0) it holds that:

$$\delta E_0 = \sum_a \Big([\delta a|h|a] + [a|h|\delta a]\Big) +$$

$$\frac{1}{2}\sum_{ab}\Big([\delta aa|bb] + [a\delta a|bb] + [aa|\delta bb] + [aa|b\delta b] - [\delta ab|ba] - [a\delta b|ba] - [ab|\delta ba] - [ab|b\delta a]\Big). \tag{4.5}$$

Where the first sum can simply be rewritten:

$$\sum_a([\delta a|h|a] + [a|h|\delta a]) = \sum_a([\delta a|h|a] + [\delta a|h|a]^*) = \sum_a [\delta a|h|a] + \text{ complex conjugate.} \tag{4.6}$$

The second sum requires the following:

$$[\delta ij|kl] = [kl|\delta ij] = [i\delta j|kl]^* = [kl|i\delta j]^*, \tag{4.7}$$

Now the sum can be rewritten as follows:

$$\frac{1}{2}\sum_{ab}([\delta aa|bb] + [\delta aa|bb]^* + [\delta aa|bb] + [\delta aa|bb]^* - [\delta ab|ba] - [\delta ab|ba]^* - [\delta ab|ba] - [\delta ab|ba]^*)$$

$$= \sum_{ab}([\delta aa|bb] - [\delta ab|ba]) + \text{ complex conjugate.} \tag{4.8}$$

Now the second term of $\delta\mathcal{L}$ can be rewritten:

$$\sum_{ab}\epsilon_{ab}\delta[a|b] = \sum_{ab}\epsilon_{ab}[\delta a|b] + [a|\delta b] = \sum_{ab}\epsilon_{ab}[\delta a|b] + \sum_{ab}\epsilon_{ab}[a|\delta b]$$

$$= \sum_{ab}\epsilon_{ab}[\delta a|b] + \sum_{ab}\epsilon_{ba}[b|\delta a] = \sum_{ab}\epsilon_{ab}[\delta a|b] + \sum_{ab}\epsilon^*_{ab}[\delta a|b]^* = \sum_{ab}\epsilon_{ab}[\delta a|b] + \Big(\sum_{ab}\epsilon_{ab}[\delta a|b]\Big)^*$$

$$= \sum_{ab}\epsilon_{ab}[\delta a|b] + \text{ complex conjugate.} \tag{4.9}$$

The result is

$$\delta\mathcal{L} = \sum_a [\delta a|h|a] + \sum_{ab}([\delta aa|bb] - [\delta ab|ba]) - \sum_{ab}\epsilon_{ab}[\delta a|b] + \text{ complex conjugates} = 0. \tag{4.10}$$

At this point the Coulomb and Exchange operators are introduced:

$$\mathcal{J}_b(1)\chi_a(1) = \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_b(2)\chi_a(1) \tag{4.11}$$

$$\mathcal{K}_b(1)\chi_a(1) = \int dx_2 \chi_b^*(2) r_{12}^{-1} \chi_a(2)\chi_b(1) \tag{4.12}$$

Using this we write out $\delta\mathcal{L}$:

$$\delta\mathcal{L} = \sum_a \int dx_1 \delta\chi_a^*(1)\Big[\big(h(1) + \sum_b(\mathcal{J}_b(1) - \mathcal{K}_b(1))\chi_a(1) - \sum_b \epsilon_{ab}\chi_b(1)\Big] + \text{ complex conjugate} = 0 \tag{4.13}$$

Both the sum over a above and its complex conjugate have to be zero for the equation to hold and since $\delta\chi_a^*$ is arbitrary the term in the square brackets needs to be zero:

$$\Big[h(1) + \sum_b(\mathcal{J}_b(1) - \mathcal{K}_b(1))\Big]\chi_a(1) = \sum_b \epsilon_{ab}\chi_b(1) \tag{4.14}$$

The term in the square brackets is the definition of the Fock operator f and the non-canonical Hartree-Fock equations follow directly:

$$f|\chi_a\rangle = \sum_{b=1}^{N} \epsilon_{ab}|\chi_b\rangle \tag{4.15}$$

Before the canonical Hartree-Fock equations are derived a simplified integral notation for the Fock operator has to be introduced. For this the permutation operator $\mathcal{P}_{12}$ is needed, it is defined by:

$$\mathcal{P}_{12}\chi_1(x_1)\chi_2(x_2) = \chi_2(x_1)\chi_1(x_2). \tag{4.16}$$

With this it is possible to write the Exchange operator in the same form as the Coulomb operator with $r_{12}^{-1}\mathcal{P}_{12}$ in place of r_{12}^{-1}. The Fock operator in integral form can now be written as:

$$f(x_1) = h_1(x_1) + \sum_b \int dx_2 \chi_b^* r_{12}^{-1}(1 - \mathcal{P}_{12})\chi_b. \tag{4.17}$$

4.1.2 Canonical Hartree-Fock equations

Starting from the non-canonical Hartree-Fock equations the Canonical Hartree-Fock equations can be derived using an unitary transformation on the spin orbitals. A unitary transformation satisfies:

$$U^\dagger U = UU^\dagger = I, \tag{4.18}$$

and thus also

$$1 = \det(U^\dagger U) = \det(U^\dagger)\det(U) = \det(U)^*\det(U) = |\det(U)|^2 = 1. \tag{4.19}$$

Therefore it follows that $\det(U) = e^{i\phi}$. This is relevant because it means that after the unitary transformation of the spin orbitals $|\Psi_0\rangle$ only changes by a phase factor. This is easily shown. We start with the transformed spin orbitals:

$$\chi'_a = \sum_b U_{ab}\chi_b \tag{4.20}$$

Using the matrix A of the relevant spin orbitals the Hartree-Fock ground state can be written as:

$$|\Psi_0\rangle = \frac{1}{\sqrt{N!}}\det(A). \tag{4.21}$$

The transformed ground state can be rewritten as follow:

$$|\Psi'_0\rangle = \frac{1}{\sqrt{N!}}\det(A') = \frac{1}{\sqrt{N!}}\det(U)\det(A) = \det(U)|\Psi_0\rangle. \tag{4.22}$$

This shows $|\Psi_0\rangle$ only changes by a phase factor. The next step is showing that the Fock operator is invariant under unitary transformation. It suffices to show that the Coulomb and Exchange operator are invariant under unitary transformation. First the Coulomb operator:

$$\begin{aligned}
\sum_b \mathcal{J}'_b(1)\chi_a(1) &= \sum_b \int dx_2 \chi_b'^*(2) r_{12}^{-1}\chi'_b(2)\chi_a(1) \\
&= \sum_b \int dx_2 (\sum_c U_{cb}^*\chi_c^*(2)) r_{12}^{-1} (\sum_d U_{db}\chi_d(2))\chi_a(1) \\
&= \sum_{cd} \big[\sum_b U_{cb}^* U_{db}\big] \int dx_2 \chi_c^*(2) r_{12}^{-1}\chi_d(2)\chi_a(1) \\
= \sum_{cd} \delta_{dc} \int dx_2 \chi_c^*(2) r_{12}^{-1}\chi_d(2)\chi_a(1) &= \sum_c \int dx_2 \chi_c^*(2) r_{12}^{-1}\chi_c(2)\chi_a(1) = \sum_c \mathcal{J}_c(1)\chi_a(1).
\end{aligned} \tag{4.23}$$

and now the Exchange operator:

$$\sum_b \mathcal{K}'_b(1)\chi_a(1) = \sum_b \int dx_2 \chi_b'^*(2) r_{12}^{-1}\chi'_a(2)\chi_b(1)$$

$$= \sum_b \int dx_2 (\sum_c U^*_{cb} \chi^*_c(2)) r_{12}^{-1} \chi_a(2) (\sum_d U_{db} \chi_d(1))$$

$$= \sum_{cd} [\sum_b U^*_{cb} U_{db}] \int dx_2 \chi^*_c(2) r_{12}^{-1} \chi_a(2) \chi_d(1)$$

$$= \sum_{cd} \delta_{dc} \int dx_2 \chi^*_c(2) r_{12}^{-1} \chi_a(2) \chi_d(1) = \sum_c \int dx_2 \chi^*_c(2) r_{12}^{-1} \chi_a(2) \chi_c(1) = \sum_c \mathcal{K}_c(1) \chi_a(1). \tag{4.24}$$

Using the fact that ϵ is Hermitian the canonical Hartree-Fock equations are found quickly. Because ϵ is Hermitian we can diagonalize it with some unitary transformation:

$$\epsilon' = U^\dagger \epsilon U, \tag{4.25}$$

and because the Fock operator is invariant under unitary transformation we get the following by using the unitary transformation on the non-canonical equations:

$$f|\chi'_a\rangle = \epsilon'_a |\chi'_a\rangle. \tag{4.26}$$

which are the canonical Hartree-Fock equations when we remove the primes:

$$f|\chi_a\rangle = \epsilon_a |\chi_a\rangle. \tag{4.27}$$

The canonical Hartree-Fock equations are once again some form of eigenvalue problem, now with the orbital energies ϵ_a as eigenvalues. Note that this is not an eigenvalue problem, since this is a set of N equations that have to be solved at the same time, with f depending on all orbitals instead of just the one it is working on.

4.1.3 Orbital Energies and Koopmans' Theorem

Before continuing with the Hartree-Fock method, some results can be derived from the canonical HF equations. The main interest in this section is Koopmans' Theorem, which states the following: Given an N-electron Hartree-Fock single determinant $|^N\Psi_0\rangle$ with occupied and virtual spin orbital energies ϵ_c and ϵ_r, then the ionization potential to produce an $(N-1)$-electron single determinant $|^{N-1}\Psi_c\rangle$ with identical spin orbitals, obtained by removing an electron from spin orbital χ_c, and the electron affinity to produce an $(N+1)$-electron single determinant $|^{N+1}\Psi_r\rangle$ with identical spin orbitals obtained by adding an electron to spin orbital χ_r , are just $-\epsilon_c$ and $-\epsilon_r$, respectively [14]. This will be used later on in a trick to approximate orbitals as constants to reduce dimensionality. First an expression for the orbital energies has to be derived. Start with multiplying the canonical Hartree-Fock equations by $\langle\chi_i|$:

$$\langle\chi_i|f|\chi_a\rangle = \epsilon_a \langle\chi_i|\chi_a\rangle = \epsilon_a \delta_{ia} \tag{4.28}$$

Here ϵ_a is the orbital energy of $|\chi_a\rangle$. Use this to get an expression for ϵ_a:

$$\epsilon_a = \langle\chi_a|f|\chi_a\rangle = \langle\chi_a|h + \sum_b (\mathcal{J}_b - \mathcal{K}_b)|\chi_a\rangle$$

$$= \langle\chi_a|h|\chi_a\rangle + \sum_b (\langle\chi_a|\mathcal{J}_b|\chi_a\rangle - \langle\chi_a|\mathcal{K}_b|\chi_a\rangle)$$

$$= \langle a|h|a\rangle + \sum_b (\langle ab|ab\rangle - \langle ab|ba\rangle) = \langle a|h|a\rangle + \sum_b \langle ab||ab\rangle \tag{4.29}$$

Here the sums are over occupied orbitals only and the notational convention $\langle ab||ab\rangle \equiv \langle ab|ab\rangle - \langle ab|ba\rangle$ is used. Note that the expression for a virtual orbital sums over N terms, while the one for a occupied orbital effectively sums over $N-1$ since $\langle aa||aa\rangle = 0$. To show Koopmans' Theorem the energy differences between $|^N\Psi_0\rangle$ and $|^{N-1}\Psi_c\rangle$ and between $|^N\Psi_0\rangle$ and $|^{N+1}\Psi_r\rangle$ have to be calculated. It should be clear that these energies are given by:

$$^NE_0 = \langle {}^N\Psi_0|H_e|{}^N\Psi_0\rangle, \quad {}^{N-1}E_c = \langle {}^{N-1}\Psi_c|H_e|{}^{N-1}\Psi_c\rangle, \quad {}^{N+1}E_r = \langle {}^{N+1}\Psi_r|H_e|{}^{N+1}\Psi_r\rangle$$

As we saw before these energy expectation values are of the following form, where the sum goes over all occupied states (Equation 3.29 after notation change):

$$E = \sum_a^{\text{occupied}} \langle a|h|a\rangle + \frac{1}{2} \sum_{ab}^{\text{occupied}} \langle ab||ab\rangle$$

Now deriving the Ionization potential is easy as it is just the terms that involve c:

$$IP = {}^{N-1}E_c - {}^{N}E_0 = -\langle a|h|a\rangle - \frac{1}{2}\sum_a \langle ac||ac\rangle - \frac{1}{2}\sum_b \langle cb||cb\rangle$$

$$= -\langle a|h|a\rangle - \sum_b \langle cb||cb\rangle = -\epsilon_c$$

Analogously it can be shown that for the Electron Affinity it holds that:

$$EA = {}^{N}E_0 - {}^{N+1}E_r = -\epsilon_r$$

with this Koopmans' Theorem is shown and as hinted at before it will be used to remove active orbitals and electrons from the system in section 7.1.

4.1.4 Restricted (closed-shell) Hartree-Fock

The derivation for the remainder of the Hartree-Fock method will be under the assumption that each spatial orbital either has both related spin orbitals occupied or neither. This approximation is called restricted or closed-shell. Not only does it remove spin, it also removes half of the terms for each sum in the expression. The reason this approximation is used for the derivation is because the non restricted derivation is very similar. It separates the terms for both spin types and is analogous but with more bookkeeping. Therefore the deviation for the restricted equations is shown, together with the final result for the unrestricted method. In the following its is assumed without loss of generality that N is even. The first step is recalling the definition of the spin orbitals:

$$\chi_i = \psi_i(r)\alpha(\omega), \chi_{i+1} = \psi_i(r)\beta(\omega) \text{ for } i \text{ even} \tag{4.30}$$

For spatial orbital ψ_i with spin-function β we write $\bar{\psi}_i$ and now the ground state looks like:

$$|\Psi_0\rangle = |\chi_1\chi_2\cdots\chi_{N-1}\chi_N\rangle = |\psi_1\bar{\psi}_1\cdots\psi_a\bar{\psi}_a\cdots\psi_{N/2}\bar{\psi}_{N/2}\rangle. \tag{4.31}$$

The next step is integrating out the spin functions from the original Hartree-Fock equation and simplifying the remaining terms back into recognizable shape.

$$f(x_1)\chi_i(x_1) = \epsilon_i\chi_i(x_i) \tag{4.32}$$

The following part is the same for both spin functions so without loss of generality χ_i is assumed to have spin function α.

$$f(x_1)\psi_i(r_1)\alpha(\omega_1) = \epsilon_i\psi_i(r_1)\alpha(\omega_1) \tag{4.33}$$

By multiplying by $\alpha^*(\omega_1)$ on the left and integrating over ω_1 the spin can be eliminated from the right hand side:

$$\left[\int d\omega_1\alpha^*(\omega_1)f(x_1)\alpha(\omega_1)\right]\psi_i(r_1) = \epsilon_i\psi_i(r_1) \tag{4.34}$$

The result is an expression of the form $f(r_1)\psi_i(r_1) = \epsilon_i\psi_i(r_1)$ and the next step is evaluating this operator $f(r_1) = \int d\omega_1\alpha^*(\omega_1)f(x_1)\alpha(\omega_1)$, the closed-shell Fock operator. This is done by first substituting the spin orbital Fock operator:

$$\left[\int d\omega_1\alpha^*(\omega_1)f(x_1)\alpha(\omega_1)\right]\psi_i(r_1) = \left[\int d\omega_1\alpha^*(\omega_1)h(r_1)\alpha(\omega_1)\right]\psi_i(r_1)$$

$$+\left[\sum_a \int d\omega_1 dx_2\alpha^*(\omega_1)\chi_a^*(x_2)r_{12}^{-1}(1-\mathcal{P})\chi_a(x_2)\alpha(\omega_1)\right]\psi_i(r_1) \tag{4.35}$$

Next the integral over ω_1 in the term with h is worked out and the permutation operator is used to write the exchange term separately:

$$f(r_1)\psi_i(r_1) = h(r_1)\psi_i(r_1) + \sum_a \int d\omega_1 dx_2\alpha^*(\omega_1)\chi_a^*(x_2)r_{12}^{-1}\chi_a(x_2)\alpha(\omega_1)\psi_i(r_1)$$

$$-\sum_a \int d\omega_1 dx_2\alpha^*(\omega_1)\chi_a^*(x_2)r_{12}^{-1}\chi_a(x_1)\alpha(\omega_2)\psi_i(r_2) = \epsilon_i\psi_i(r_1) \tag{4.36}$$

Now the spin orbitals are split based on their spin function and get two sums over $N/2$ instead of each single sum over N.

$$
\begin{aligned}
f(r_1)\psi_i(r_1) &= h(r_1)\psi_i(r_1) \\
&+ \sum_a^{N/2} \int d\omega_1 d\omega_2 dx_2 \alpha^*(\omega_1)\psi_a^*(r_2)\alpha^*(\omega_2) r_{12}^{-1}\psi_a(r_2)\alpha(\omega_2)\alpha(\omega_1)\psi_i(r_1) \\
&+ \sum_a^{N/2} \int d\omega_1 d\omega_2 dx_2 \alpha^*(\omega_1)\psi_a^*(r_2)\beta^*(\omega_2) r_{12}^{-1}\psi_a(r_2)\beta(\omega_2)\alpha(\omega_1)\psi_i(r_1) \\
&- \sum_a^{N/2} \int d\omega_1 d\omega_2 dx_2 \alpha^*(\omega_1)\psi_a^*(r_2)\alpha^*(\omega_2) r_{12}^{-1}\psi_a(r_1)\alpha(\omega_1)\alpha(\omega_2)\psi_i(r_2) \\
&- \sum_a^{N/2} \int d\omega_1 d\omega_2 dx_2 \alpha^*(\omega_1)\psi_a^*(r_2)\beta^*(\omega_2) r_{12}^{-1}\psi_a(r_1)\beta(\omega_1)\alpha(\omega_2)\psi_i(r_2) \\
&= \epsilon_i \psi_i(r_1)
\end{aligned} \tag{4.37}
$$

Integrating over ω_1 and ω_2 the last term vanishes and in the others the spin dependency is removed. The two coulomb terms are equal after integration and the result is:

$$
\begin{aligned}
f(r_1)\psi_i(r_1) = h(r_1)\psi_i(r_1) &+ 2\left[\sum_a^{N/2} \int dr_2 \psi_a^*(r_2) r_{12}^{-1}\psi_a(r_2)\right]\psi_i(r_1) \\
&- \left[\sum_a^{N/2} \int dr_2 \psi_a^*(r_2) r_{12}^{-1}\psi_i(r_2)\right]\psi_a(r_1) = \epsilon_i\psi_i(r_1)
\end{aligned} \tag{4.38}
$$

Now the closed-shell Fock operator is found:

$$
f(r_1) = h(r_1) + \sum_a^{N/2} \int dr_2 \psi_a^*(r_2)(2 - \mathcal{P}_{12}) r_{12}^{-1}\psi_a(r_2), \tag{4.39}
$$

or in other notation:

$$
f(1) = h(1) + \sum_a^{N/2} 2\mathcal{J}_a(1) - \mathcal{K}_a(1). \tag{4.40}
$$

Here $\mathcal{J}_a$ and $\mathcal{K}_a$ are analogously defined for spatial orbitals compared to how they were defined for spin orbitals in Equation 4.11 and Equation 4.12. At this point the Hartree-Fock theory ends, as the next steps are part of the standard solving method and not derivations of general equations. The result for unrestricted Hartree-Fock is shown, once again with the remark that the derivation is very similar and does not add much to the restricted derivation, except tedious algebra. It is then once more ignored in the remainder of the section.

4.2 Solving Hartree-Fock

4.2.1 Roothaan Equations

With the spin eliminated $f(r_1)\psi_i(r_1) = \epsilon_i\psi_i(r_1)$ has to be solved and to do this the spatial orbitals will have to be defined. The spatial orbitals are assumed to be in the space spanned by a set of K know basic functions $\{\phi_\mu(r)|\mu = 1, 2, \ldots, K\}$:

$$
\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu \tag{4.41}
$$

Switching index the equation to be solved changes into:

$$
f(1)\sum_\nu^K C_{\nu i}\phi_\nu(1) = \epsilon_i \sum_\nu^K C_{\nu i}\phi_\nu(1) \tag{4.42}
$$

The Roothaan Equations follow from writing this expression into matrix form and for that the left has to be multiplied by $\phi_\mu^*(1)$ and integrated:

$$\sum_\nu^K C_{\nu i} \int dr_1 \phi_\mu^*(1) f(1) \phi_\nu(1) = \epsilon_i \sum_\nu^K C_{\nu i} \int dr_1 \phi_\mu^*(1) \phi_\nu(1) \tag{4.43}$$

We define matrices using the two integrals; the overlap matrix $\mathbf{S}$ and the Fock matrix $\mathbf{F}$. We define them by their elements:

$$S_{\mu\nu} = \int dr_1 \phi_\mu^*(1) \phi_\nu(1) \tag{4.44}$$

$$F_{\mu\nu} = \int dr_1 \phi_\mu^*(1) f(1) \phi_\nu(1) \tag{4.45}$$

We rewrite the integrated Hartree-Fock equation into the Roothaan equations:

$$\sum_\nu F_{\mu\nu} C_{\nu i} = \epsilon_i \sum_\nu S_{\mu\nu} C_{\nu i} \tag{4.46}$$

If we make the matrix $\boldsymbol{\epsilon}$ by putting $\epsilon_1, \ldots, \epsilon_K$ on the diagonal and make the matrix C from the $C_{\nu i}$ as defined above we can write this in matrix form:

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{4.47}$$

Before continuing the unrestricted case is considered one more time. Up to this point it has not deviated from the restricted case much and instead of the Roothaan equations, the Pople-Nesbet-Berthier equations are found:

$$\begin{aligned} \mathbf{F}^\alpha \mathbf{C}^\alpha &= \mathbf{S}\mathbf{C}^\alpha \boldsymbol{\epsilon}^\alpha \\ \mathbf{F}^\beta \mathbf{C}^\beta &= \mathbf{S}\mathbf{C}^\beta \boldsymbol{\epsilon}^\beta \end{aligned} \tag{4.48}$$

Where α and β correspond to the spin up and spin down orbitals. The coupling here comes from the fact that the Fock matrices have terms related to orbitals of the other spin as well. This can be explained intuitively because the Fock matrix has to consider the average field of all electrons, not just the ones with the same spin.

4.2.2 Self-Consistent-Field method

The reason why this problem is hard to solve is that $\mathbf{F}$ depends on $\mathbf{C}$ ($\mathbf{F} = \mathbf{F}(\mathbf{C})$). This makes the problem non-linear and therefore we will solve it using an iterative method where we guess $\mathbf{F}$, use that guess to calculate $\mathbf{C}$ and use that to calculate the next $\mathbf{F}$. First we will introduce a charge density matrix $\mathbf{P}$ that is directly related to $\mathbf{C}$. The field defined by $\mathbf{P}$ gives this method its name. With this charge density matrix we can split the parts of $\mathbf{F}$ that depend on $\mathbf{P}$ (or $\mathbf{C}$) and those that do not. Then we show that we can remove $\mathbf{S}$ from the equations by changing the basis of the ϕ_μ's to a orthogonal one. Lastly the SCF-method is written out in a step by step overview.

The charge density/probability density function of an electron described by spatial wave function $\psi_a(r)$ is of course $|\psi_a(r)|^2$. In a closed-shell system we can write the total charge density:

$$\rho(r) = 2 \sum_a^{N/2} |\psi_a(r)|^2 \tag{4.49}$$

which is of course the sum of all charge densities of all electrons in the system. Now $\rho(r)dr$ is the probability of finding a electron within dr of r. Rewriting into the basis of ϕ_μ's we get:

$$\rho(r) = 2 \sum_a^{N/2} \psi_a^*(r) \psi_a(r)$$

$$\rho(r) = 2 \sum_a^{N/2} \sum_\nu C_{\nu a}^* \phi_\nu^*(r) \sum_\mu C_{\mu a} \phi_\mu(r)$$

$$\rho(r) = \sum_{\mu\nu} \left[2 \sum_a^{N/2} C_{\mu a} C_{\nu a}^* \right] \phi_\nu^*(r) \phi_\mu(r)$$

$$\rho(r) = \sum_{\mu\nu} P_{\mu\nu} \phi_\nu^*(r) \phi_\mu(r) \tag{4.50}$$

Where the matrix $\mathbf{P}$ defined by the elements $P_{\mu\nu}$ is the matrix we eluded to before. Next we will show that it is more natural to define $\mathbf{F}$ as a function of $\mathbf{P}$ than as a function of $\mathbf{C}$.

4.2.3 The Fock Matrix

The elements of the Fock Matrix are (as we saw before) given by:

$$F_{\mu\nu} = \int dr_1 \phi_\mu^*(1) f(1) \phi_\nu(1)$$

$$F_{\mu\nu} = \int dr_1 \phi_\mu^*(1) h(1) \phi_\nu(1) + \sum_a^{N/2} \int dr_1 \phi_\mu^*(1) \big[2\mathcal{J}_a(1) - \mathcal{K}_a(1)\big] \phi_\nu(1) \tag{4.51}$$

We now define the Core Hamiltonian $H_{\mu\nu}^{\text{core}} = \int dr_1 \phi_\mu^*(1) h(1) \phi_\nu(1)$. which does not depend on $\mathbf{F}$ or $\mathbf{C}$. This term will be consistent throughout our iteration because of this and is thus nice to split off. We continue with this substitution and write out the second term as well:

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_a^{N/2} 2(\mu\nu|aa) - (\mu a|a\nu) \tag{4.52}$$

We can now get clear dependency on $\mathbf{P}$ by writing the orbitals a introduced by the Coulomb and Exchange operators into the same basis:

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_a^{N/2} \sum_{\lambda\sigma} C_{\lambda a} C_{\sigma a}^* \big[2(\mu\nu|\sigma\lambda) - (\mu\lambda|\sigma\nu)\big] \tag{4.53}$$

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + G_{\mu\nu} \tag{4.54}$$

With this we have written the Fock Matrix in a form where we can split off the part that does not change during the iteration easily ($H_{\mu\nu}^{\text{core}}$) such that we only calculate it once. Additionally we have a relatively simple dependency on $\mathbf{P}$ that we can use to iterate. However the integrals $(\mu\nu|\sigma\lambda)$ and $(\mu\lambda|\sigma\nu)$ are very big in number. Because of this calculating all of them is the big problem in Hartree-Fock calculation.

4.2.4 Orthogonalization of the basis (removal of S)

We step back and look at the Roothaan equations in matrix form:

$$\mathbf{F}(\mathbf{C})\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\epsilon} \tag{4.55}$$

Ideally the overlap matrix $\mathbf{S}$ is the identity matrix. In this case it is removed from the equation and the result is of a standard matrix eigenvalue problem form:

$$\mathbf{F}\mathbf{C} = \mathbf{C}\boldsymbol{\epsilon} \tag{4.56}$$

This is the same as requiring the basis functions to be orthonormal to each other. It is clear that we can find a transformation matrix $\mathbf{X}$ such that the ϕ_μ's are orthogonal after transformation:

$$\phi'_\mu = \sum_\nu X_{\nu\mu} \phi_\nu \text{ with } \int dr_1 \phi_\mu'^*(1) \phi'_\nu(1) = \delta_{\mu\nu} \ \forall \mu, \nu = 1, 2, \ldots, K \tag{4.57}$$

We inspect the properties of $\mathbf{X}$ by substituting the definition of ϕ'_μ:

$$\int dr_1 \phi_\mu'^*(1) \phi'_\nu(1) = \int dr_1 \sum_\lambda X_{\lambda\mu}^* \phi_\lambda^*(r) \sum_\sigma X_{\sigma\nu} \phi_\sigma \tag{4.58}$$

We interchange integration and sums and take the X's out of the integral to get:

$$= \sum_\lambda \sum_\sigma X_{\lambda\mu}^* \int dr_1 \phi_\lambda^*(r) \phi_\sigma X_{\sigma\nu}$$

$$= \sum_\lambda \sum_\sigma X_{\lambda\mu}^* S_{\lambda\sigma} X_{\sigma\nu} = \delta_{\mu\nu} \tag{4.59}$$

Which in matrix notation gives us the following relation $\mathbf{X}$ should satisfy:

$$\mathbf{X}^\dagger \mathbf{S} \mathbf{X} = \mathbf{1} \tag{4.60}$$

Before we show a way to obtain such an $\mathbf{X}$ we can define, since $\mathbf{S}$ is Hermitian, a matrix $\mathbf{U}$ that diagonalizes $\mathbf{S}$:

$$\mathbf{U}^\dagger \mathbf{S} \mathbf{U} = \mathbf{s} \tag{4.61}$$

Where $\mathbf{s}$ is the diagonal matrix with the eigenvalues of $\mathbf{S}$. We use the following definition of $\mathbf{X}$, doing this is called canonical orthogonalization:

$$\mathbf{X} = \mathbf{U}\mathbf{s}^{-1/2} \tag{4.62}$$

With this we mean that the columns of $\mathbf{U}$ are divided by the square root of the corresponding eigenvalue:

$$X_{ij} = U_{ij}/s_j^{1/2} \tag{4.63}$$

Now we have to worry about the linear dependency of the basis set. If there is linear dependence of the basis functions some of the s_j's will approach zero. However we have a trick: first we sort the eigenvalues of $\mathbf{S}$ from big to small in $\mathbf{s}$. We can do this when we also sort the columns of $\mathbf{U}$ (and thus also $\mathbf{X}$) as well. Now we can remove the m eigenvalues we think are to small and truncate $\mathbf{X}$ by removing the last m columns. The result is a set of $K - m$ orthonormal basis functions that span the same space as the original set of K basis functions. With this fix we can assume that there is no linear dependence and continue on by defining $\mathbf{C}'$ and $\mathbf{F}'$ and deriving the transformed Roothaan equations. Note that in the first step we require $\mathbf{X}$ to be invertible, which should be the case if we have no linear dependencies:

$$\mathbf{C}' = \mathbf{X}^{-1}\mathbf{C} \tag{4.64}$$

The Roothaan equations in matrix form become:

$$\mathbf{F}\mathbf{X}\mathbf{C}' = \mathbf{S}\mathbf{X}\mathbf{C}'\boldsymbol{\epsilon} \tag{4.65}$$

We multiply on the left by $\mathbf{X}^\dagger$ and get:

$$(\mathbf{X}^\dagger\mathbf{F}\mathbf{X})\mathbf{C}' = (\mathbf{X}^\dagger\mathbf{S}\mathbf{X})\mathbf{C}'\boldsymbol{\epsilon} \tag{4.66}$$

Defining $\mathbf{F}' = \mathbf{X}^\dagger\mathbf{F}\mathbf{X}$ we get:

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\boldsymbol{\epsilon} \tag{4.67}$$

the transformed Roothaan Equations.

4.2.5 SCF Method

Now the SCF Method can be written out in a step by step structure. Below is a list of the steps and a flowchart is included in Figure 4.1.

1. Specify the system: the nuclear coordinates $\{R_A\}$, the atomic numbers $\{Z_A\}$, the number of electrons N and a basis set $\{\phi_\mu\}$.
2. Calculate $\mathbf{S}$, $H_{\mu\nu}^{\text{core}}$ and the two electron integrals of the shape $(\mu\nu|\lambda\sigma)$
3. Diagonalize $\mathbf{S}$ and calculate $\mathbf{X}$
4. Obtain a guess for $\mathbf{P}$
5. Calculate $\mathbf{G}$ with this $\mathbf{P}$
6. Add $\mathbf{G}$ to $H_{\mu\nu}^{\text{core}}$ to obtain $\mathbf{F}$
7. Transform $\mathbf{F}$ into $\mathbf{F}' = \mathbf{X}^\dagger\mathbf{F}\mathbf{X}$
8. Diagonalize $\mathbf{F}'$ to obtain $\mathbf{C}'$ and $\boldsymbol{\epsilon}$
9. Calculate $\mathbf{C} = \mathbf{X}\mathbf{C}'$
10. Form a new $\mathbf{P}$ using this $\mathbf{C}$
11. Check if the scheme has converged (this $\mathbf{P}$ is the same as (or close by) the one before). If not return to step 5 with the new $\mathbf{P}$.
12. The output that represents the solution is the matrices $\mathbf{C}$,$\mathbf{P}$,$\mathbf{F}$

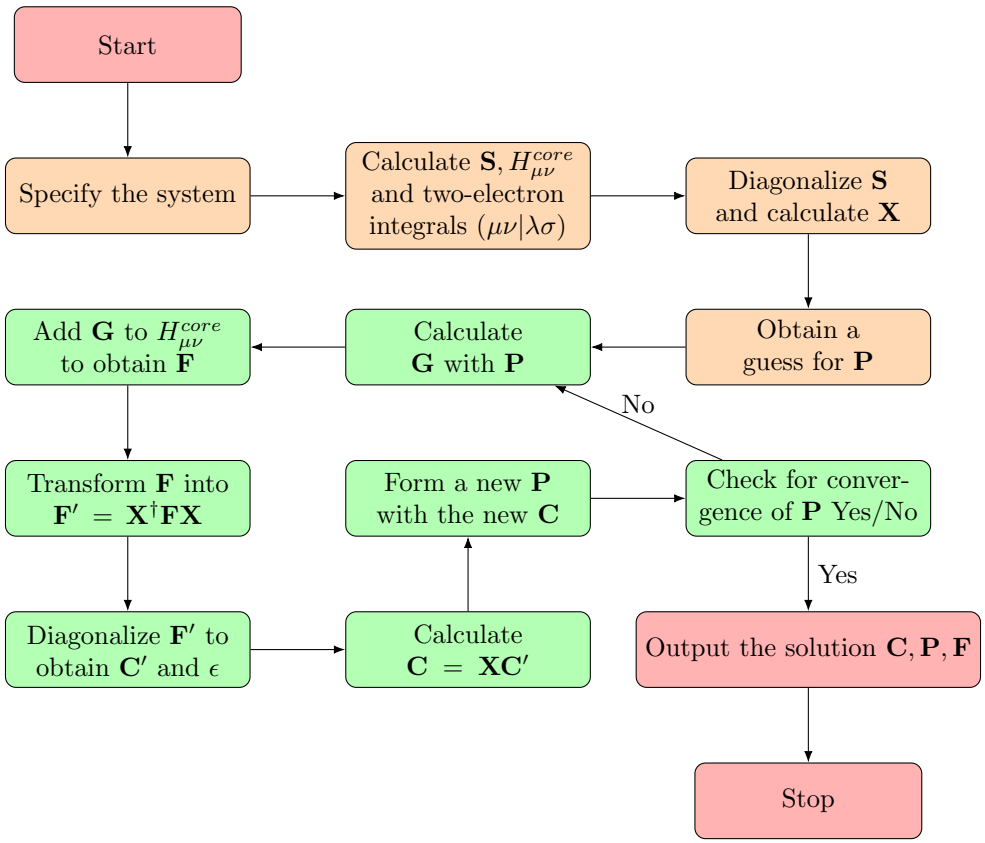


Figure 4.1: Flow chart of SCF HF. Orange Blocks are preparation steps and green blocks are the iteration cycle.

4.2.6 Orbital choice/Orbital basis functions

A subject that has not been discussed yet, is the orbital choice. All choices that will discussed are based on atomic orbitals. Atomic orbitals are based on the exact solutions of the hydrogen orbitals. They are either a Slater or a Gaussian type function. The $1s$ Slater type function is

$$\phi_{1s}^{SF}(\zeta, r-R) = \left(\frac{\zeta^3}{\pi}\right)^{1/2} e^{-\zeta|r-R|}, \tag{4.68}$$

and the Gaussian type $1s$ function is

$$\phi_{1s}^{GF}(\alpha, r-R) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha|r-R|^2}. \tag{4.69}$$

Here ζ and α are parameters of the function and R is the location of the orbital. Slater Type Orbitals (STO) are better representations of wavefunctions than Gaussian functions, but the two electron integrals are not calculable with STO. The solution to this problem that is used is approximating STO basis functions with multiple Gaussian functions. Gaussian functions have easily calculable integrals and using them was first proposed by Boys in 1950 [18]. These orbitals are called STO-NG; they are Slater type orbitals approximated by N Gaussian functions. These basis functions are called contracted basis functions. Looking once again at the $1s$ Slater type function:

$$\phi_{1s}^{SF}(\zeta, r-R) \approx \sum_{i=1}^{N} d_{i,1s}\phi_{1s}^{GF}(\alpha_{i,1s}, r-R_i) \tag{4.70}$$

where the R_i's are often equal to R. If you fix N the optimal values of the α_i's and d_i's are fixed as well and can be looked up in databases such as `basissetexchange.org` [19]. Only 1s, 2p, 3d etc Gaussians are used. The 2s 3p 3s etc. Gaussians do not simplify the integrals and can be closely approximated by Gaussians that do. If this is done with only 1s Gaussians (Lobe Basis set) the R_i's are not equal to R, because you place them at different points in space to approximate higher order orbitals. In the project we often use minimal basis sets: a single STO-NG function for each atomic orbital. Instead you may want to have a loose and a tight version of each orbital and let the algorithm choose which one is best. Typically this is only used for valence orbitals.

There are several ways to improve an orbital basis set. First you can change an STO-NG orbital set into an STO-(N+1)G basis set of the same orbitals. The functions will be more similar to a Slater type orbital and thus

better. An example is switching from an STO-3G minimal basis to an STO-6G minimal basis. A second way to improve an basis set is by taking an Gaussian out of one of the contracted basis functions and making it its own basic function. (E.g. take the Gaussian with the highest d_i use a Gaussian with that α_i as separate basis function. Then recalculate the d_i's for the other contracted basis function which now has $N-1$ Gaussians). Finally you can add orbitals with higher quantum numbers than required. For example adding d-type functions to the atoms Li through F and possibly also p-type functions to H and He. One basically adds the next few atomic orbitals that are not considered to the basis set. These functions are called polarization functions and they are used because they approximate the effect of an external electric field on the electron cloud of the atom.

Next to STO-NG, notation is done in the following form: X-XG**, where the dash separates the core and valence electrons. X can be multiple numbers each represents a basis function, the number itself is the amount of Gaussians used to approximate the Slater orbitals. E.g 4-31G means core electrons are contracted from 4 Gaussians while valence electrons have two basis functions for each orbital: one with 3 Gaussians and a separate single Gaussian. If asterisks are added a single one means that d-type orbitals are added to Li-F and two asterisks mean that d-type orbitals are added to Li-F and p-type orbitals are added to H (and He). In the main results of this report only orbital sets created using an STO-3G orbital basis set are considered. A minimal basis set is required for the system to stay small enough. Linear combinations of the chosen atomic orbitals can be used as molecular orbital sets. These can be made in different ways and the one used in this project is letting Hartree-Fock optimize the molecular orbitals. It is argued however that the best method instead is to linearize the one-electron integral matrix to get molecular orbitals, as these orbitals are computationally nicer to handle [9].

4.3 Full CI

Compared to Hartree-Fock, the derivation Full CI method is a lot shorter. Recall the basis set of the orbital space from Equation 3.10:

$$|\Phi\rangle = c_0|\Phi_0\rangle + \sum_{ar} c_a^r|\Phi_a^r\rangle + \sum_{\substack{a<b\\r<s}} c_{ab}^{rs}|\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c\\r<s<t}} c_{abc}^{rst}|\Phi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d\\r<s<t<u}} c_{abcd}^{rstu}|\Phi_{abcd}^{rstu}\rangle + \dots. \tag{4.71}$$

FCI creates a Hamiltonian matrix for this basis set and finding the lowest eigenvalue gives the solution of the problem due to the variational principle. The creation of the FCI matrix is rather straightforward. First, for clarity the basis is rewritten as a block vector:

$$|\Phi\rangle = c_0|\Phi_0\rangle + c_S|S\rangle + c_D|D\rangle + c_T|T\rangle + c_Q|Q\rangle + \dots, \tag{4.72}$$

where $|S\rangle$ is shorthand for all single excited Slater determinants, $|D\rangle$ is shorthand for all doubly excited Slater determinants and so on. The coefficients $c_S, C_D, \dots$ are of course shorthand for the original coefficients. The FCI matrix is just the Hamiltonian matrix and has elements $H_{i,j} = \langle\phi_i|H_e|\phi_j\rangle$, where H_e is the electronic Hamiltonian and ϕ_i and ϕ_j basis vectors. Using the rewritten block basis the FCI matrix looks like this:

$$\begin{array}{c} \\ \langle\Phi_0| \\ \langle S| \\ \langle D| \\ \langle T| \\ \langle Q| \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} |\Phi_0\rangle & |S\rangle & |D\rangle & |T\rangle & |Q\rangle & \cdots \end{array} \\ \begin{bmatrix} \langle\Phi_0|H_e|\Phi_0\rangle & \langle\Phi_0|H_e|S\rangle & \langle\Phi_0|H_e|D\rangle & \langle\Phi_0|H_e|T\rangle & \langle\Phi_0|H_e|Q\rangle & \cdots \\ \langle S|H_e|\Phi_0\rangle & \langle S|H_e|S\rangle & \langle S|H_e|D\rangle & \langle S|H_e|T\rangle & \langle S|H_e|Q\rangle & \cdots \\ \langle D|H_e|\Phi_0\rangle & \langle D|H_e|S\rangle & \langle D|H_e|D\rangle & \langle D|H_e|T\rangle & \langle D|H_e|Q\rangle & \cdots \\ \langle T|H_e|\Phi_0\rangle & \langle T|H_e|S\rangle & \langle T|H_e|D\rangle & \langle T|H_e|T\rangle & \langle T|H_e|Q\rangle & \cdots \\ \langle Q|H_e|\Phi_0\rangle & \langle Q|H_e|S\rangle & \langle Q|H_e|D\rangle & \langle Q|H_e|T\rangle & \langle Q|H_e|Q\rangle & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array} \tag{4.73}$$

This matrix can be simplified by using Brillouin's Theorem [20]. The theorem states that for an Hartree-Fock ground state $|\Phi_0\rangle$ and $|\Phi_a^r\rangle$ a singly excited state created by removing orbital a from $|\Phi_0\rangle$ and adding virtual orbital r. Then it holds that:

$$\langle\Phi_0|H_e|\Phi_a^r\rangle = 0. \tag{4.74}$$

It can be proven that this equation is in fact equivalent to the Hartree-Fock equations and it is provable directly from the variational principle when restricted to Slater determinants [21]. A proof of just the theorem can be found in [22].

Another simplification comes from the fact that any two terms that differ by more than two spin orbitals are zero. This will not only cancel out entire blocks, it will also show that the blocks that are still left are sparse.

For example $\langle \Phi_{ab}^{rs} | H_e | \Phi_{cdef}^{tuvw} \rangle$ is only non-zero when $a, b \in \{c, d, e, f\}$ and $r, s \in \{t, u, v, w\}$. The proof about the matrix elements is not too hard, but some of the theory from the next section is used in it. Therefore the proof is included in Appendix A instead. Using both results the FCI matrix looks like this:

$$\begin{array}{c} \\ \langle \Phi_0 | \\ \langle S | \\ \langle D | \\ \langle T | \\ \langle Q | \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccccc} |\Phi_0\rangle & |S\rangle & |D\rangle & |T\rangle & |Q\rangle & \cdots \end{array} \\ \begin{bmatrix} \langle \Phi_0 | H_e | \Phi_0 \rangle & 0 & \langle \Phi_0 | H_e | D \rangle & 0 & 0 & \cdots \\ 0 & \langle S | H_e | S \rangle & \langle S | H_e | D \rangle & \langle S | H_e | T \rangle & 0 & \cdots \\ \langle D | H_e | \Phi_0 \rangle & \langle D | H_e | S \rangle & \langle D | H_e | D \rangle & \langle D | H_e | T \rangle & \langle D | H_e | Q \rangle & \cdots \\ 0 & \langle T | H_e | S \rangle & \langle T | H_e | D \rangle & \langle T | H_e | T \rangle & \langle T | H_e | Q \rangle & \cdots \\ 0 & 0 & \langle Q | H_e | D \rangle & \langle Q | H_e | T \rangle & \langle Q | H_e | Q \rangle & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{array} \tag{4.75}$$

It can easily be seen that this matrix scales very quickly in size. Since there are $\binom{2K}{N}$ Slater determinants, the system scales at least as $\left(\frac{2K}{N}\right)^N$. However a bigger system does not require only a bigger N, as there are more electrons, it also requires a bigger K as more orbitals are needed to get results. Additionally going from a small basis set to a bigger one will increase K even further. On top of that the matrix has a size of $\binom{2K}{N}^2$ since it has length and height $\binom{2K}{N}$ and will scale at least as $\left(\frac{2K}{N}\right)^{2N}$. A result of this is that even for water the FCI solution of the basis set 6-31G* took too long to calculate to be worth it for the next section where HF and FCI will be compared for water.

Lastly [14] introduces an approximation to the Full CI basis that is interesting, since the method is very similar to the main trick attempted in this project. The argument is the following: $|\Phi_0\rangle$ will most likely be a very important part of $|\Phi\rangle$. Now the intermediate normalized form is introduced by setting c_0 from Equation 4.71 to 1.

$$|\Phi\rangle = |\Phi_0\rangle + \sum_{ar} c_a^r |\Phi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c_{ab}^{rs} |\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c_{abc}^{rst} |\Phi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d \\ r<s<t<u}} c_{abcd}^{rstu} |\Phi_{abcd}^{rstu}\rangle + \dots . \tag{4.76}$$

This form is not normalized and after renormalizing it by multiplying each term with some constant c', the book stops and says that means this this notation can easily be brought back to the compete basis. This is true, but looking at the individual functions each $c_{a\dots}^{r\dots}$ has been changed into $c' c_{a\dots}^{r\dots}$. This normalized function looks like this and varying the $c_{a\dots}^{r\dots}$ still gives the full basis:

$$|\Phi\rangle = c'|\Phi_0\rangle + \sum_{ar} c' c_a^r |\Phi_a^r\rangle + \sum_{\substack{a<b \\ r<s}} c' c_{ab}^{rs} |\Phi_{ab}^{rs}\rangle + \sum_{\substack{a<b<c \\ r<s<t}} c' c_{abc}^{rst} |\Phi_{abc}^{rst}\rangle + \sum_{\substack{a<b<c<d \\ r<s<t<u}} c' c_{abcd}^{rstu} |\Phi_{abcd}^{rstu}\rangle + \dots . \tag{4.77}$$

Most importantly for later the coefficient for the Hartree-Fock solution has increased. This is easy to show since c' has to be lower than 1 (as it is the new coefficient for c_0). This means all coefficients have decreased, while their sum of absolute squares is still normalized. This means that $c' > c_0$ and therefore the coefficient for the Hartree-Fock ground state has increased from the starting point.

4.4 Comparison of HF and FCI for H_2O

In this section the things that have been discussed in the previous sections will be illustrated with some example calculations on H_2O. First different basis sets are compared. For this calculation (and most others in this report), the geometry of the water molecule has a fixed bond angle equal to the optimal angle according to STO-3G Hartree-Fock. This is a reasonable choice, as when looking at a molecule without experimental results, it is the best estimation possible when only using Hartree-Fock in the preparation. This angle is $\sim 100°$, which is a slight bit off from the experimentally measured 104.45°. The bond length is varied around the HF optimal value, which is also quite close to the actual bond length of 0.9584Åat $\sim$ 1Å.

4.4.1 Comparison of basis sets

As discussed the choice of basis set matters a lot. It determines how good the subspace you are looking in is. Better (often bigger) basis sets have significantly better results than small ones. This is shown in Figure 4.2. As expected FCI will always outperform HF in the same basis set, but different basis sets can be so good that HF outperforms the FCI in a worse basis set. In fact the basis set of interest in this report, STO-3G, is outperformed by HF in all other basis sets in Figure 4.2. The reason this basis set is used, is simply because of

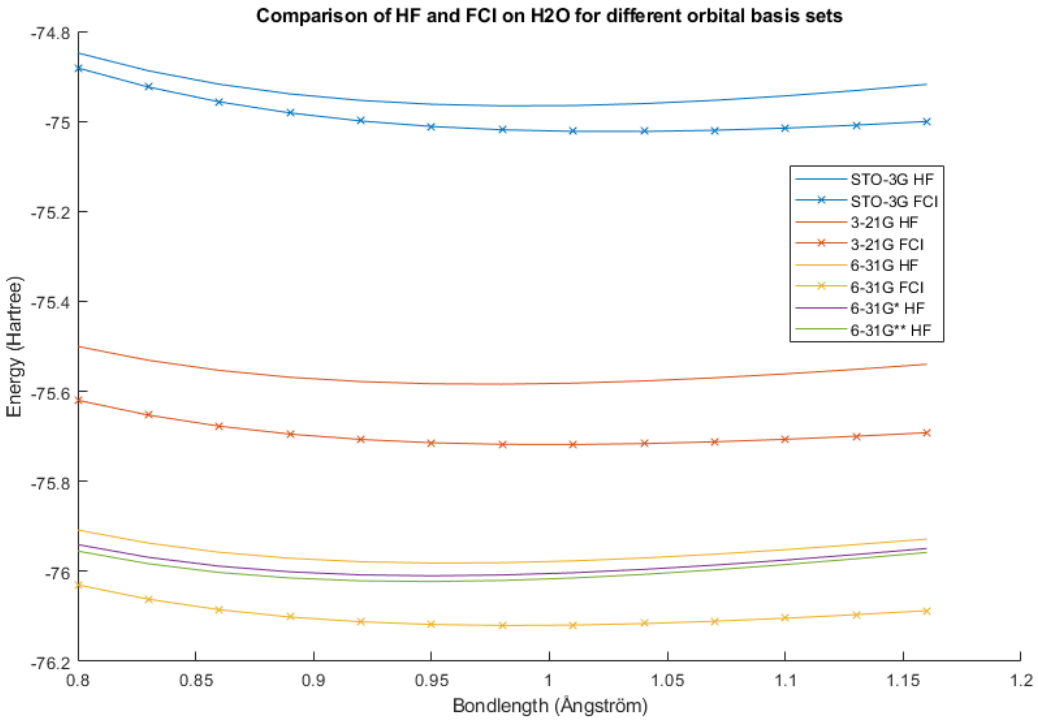


Figure 4.2: Comparison of different basis sets on H_2O. FCI and HF of the same basis set have the same color, but FCI has marks. 6-31G* and 6-31G** FCI are missing because the running time is already too long.

the requirements for higher basis sets. There are not really any molecules that are simulatable with VQE on a 6-31G basis set except H2, which stays too small to be of the size of interest.

4.4.2 Removal of one H atom

By pulling away one of the H atoms and letting the other stay in place it is possible to illustrate systematic errors that Hartree-Fock has due to the lack of electron correlation. In Figure 4.3 the energy of the system with constant bond angle of 100° and fixed bond length of 1Å for one H atom and variable bond length for the other H atom is shown for STO-3G HF and STO-3G FCI. It can be seen that the asymptotic behavior of the HF line is different to the one of FCI. This is a systematic error due to the lack of electron correlation in the single Slater determinant space. This illustrates why using HF on complex systems is considered not accurate enough.

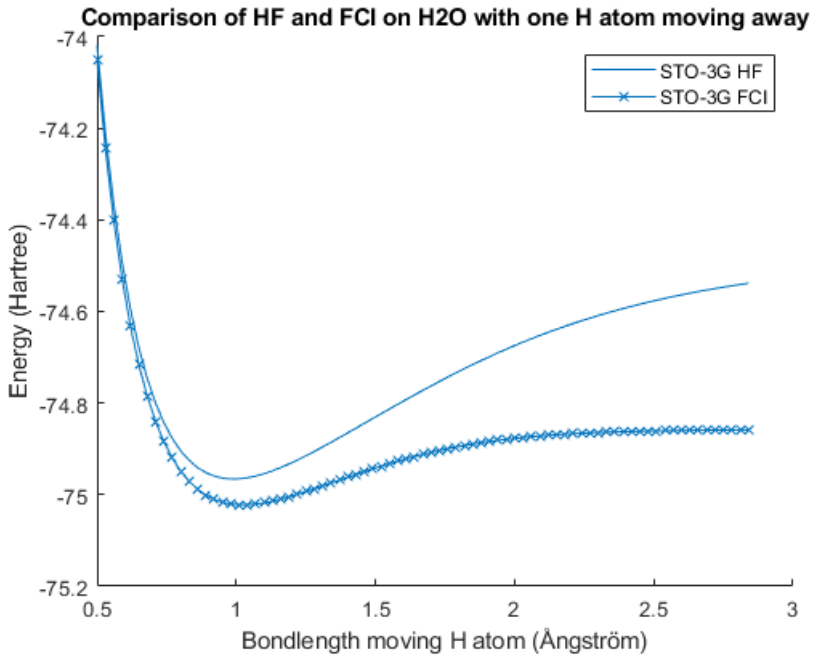


Figure 4.3: Comparison of asymptotic behavior for water when removing one of the H atoms from the molecule. Both STO-3G HF and STO-3G FCI are shown.

Chapter 5

VQE Introduction and Theory

The Variational Quantum Eigensolver is a hybrid algorithm that uses a quantum chip to speed up finding minimal eigenvalues of an operator. In this report the only operator of interest is the electronic Hamiltonian and its energy, but many problems can be tackled using the algorithm. As said, the algorithm is a hybrid one. This means that part of it is done on a normal (classical) computer and part of it is done on a quantum system (quantum chip/processor). The quantum chip does not have to be big in order to get results and VQE scales well with system size, contrary to classical methods. The algorithm works fairly intuitively where the part that is hard for the classical computer (sampling trial states) is instead done with the quantum processor. The optimization is still done classically, however using energies calculated with results from the quantum chip instead of from the classical computer it self. The algorithm can be described in the following steps.

1. The Hamiltonian (operator) of interest is transformed into a sum of Pauli operators. This is done with some fermionic transformation such as the Jordan-Wigner transform or the Bravyi-Kitaev transform. The theory behind this is discussed in this chapter.
2. The classical computer picks some quantum state $|q(\vec{\theta})\rangle$ to be sampled, fully defined by the rotation matrix $\vec{\theta}$.
3. The quantum state $|q(\vec{\theta})\rangle$ is created on the quantum chip. This is done by some system of rotational gates and entangling gates. It is important that all states can be reached and therefore in this project the depth of the circuit can be increased to allow more rotational and entangling gates, increasing the complexity of the problem, but restricting the search space less. State preparation is discussed in section 6.2.
4. The quantum state $|q(\vec{\theta})\rangle$ is measured on the quantum chip (see section 6.3). The classical computer calculates the energy E_q of $|q(\vec{\theta})\rangle$ using the measurements and one- and two-electron integrals.
5. The minimization algorithm determines the next step until convergence (return to 2). The convergence criteria can be chosen in different ways. Common criteria are: some fixed amount of iteration steps, some fixed amount of running time, convergence of the classical optimizer or getting within chemical accuracy of the solution (if the solution is known beforehand). Chemical accuracy is achieved if the error is below 0.0016 Hartree and is considered the error required to make accurate chemical predictions. Classical optimizers are discussed in section 6.4.

A simplified flowchart is given in Figure 5.1. Before moving on to the VQE algorithm some mathematics have to be done (specifically the orange block in the flow chart). As said, this chapter will discuss writing the electronic Hamiltonian in a form where qubits can represent quantum states $|q\rangle$ relating to electron occupation. This is done by first deriving the second quantization form of the Hamiltonian and then using the Jordan-Wigner transformation to transform this Hamiltonian into a qubit Hamiltonian. This chapter is based on [14] and [23], with some additional explanation and derivation steps.

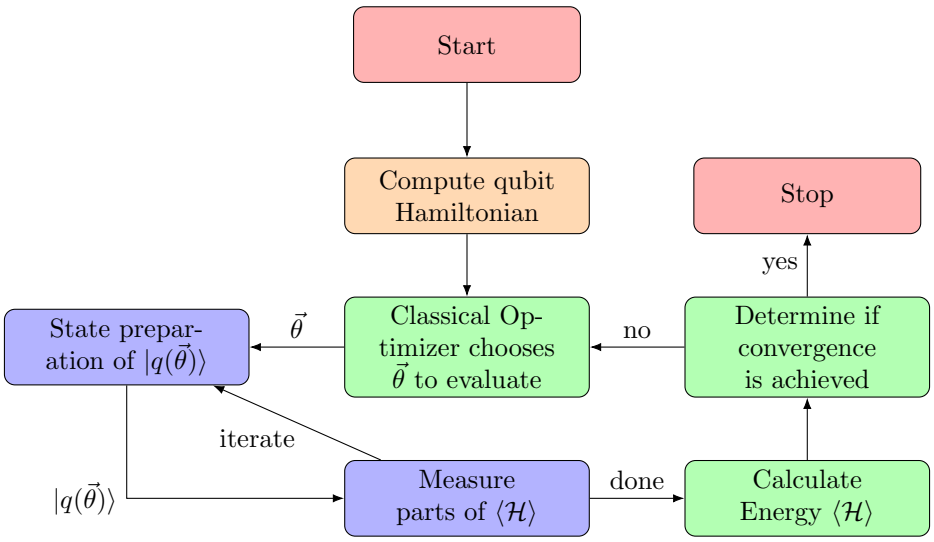


Figure 5.1: Simple flow chart of VQE. The orange block is a preparation step, the green blocks happen on the classical computer and the blue blocks are on the quantum processor.

5.1 Second Quantization

Since the antisymmetry principle is an axiom of quantum mechanics it needs to be "forced" onto either the Hamiltonian or wavefunction in some way, as the Hamiltonian can otherwise have non-antisymmetric wavefunctions as solutions. The normal Electronic Hamiltonian has this problem. Antisymmetry can be forced onto the Hamiltonian by defining it using the annihilation and creation operators, which have nice properties that ensure antisymmetry without having to force antisymmetry on the solution using Slater determinants.

5.1.1 Annihilation and Creation operators

Define the creation operator $a_i^\dagger$ by its action on an arbitrary Slater determinant: $a_i^\dagger|\chi_k\cdots\chi_l\rangle = |\chi_i\chi_k\cdots\chi_l\rangle$ with i and j not occupied. The following anticommutator relation of two creation operators follows:

$$\{a_i^\dagger, a_j^\dagger\} = a_i^\dagger a_j^\dagger + a_j^\dagger a_i^\dagger = 0. \tag{5.1}$$

This can be shown by introducing some arbitrary Slater determinant $|\chi_k\cdots\chi_l\rangle$:

$$(a_i^\dagger a_j^\dagger + a_j^\dagger a_i^\dagger)|\chi_k\cdots\chi_l\rangle = |\chi_i\chi_j\chi_k\cdots\chi_l\rangle + |\chi_j\chi_i\chi_k\cdots\chi_l\rangle = 0. \tag{5.2}$$

Where the last step is due to the antisymmetry of the Slater determinant. Notice that this in specific means that $a_i^\dagger a_i^\dagger = -a_i^\dagger a_i^\dagger = 0$ if i is not occupied. This leads to Pauli's exclusion principle: a second electron can not be added to the same spin orbital. Written differently this means that:

$$0 = a_i^\dagger a_i^\dagger|\chi_k\cdots\chi_l\rangle = a_i^\dagger|\chi_i\chi_k\cdots\chi_l\rangle, \tag{5.3}$$

and due to antisymmetry of the Slater determinant

$$a_i^\dagger|\chi_k\cdots\chi_l\rangle = 0 \text{ if } i \in \{k,\ldots,l\}. \tag{5.4}$$

With this knowledge it is clear that the case where i or j are occupied in Equation 5.1 is trivial and thus that it holds for all i and j.

Next is the annihilation operator a_i defined by: $a_i|\chi_i\chi_k\cdots\chi_l\rangle = |\chi_k\cdots\chi_l\rangle$. It is important to note that the annihilation operator can only work on the leftmost spin orbital. For example $a_i|\chi_k\chi_i\chi_l\rangle = -|\chi_i\chi_k\chi_l\rangle = -|\chi_k\chi_l\rangle \neq |\chi_k\chi_l\rangle$. The notation a_i and $a_i^\dagger$ suggests that the operators are adjoints ($(a_i^\dagger)^\dagger = a_i$), which is true and can be shown as follows:

$$|K\rangle = |\chi_i\chi_j\rangle = a_i^\dagger|\chi_j\rangle. \tag{5.5}$$

Take the adjoint

$$\langle K| = \langle\chi_j|(a_i^\dagger)^\dagger \tag{5.6}$$

and multiply by $|K\rangle$ on the right to get:

$$\langle K|K\rangle = 1 = \langle \chi_j|(a_i^\dagger)^\dagger|\chi_i\chi_j\rangle. \tag{5.7}$$

Since $\langle \chi_j|\chi_j\rangle = 1$ for consistency it is necessary that $(a_i^\dagger)^\dagger|\chi_i\chi_j\rangle = |\chi_j\rangle = a_i|\chi_i\chi_j\rangle$ and thus $(a_i^\dagger)^\dagger = a_i$. This means that a_i acts as a creation operator to the left and $a_i^\dagger$ acts as an annihilation operator to the left. From the anticommutator relation of the creation operator we can directly get the anticommutator relation of the annihilation operator by using that for two operators A and B it holds that $(AB)^\dagger = B^\dagger A^\dagger$. This results in

$$0 = \{a_i^\dagger, a_j^\dagger\} = a_i^\dagger a_j^\dagger + a_j^\dagger a_i^\dagger = a_j a_i + a_i a_j = \{a_j.a_i\}. \tag{5.8}$$

Analogous with the creation operator one can find $a_i a_i = -a_i a_i = 0$. This means the same spin orbital cannot be removed twice. The result is that, analogously with the creation operator not being able to make an orbital twice, we can not remove a orbital that is not there.

$$a_i|\chi_k \cdots \chi_l\rangle = 0 \text{ if } i \notin \{k, \ldots, l\} \tag{5.9}$$

Lastly it is necessary to know what happens if we use both operators together. This starts with determining $\{a_i, a_j^\dagger\}$. Consider the case where $i = j$. Once more take an arbitrary Slater determinant $|\chi_k \cdots \chi_l\rangle$. If χ_i is not occupied

$$(a_i a_i^\dagger + a_i^\dagger a_i)|\chi_k \cdots \chi_l\rangle = a_i a_i^\dagger|\chi_k \cdots \chi_l\rangle + 0 = a_i|\chi_i\chi_k \cdots \chi_l\rangle = |\chi_k \cdots \chi_l\rangle. \tag{5.10}$$

If χ_i is occupied

$$(a_i a_i^\dagger + a_i^\dagger a_i)|\chi_k \cdots \chi_i \cdots \chi_l\rangle = 0 + a_i^\dagger a_i|\chi_k \cdots \chi_i \cdots \chi_l\rangle = -a_i^\dagger a_i|\chi_i \cdots \chi_k \cdots \chi_l\rangle =$$

$$- a_i^\dagger|\cdots \chi_k \cdots \chi_l\rangle = -|\chi_i \cdots \chi_k \cdots \chi_l\rangle = |\chi_k \cdots \chi_i \cdots \chi_l\rangle. \tag{5.11}$$

Note that if χ_i is the first orbital the same happens but trivially. Since in all cases the arbitrary determinant is returned it follows that $\{a_i, a_i^\dagger\} = a_i a_i^\dagger + a_i^\dagger a_i = 1$. Next consider the case where $i \neq j$. Again $(a_i a_j^\dagger + a_j^\dagger a_i)|\chi_k \cdots \chi_l\rangle$ is evaluated. Note that this is zero unless χ_i appears and χ_j does not appear. Otherwise the outcome is zero due to $a_j^\dagger$ trying to create an orbital that is already there. Similarly the outcome would be zero if a_i tried to annihilate an orbital that is not there. However even if $i \in \{k, \ldots, l\}$ and $j \notin \{k, \ldots, l\}$ the outcome is still zero:

$$(a_i a_j^\dagger + a_j^\dagger a_i)|\chi_k \cdots \chi_i \cdots \chi_l\rangle = -(a_i a_j^\dagger + a_j^\dagger a_i)|\chi_i \cdots \chi_k \cdots \chi_l\rangle =$$

$$-a_i|\chi_j\chi_i \cdots \chi_k \cdots \chi_l\rangle - a_j^\dagger|\cdots \chi_k \cdots \chi_l\rangle = a_i|\chi_i\chi_j \cdots \chi_k \cdots \chi_l\rangle - |\chi_j \cdots \chi_k \cdots \chi_l\rangle$$
$$= |\chi_j \cdots \chi_k \cdots \chi_l\rangle - |\chi_j \cdots \chi_k \cdots \chi_l\rangle = 0. \tag{5.12}$$

Combining both results gives the anticommutation relation: $\{a_i, a_j^\dagger\} = a_i a_j^\dagger + a_j^\dagger a_i = \delta_{ij}$. This means creation and annihilation operators of different spin orbitals can be interchanged if we change the sign: $a_i a_j^\dagger = -a_j^\dagger a_i$ if $i \neq j$. Conversely if the operators are of the same spin orbital the formula is $a_i a_i^\dagger = 1 - a_i^\dagger a_i$.

5.1.2 Vacuum state and Occupation number representation

In the following the actions of the operators on any vector in the occupation number representation will be defined. To do this (and rewrite the Hamiltonian in its second quantization) the vacuum state $|\ \rangle$ is needed. This state represents a state of the system with no electrons. It has the following properties:

- It is normalized: $\langle\ |\ \rangle = 1$.
- Electrons cannot be removed from it: $a_i|\ \rangle = \langle\ |a_i^\dagger = 0$.
- Electrons can be added to it: $a_i^\dagger|\ \rangle = |\chi_i\rangle, \langle\ |a_i = \langle\chi_i|$
- Any Slater Determinant can be created by applying multiple creation operators: $a_k^\dagger \cdots a_l^\dagger|\ \rangle = |\chi_k \cdots \chi_l\rangle$

Now any Slater determinant with spin orbitals $\{\chi_1, \ldots, \chi_N\}$ can be equated to a state $|\alpha\rangle$ defined by the vector $(\alpha_1, \ldots, \alpha_N) \in \{0,1\}^N$, using the following definition of $|\alpha\rangle$:

$$|\alpha\rangle = (a_1^\dagger)^{\alpha_1} \cdots (a_N^\dagger)^{\alpha_N} |\ \rangle \tag{5.13}$$

With this the action of a_i and $a_i^\dagger$can be fully defined. In order to do so define $|\alpha^{(i)}\rangle$ as $|\alpha\rangle$ with α_i flipped ($0 \to 1$ or $1 \to 0$):

$$a_i|\alpha\rangle = \begin{cases} 0 & \text{if } \alpha_i = 0 \\ -(-1)^{s_\alpha^i}|\alpha^{(i)}\rangle & \text{if } \alpha_i = 1 \end{cases} \tag{5.14}$$

and

$$a_i^\dagger|\alpha\rangle = \begin{cases} -(-1)^{s_\alpha^i}|\alpha^{(i)}\rangle & \text{if } \alpha_i = 0 \\ 0 & \text{if } \alpha_i = 1. \end{cases} \tag{5.15}$$

Where $s_\alpha^i = \sum_{k=1}^{i-1} \alpha_k$ (for $i = 1$ it is 1). Note that in the first definition a one is changed into a zero in $|\alpha^{(i)}\rangle$ and in the second definition a zero into a one. This will be derived for a_i with the comment that $a_i^\dagger$ can be done analogously. The first case where $\alpha_i = 0$ is trivial. The case where $\alpha_i = 1$ goes as follows:

$$a_i|\alpha\rangle = a_i|\chi_k \cdots \chi_i \cdots \chi_l\rangle = \pm a_i|\chi_i\chi_k \cdots \chi_l\rangle = \pm|\chi_k \cdots \chi_l\rangle \tag{5.16}$$

It is clear that χ_i has to move to the front without changing the order of the χ's in front to get $|\alpha^{(i)}\rangle$ back at the end. This is done by doing exactly $s_\alpha^i = \sum_{k=1}^{i-1} \alpha_k$ spin orbital switches, corresponding to the $(-1)^{s_\alpha^i}$ term.

5.1.3 Second Quantization of the Hamiltonian

As seen earlier the Hamiltonian can be split into an one-electron operator and a two-electron operator part. Instead of deriving the second quantization it will be shown that the correct outcomes follow after starting with the correct expressions. These expressions are:

$$\mathcal{O}_1 = \sum_{ij} \langle i|h|j\rangle a_i^\dagger a_j, \tag{5.17}$$

$$\mathcal{O}_2 = \frac{1}{2} \sum_{ijkl} \langle ij|kl\rangle a_i^\dagger a_j^\dagger a_l a_k, \tag{5.18}$$

where the sums run over the spin orbitals $\{\chi_i\}$. Now it has to be shown that evaluating these operators corresponds to the evaluation of the one- and two-electron parts of the (Hartree-Fock) Hamiltonian. Evaluating with a single Slater determinant is enough, as we can build the FCI space using only Slater determinants. The determinant $|\Psi_0\rangle = |\chi_1 \cdots \chi_a \chi_b \cdots \chi_N$ is taken.

$$\langle \Psi_0|\mathcal{O}_1|\Psi_0\rangle = \sum_a [a|h|a] \tag{5.19}$$

$$\langle \Psi_0|\mathcal{O}_2|\Psi_0\rangle = \frac{1}{2} \sum_{ab} [aa|bb] - [ab|ba] \tag{5.20}$$

First is the one-electron operator:

$$\langle \Psi_0|\mathcal{O}_1|\Psi_0\rangle = \sum_{ij} \langle i|h|j\rangle \langle \Psi_0|a_i^\dagger a_j|\Psi_0\rangle \tag{5.21}$$

Since a_j and $a_i^\dagger$ are trying to annihilate an electron (to the right and to the left respectively), the indices i and j must belong to the set $a, b, ...$ corresponding to $|\Psi_0\rangle$. This reduces the sum to those indices only:

$$\langle \Psi_0|\mathcal{O}_1|\Psi_0\rangle = \sum_{ab} \langle a|h|b\rangle \langle \Psi_0|a_a^\dagger a_b|\Psi_0\rangle. \tag{5.22}$$

Now $a_a^\dagger a_b = \delta_{ab} - a_b a_a^\dagger$ is used to move $a_a^\dagger$ to the right:

$$\langle \Psi_0|a_a^\dagger a_b|\Psi_0\rangle = \delta_{ab}\langle \Psi_0|\Psi_0\rangle - \langle \Psi_0|a_b a_a^\dagger|\Psi_0\rangle \tag{5.23}$$

It immediately follows that this is equal to δ_{ab} since $\langle\Psi_0|\Psi_0\rangle = 1$ and the second term is zero since $a_a^\dagger$ is attempting to create an electron in a occupied spin orbital. The expected result follows:

$$\langle\Psi_0|\mathcal{O}_1|\Psi_0\rangle = \sum_{ab}\langle a|h|b\rangle\delta_{ab} = \sum_{a}\langle a|h|a\rangle. \tag{5.24}$$

Next up is the two-electron operator where immediately the indices $ijkl$ are replaced with $abcd$ using the same argument as used for the one-electron operator.

$$\langle\Psi_0|\mathcal{O}_2|\Psi_0\rangle = \frac{1}{2}\sum_{abcd}\langle ab|cd\rangle\langle\Psi_0|a_a^\dagger a_b^\dagger a_d a_c|\Psi_0\rangle \tag{5.25}$$

Again the anticommutator relation ($a_a^\dagger a_b = \delta_{ab} - a_b a_a^\dagger$) is used to move $a_a^\dagger$ and $a_b^\dagger$ to the far right such that the term vanishes.

$$\begin{aligned}\langle\Psi_0|a_a^\dagger a_b^\dagger a_d a_c|\Psi_0\rangle = \delta_{bd}\langle\Psi_0|a_a^\dagger a_c|\Psi_0\rangle - \langle\Psi_0|a_a^\dagger a_d a_b^\dagger a_c|\Psi_0\rangle = \\ \delta_{bd}\delta_{ac}\langle\Psi_0|\Psi_0\rangle - \delta_{bd}\langle\Psi_0|a_c a_a^\dagger|\Psi_0\rangle - \delta_{bc}\langle\Psi_0|a_a^\dagger a_d|\Psi_0\rangle + \langle\Psi_0|a_a^\dagger a_d a_c a_b^\dagger|\Psi_0\rangle \\ = \delta_{bd}\delta_{ac} - \delta_{bc}\delta_{ad}\langle\Psi_0|\Psi_0\rangle + \langle\Psi_0|a_d a_a^\dagger|\Psi_0\rangle = \delta_{bd}\delta_{ac} - \delta_{bc}\delta_{ad}\end{aligned} \tag{5.26}$$

Where the last few steps follow from the terms with creation operators on the far right vanishing as before and $\langle\Psi_0|\Psi_0\rangle = 1$. Looking at the operator again:

$$\langle\Psi_0|\mathcal{O}_2|\Psi_0\rangle = \frac{1}{2}\sum_{abcd}\langle ab|cd\rangle(\delta_{bd}\delta_{ac} - \delta_{bc}\delta_{ad}) = \frac{1}{2}\sum_{ab}\langle ab|ab\rangle - \langle ab|ba\rangle \tag{5.27}$$

Which is the correct result after a change of notation:

$$\langle\Psi_0|\mathcal{O}_2|\Psi_0\rangle = \frac{1}{2}\sum_{ab}[aa|bb] - [ab|ba] \tag{5.28}$$

This means that indeed for the Hamiltonian it holds that:

$$\mathcal{H} = \sum_{ij}\langle i|h|j\rangle a_i^\dagger a_j + \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle a_i^\dagger a_j^\dagger a_l a_k \tag{5.29}$$

5.2 Jordan-Wigner Transform

Next the Jordan-Wigner transform [24] is introduced. It maps a_i and $a_i^\dagger$ working on a vector $|\alpha\rangle$ to tensor products of quantum gates operating on a series of qubits that represent $|\alpha\rangle$. This is in fact fairly simple. Recall the definition of a_i:

$$a_i|\alpha\rangle = \begin{cases} 0 & \text{if } \alpha_i = 0 \\ -(-1)^{s_\alpha^i}|\alpha^{(i)}\rangle & \text{if } \alpha_i = 1 \end{cases} \tag{5.30}$$

The qubits will represent orbitals. Using a_i on qubit i should give appropriate outcomes. If qubit i is in state $|0\rangle$ the output has to be $|0\rangle$ for all qubits. Additionally if the qubit is in state $|1\rangle$ the output has to be $-(-1)^{s_\alpha^i}$ times the input with the ith qubit flipped. We can do this as follows:

$$a_i = -\Big(\otimes_{k=1}^{i-1} Z_k\Big) \otimes \sigma_i \tag{5.31}$$

Where Z_k is the Pauli operator given by $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ working on qubit k and $\sigma_i \equiv |0\rangle\langle 1|$ working on qubit i. It is clear that the first term with the Z operators takes care of the $(-1)^{s_\alpha^i}$ factor. The σ_i makes the ith qubit flip if it is $|1\rangle$ ($|0\rangle\langle 1|1\rangle = |0\rangle$) and makes the entire thing 0 if the qubit is $|0\rangle$ ($|0\rangle\langle 1|0\rangle = |0\rangle \cdot 0 = 0$). It should be clear that $a_i^\dagger$ is analogue to a_i:

$$a_i^\dagger = -\Big(\otimes_{k=1}^{i-1} Z_k\Big) \otimes \sigma_i^\dagger, \tag{5.32}$$

where $\sigma_i^\dagger = |1\rangle\langle 0|$ instead.

The Jordan-Wigner transform on the second quantization results in a qubit version of the occupation number representation. The 2^N possible states $|\alpha\rangle$ are the basis states of the qubit system. Additionally the Jordan-Wigner transform makes the Hamiltonian a sum of N-qubit Pauli operators. First define the single qubit Pauli operators by their matrices (the Pauli matrices):

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{5.33}$$

If multiple qubits are considered I_k, X_k, Y_k and Z_k are written to indicate the Pauli operator working on qubit k as previously. Define a N-qubit Pauli operator $\mathcal{P}_N$ as an operator that does a single Pauli operator on each of the N qubits, up to a constant of ± 1 or $\pm i$.

$$\mathcal{P}_N = \bigotimes_{k=1}^{N} i^{a_k} \sigma^{(k)}, \quad a_k \in \mathbb{N}, \quad \sigma^{(k)} \in \{I_k, X_k, Y_k, Z_k\} \tag{5.34}$$

Since σ_i and $\sigma_i^\dagger$ in the Jordan-Wigner transform can be written as a function of Pauli operators

$$\sigma_i = \frac{X + iY}{2}, \quad \sigma_i^\dagger = \frac{X - iY}{2}, \tag{5.35}$$

the Jordan-Wigner transform transforms the annihilation and creation operators into sums of N-qubit Pauli operators, since all unaffected qubits can be given I_k to fill the N-qubit operator. Doing this and splitting up the $\sigma_i/\sigma_i^\dagger$ into X_i and Y_i gives:

$$a_i = -\Big(\otimes_{k=1}^{i-1} Z_k \Big) \otimes \sigma_i \otimes \Big(\otimes_{k=i+1}^{N} I_k \Big) = \frac{1}{2}\mathcal{P}_{N,i}^{(1)} + \frac{1}{2}\mathcal{P}_{N,i}^{(2)} \tag{5.36}$$

$$a_i^\dagger = -\Big(\otimes_{k=1}^{i-1} Z_k \Big) \otimes \sigma_i^\dagger \otimes \Big(\otimes_{k=i+1}^{N} I_k \Big) = \frac{1}{2}\mathcal{P}_{N,i}^{(1)} - \frac{1}{2}\mathcal{P}_{N,i}^{(2)} \tag{5.37}$$

where

$$\mathcal{P}_{N,i}^{(1)} = -\Big(\otimes_{k=1}^{i-1} Z_k \Big) \otimes X_i \otimes \Big(\otimes_{k=i+1}^{N} I_k \Big), \quad \mathcal{P}_{N,i}^{(2)} = -\Big(\otimes_{k=1}^{i-1} Z_k \Big) \otimes iY_i \otimes \Big(\otimes_{k=i+1}^{N} I_k \Big). \tag{5.38}$$

Because products of single Pauli operators are themselves Pauli operators op to a constant $\pm i$ (products of I are trivial and $XY = -YX = iZ$, $XZ = -ZX = -iY$ and $YZ = -ZY = iX$), the products of N-qubit Pauli operators are also N-qubit Pauli operators. This means that the second quantization Hamiltonian (Equation 5.29), after the Jordan-Wigner transform, can be written as

$$\mathcal{H}_{JW} = \sum_k c_k \mathcal{P}_k \tag{5.39}$$

Where the constants c_k are partly due to the $\mathcal{P}_k$'s and party due to $\langle i|h|j\rangle$ or $\langle ij|kl\rangle$. How this Hamiltonian can be measured will be discussed in a later section.

There are other transforms that use different representations to gain computational advantages. The most prominent alternative to the Jordan-Wigner transform is the Bravyi-Kitaev transform. It is more complex and harder to implement but shown to be superior to Jordan-Wigner in terms of required quantum gates for implementation [25, 26]. It was not considered in this project because of the difficulty to implement and because the difference between both methods scales with the system size. It is not that significant for systems of the size used in this report.

Chapter 6

Implementing the VQE Algorithm

Before going in depth again it may be good to reiterate what exactly is going on. The problem of finding the total ground state energy of an N-electron, M-nuclei system has been reduced to finding the smallest eigenvalue of the second quantization of the electronic Hamiltonian $\mathcal{H}$ (Equation 5.29), or equivalently the qubit Hamiltonian $\mathcal{H}_{JW}$ (Equation 5.39). Finding this ground state is what the VQE algorithm is used for. After finding the groundstate $|\Phi_0\rangle$ the total energy will be calculated using:

$$E = V_{NR} + \langle \Phi_0 | \mathcal{H} | \Phi_0 \rangle \tag{6.1}$$

here V_{NR} is the nuclear repulsion term that was dropped all the way back in section 3.1, using the Born-Oppenheimer approximation.

In this section the details of how VQE was used in the project are explained and alternative methods are given. Specifically this section covers the necessary knowledge about the CQT VQE MATLAB library, as first created by Robert de Keijzer [27, 28], based on [8].

6.1 Hamiltonian Generation

The first step in the algorithm is calculating the one- and two-electron integrals. Because this is a simulation however, the full Hamiltonian will have to be calculated to determine state energies in the VQE algorithm. The first part is done in Python. OpenFermion and PSI4 are used to first define a geometry of the molecule that is being analysed and then to calculate the molecular orbitals. The geometry includes the atoms of the molecule, their location in space, some other molecular properties and the basis set used for the spin orbitals. As mentioned before, unless mentioned otherwise the geometry used is one of H_2O, with a bond angle of 100° and a bond length varied around 1Å. The chosen basis orbitals are, from now on, STO-3G atomic orbitals. For H_2O this means there are 7 atomic orbitals: H1 1s, H2 1s and O 1s,2s,2px,2py,2pz. The molecular orbitals are then determined using the Hartree-Fock solution of the system. They are called in order of increasing energy, $1a_1, 2a_1, 1b_1, 3a_1, 1b_2, 4a_1$ and $2b_1$ [11]. An illustration of the molecular orbitals of H_2O can be seen in Figure 6.1 There will be the same amount of molecular orbitals as there were atomic ones, and each molecular orbital will translate to 2 spin orbitals and thus 2 qubits.

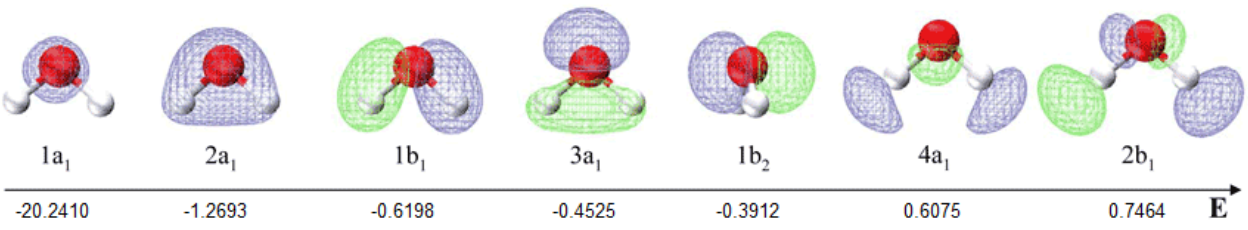


Figure 6.1: Image showing the seven molecular orbitals of H_2O that STO-3G HF creates, sorted by increasing energy left to right. Figure from [29] (edited), energies from [11].

Once this is done the electron integrals can be calculated, recall that these are all that is needed to create the qubit Hamiltonian. This can be done using any subset of the molecular orbitals, where orbitals can be ignored, left over to VQE or fixed as occupied. This will be explained in more detail in section 7.1. The output of the Python script is 4 values and 2 matrices for each geometry. The values are the HF energy, the FCI energy, the nuclear repulsion and the core coefficient. The first two are self explanatory, while the last two are constants that have to be added to the total energy. The nuclear repulsion is the constant term originally removed between Equation 3.1 and Equation 3.2, while the core coefficient is a correction term to the nuclear repulsion due to forcing orbitals to be occupied (see section 7.1). The two matrices are the NxN matrix with all one-electron integrals and the NxNxNxN matrix of two-electron integrals. By using these matrices together the 2^Nx2^N qubit Hamiltonian can be directly created by summing over all terms in Equation 5.29 and replacing the operators with their 2^Nx2^N Jordan-Wigner matrices. This final part is done in MATLAB.

6.1.1 Projecting the Hamiltonian

Next a trick is used to reduce the dimensionality of the problem further. This is nessecary as the 2^{14} by 2^{14} Hamiltonian and 2^{14} dimensional qubit statespace of the full H_2O system are too big to do calculations with. For this issue there was something present in the library, which only needed some small bug fixing and optimization to work for the water molecule. This solution is projecting the qubit statespace onto a space where all non-physical basisstates are removed. This is done with two projections. The first one is a projection onto the space with the correct number of electrons. It is given by:

$$\mathscr{P}_e^{m,N} = \prod_{\substack{j\neq N \\ j=0,\cdots,m}} \frac{N_{\text{state}} - j}{N - j}. \tag{6.2}$$

where N_{state} is the operator returning the state itself multiplied by the number of electrons in it. It is given by:

$$N_{\text{state}} = \sum_j a_j^\dagger a_j. \tag{6.3}$$

The form used is the Jordan-Wigner transformed version of this projection. The same holds for the second projection, which is based on spin. The states with the wrong spin number can also be removed. The projection for this is given by:

$$\mathscr{P}_e^{m,S} = \prod_{\substack{j\neq S \\ j=0,\cdots,m}} \frac{|N_{\text{state}\uparrow} - N_{\text{state}\downarrow}| - j}{S - j}. \tag{6.4}$$

Analogously to the previous operator the $N_{\text{state}\uparrow}$ and $N_{\text{state}\downarrow}$ operators are the operators that return the state multiplied by the amount of spin up and spin down respectively. They are defined as:

$$N_{\text{state}\uparrow} = \sum_j a_j^\dagger a_j \delta_{S_j^\uparrow}, \quad N_{\text{state}\downarrow} = \sum_j a_j^\dagger a_j \delta_{S_j^\downarrow} \tag{6.5}$$

Both operators are unitary and will change a lot of the Hamiltonian's columns and rows into all zeros. These trivial rows and columns can now be removed, without changing the eigenvalues of the Hamiltonian. Simple tests confirmed that the minimal eigenvalue was the same for the full 14 qubit Hamiltonian H and the reduced 9 qubit Hamiltonian Q, which is the trimmed version of the 14 qubit projected Hamiltonian $\mathscr{P}_e^{m,N}\mathscr{P}_e^{m,S}H\mathscr{P}_e^{m,S\dagger}\mathscr{P}_e^{m,N\dagger}$. Note that to make Q into an 9 qubit Hamiltonian trivial row/columns are still needed. This is because this scheme actually reduces the dimension to some value between 2^8 and 2^9 and it has to be filled up to 2^9 to fully describe the 9 qubit statespace.

From this point it is important to be clear about the size of the systems being considered. While the full H_2O Hamiltonian is a 14 qubit Hamiltonian, in this project it was only ever used in computation in its 9 qubit form. Similarly the $12, 10$ and 8 dimensional Hamiltonians that result from ignoring $2, 4$ and 6 spin orbitals respectively ($1, 2$ and 3 spatial orbitals), are only used in simulations as $8, 7$ and 6 qubit systems. Because it would be misleading to say simulations were done on for example a 12 qubit system, while it was 8 qubits big, calling it a 12 qubit system would not work. However if it is called the 8 qubit system it is unclear that a reduced 12 qubit system is being simulated instead of the not reduced 8 qubit one. What will be done is specifically referring to the amount of orbitals (with the not reduced numbers) or to the amount of qubits (using the returned number). This way it is always clear what system is being talked about (e.g. the 12 orbital system is the same as the 8 qubit system). In some cases when talking about the full Hamiltonian 12 and 14 qubit Hamiltonian is used to denote the 12 and 14 orbital Hamtiltonians before reduction.

6.2 State Preparation

There are two ways results are gotten in this report. The first is by taking the minimal eigenvalue of the qubit Hamiltonian. This is the exact solution within the space and, indeed, it is the same as the FCI energy for the full qubit Hamiltonian. The second way is by actually simulating VQE, and in this case the states the classical optimizer wants sampled have to be created and evaluated. The creation or preparation of these states starts with an initial state. In this report the initial state is the Hartree-Fock ground state $|\phi_0\rangle$. Because the Hartree-Fock molecular orbitals are chosen as basis set this is possible, however picking the state where the N lowest energy orbitals out of the $2K$ molecular orbitals are occupied as a starting state is possible in general. This is done in literature [9], but other initial states are used often as well such as the vacuum state $|\ \rangle$ (corresponding to qubit state $|00\ldots00\rangle$) [8, 10, 28], or the Unitary Coupled Cluster ground state (coupled cluster is a common post-HF method) [11, 30]. Then a matrix of rotation parameters t is used to rotate this initial state to the state that will be sampled. At the first step some t_0 has to be chosen before the classical optimizer starts changing it around. During exploratory testing a few combinations of initial states and initial rotations were quickly looked at. UCC was not considered as one of the visions in the project was to go from HF directly to FCI with VQE. The initial states considered were the vacuum state, the HF ground state and the inverse of the HF ground state with respect to t_0 (i.e. the state that gets rotated to the HF ground state with t_0). In the end the HF ground state was chosen as initial state and the zero rotation as t_0. The choice of using the zero rotation as starting point has also been done before [30].

There are now many methods to rotate the initial state towards an desired sample state. All methods have in common that entanglement and rotations are needed. This can be done by rotating and entangling qubits in different pairs at different depths of the circuit [9] or by alternating a single qubit rotation and a single two-qubit entanglement gate per level of the circuit [10]. Most common however is the method used by Kandala et al. [8]. Here N triples of single qubit rotations (three times one on each qubit) are alternated with a single, often N-qubit, entanglement operation. This method will be discussed in a bit more detail now. First note that any single qubit rotation (excluding phase) can be written as

$$R(\alpha,\beta,\gamma) = Z_\alpha X_\beta Z_\gamma \tag{6.6}$$

Here Z_α and X_β are shorthand notation for the rotation matrices $R_Z(\alpha)$ and $R_X(\beta)$. These matrices rotate around the Z and X axis respectively in the Bloch sphere representation. In general for a rotation of θ around one of the three axes the rotational matrix is given by [13]:

$$R_X(\theta) = X_\theta = \begin{pmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{6.7}$$

$$R_Y(\theta) = Y_\theta = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix} \tag{6.8}$$

$$R_Z(\theta) = Z_\theta = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \tag{6.9}$$

Now define the entanglement depth d as the amount of entangling operations used. Write $|q_0\rangle$ for the starting state (initial state, ansatz state) and U_{ent} for the entangler. The entangler will be taken as a constant but does not have to be in general. Now the parameter matrix $\vec{\theta}$ is introduced: its elements $\vec{\theta}(a,b)$ are the rotation angle at depth a of qubit b. Now at each rotation the rotation of all N qubits can be written as

$$Z_{\vec{\theta}_i} = \bigotimes_{j=1}^{N} Z_{\vec{\theta}(i,j)} \quad , \quad X_{\vec{\theta}_i} = \bigotimes_{j=1}^{N} X_{\vec{\theta}(i,j)} \tag{6.10}$$

Where $\vec{\theta}_i$ is row i of $\vec{\theta}$ and $\vec{\theta}(i,j)$ is the i,j element of $\vec{\theta}$. The complete rotation at depth i can be written as:

$$U_i(\vec{\theta}) = Z_{\vec{\theta}_{3i+3}} X_{\vec{\theta}_{3i+2}} Z_{\vec{\theta}_{3i+1}} \tag{6.11}$$

The coefficient vector of the final state $|q(\vec{\theta})\rangle$ is given by

$$|q(\vec{\theta})\rangle = U_d(\vec{\theta}) \times U_{ent} \times U_{d-1}(\vec{\theta}) \times U_{ent} \times \ldots U_{ent} \times U_1(\vec{\theta}) \times U_{ent} \times U_0(\vec{\theta})|\phi_0\rangle = \mathbf{U}_d(\vec{\theta})|\phi_0\rangle \tag{6.12}$$

The dimension of the matrix $\vec{\theta}$ would therefore be $3d+3$, however if the initial state used is a basisstate the first Z rotation does nothing since all qubits are on the z-axis. Then t is $3d+2$ dimensional. The only difference is now that in Equation 6.11 the indices of ZXZ switch to $\vec{\theta}_{3i+2}, \vec{\theta}_{3i+1}$ and $\vec{\theta}_{3i+0}$. Where the last term disappears

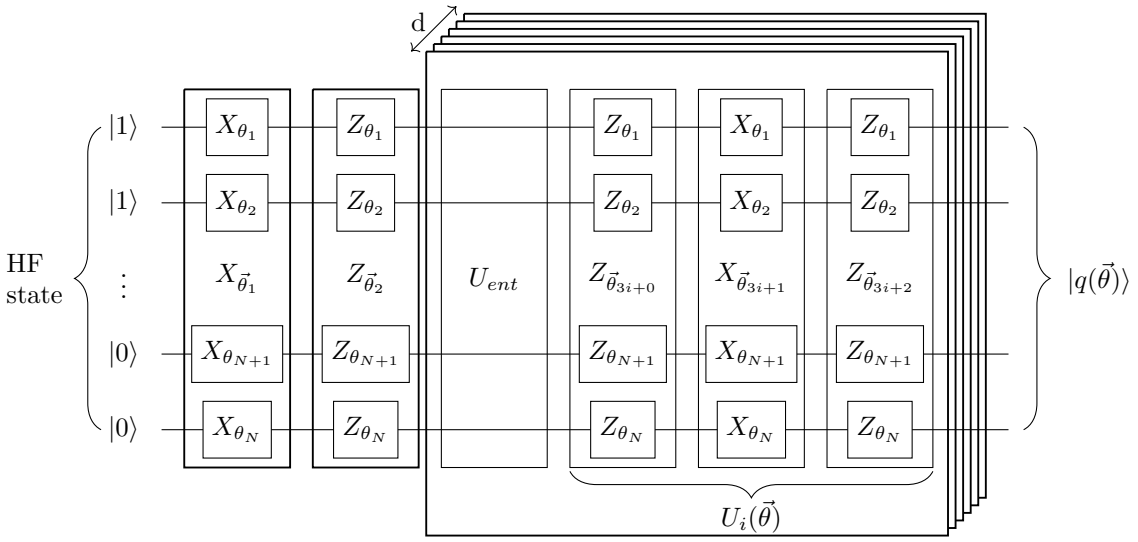


Figure 6.2: Schematic illustration of the state preparation circuit $\mathbf{U}_d(\vec{\theta})$. Based on an image from [8].

for the first rotation as there is no 0th column in $\vec{\theta}$. Specifically this is the case for the Hartree-Fock state that was used as ansatz state in this project. The full state preparation circuit (or the circuit of operator $\mathbf{U}_d(\vec{\theta})$) is shown in Figure 6.2.

Before moving on the entangler U_{ent} has to be discussed in some more detail. Its main function is to mix the states together, as each triplet of rotations can reach any point on the specific Bloch sphere of that qubit. The relations between qubits are harder to achieve. Determining which entanglers work well and how to determine how well specific entanglers mix the states in a mathematical sense are topic of interest in the field [28]. In this report the CNOT chain was used as entangler. This can be considered the most basic option as the CNOT gate is the text book example of an entangling gate:

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{6.13}$$

it flips the second qubit if the first qubit is in state $|1\rangle$. The CNOT chain links the qubits together pairwise, using $N-1$ gates to link N qubits together as shown in Figure 6.3 below. The entanglement depth d will be discussed in for each simulation in the results section.

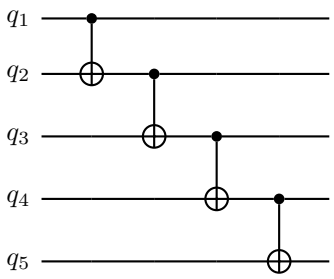


Figure 6.3: quantum circuit diagram of a CNOT chain on 5 qubits

6.3 Measurement

With the state prepared it now has to be measured. While it is possible to simulate the way VQE measures the qubit system and calculates the energy from the measurements, it only adds running time when testing systems without errors, such as in this project. Therefore instead any VQE simulation in this report is done by building

a qubit state $|q(\vec{\theta})\rangle$ from the parameters and then calculating the energy E_q exactly using $E_q = \langle q|H|q\rangle$, where H is of course the qubit Hamiltonian, which is known in the simulation. This way the classical optimization algorithm gets the same response as it would ideally get from the quantum chip after many averaging steps and it is possible to analyse convergence, running time, accuracy and other interesting properties. Because in any real application one will have to do measurements, measurement is still discussed here.

Measuring the Pauli matrices of individual qubits is straight forward. If one wants to measure $\langle I\rangle$ one can simply take the eigenvalue 1 and skip measurement all together. In the case of $\langle Z\rangle$ on can measure the qubit to be either $|0\rangle$ or $|1\rangle$ corresponding to eigenvalues -1 and 1 respectively. This is because measurement is per definition of the qubit system in the Z-direction. To measure $\langle X\rangle$ or $\langle Y\rangle$ of a qubit, one has to rotate the qubit such that the direction of interest is now in the Z-direction. This means that for $\langle X\rangle$ one needs to use $R_Y(\frac{\theta}{2})$ and for $\langle Y\rangle$ $R_X(\frac{\theta}{2})$. These rotations are called post rotations. An updated schematic of the quantum circuit, including post rotations and measurements is shown in Figure 6.4.

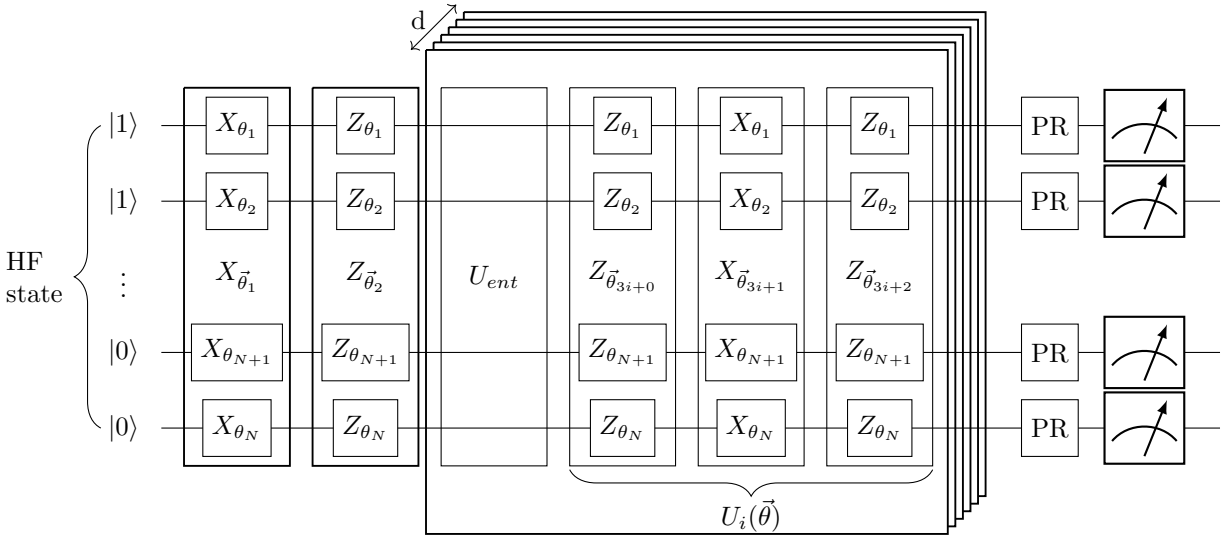


Figure 6.4: Schematic illustration of the full quantum circuit, including state preparation, post rotation and measurement. The post rotation gate can be either $I, R_X(\frac{\theta}{2})$ or $R_Y(\frac{\theta}{2})$ Based on an image from [8].

One might think that when multiple qubits are considered each qubit can be measured independently, such that only the separate $\langle X\rangle$, $\langle Y\rangle$ and $\langle Z\rangle$ expectation values have to be calculated for each qubit independently. This is only the case for two qubits as for bigger systems the entanglement and correlation between the qubits means that for example $\langle Z_1Z_2Z_3\rangle \neq \langle Z_1\rangle\langle Z_2\rangle\langle Z_3\rangle$ in general [31]. This means that each of the N-qubit Pauli operators from Equation 5.39 has to be measured. In fact many measurements have to be done to approach any of the averages $\langle \mathcal{P}_k\rangle$.

Since measurement is completely ignored in this report (in the sense that the qubit vector q is not post rotated and then measured, but instead directly evaluated by calculating $\langle q|\mathcal{H}|q\rangle$), a few tricks related to measurement from the literature will be mentioned here only for completeness. The first trick involves reducing the amount of qubits needed in the system by using symmetries in the Jordan-Wigner transformed Hamiltonian. The Hamiltonian is projected in such a way that for some qubits the Pauli operators that have to be measured can be replaced by their eigenvalues directly. This allows one to not evaluate certain qubits, allowing one to remove them from the system altogether [28, 32, 33]. Additionally there are schemes where the expected value of certain N-qubit Pauli operators (or parts of them) can be determined from others in a smart way, reducing the amount of measurements that have to be done. This is called quantum tomography. Pauli operators are grouped in sets that are evaluated together, sometimes using entangling, resulting in a significant decrease of terms that have to be measured [34, 35, 36]. Additionally many measurements are required for each Pauli operator. For some trial functions this can be sped up with methods that sacrifice system size (i.e. make the circuit bigger) to reduce the amount of measurements required to get an accurate estimated expectation value [37]. Lastly one can smartly pick which Pauli operators to evaluate and which not to, by calculating their maximum expected contribution and removing the C lowest contributing Paul operators entirely to get a decent approximation [35].

6.4 Classical Optimizers

Multiple optimizers were considered and attempted in this project. Many quickly showed to not be fit for the problem and were thus discarded after little experimenting. Only the optimizers of some interest (because others have used them for example) are discussed here. Two of them are local optimizers (SPSA and Nelder-Mead) and one is a global optimizer (fmincon). Local minimizers turned out to be incompatible with the system size very quickly, often even taking a long time to find a local minimum when starting very close to it. Multistart type global optimizers (running many local optimizers and picking the minimum of the minima) were therefore not considered at all, but they could have potential. The global optimizer implemented in the VQE library (DIRECT) was very slow for bigger systems. It had the big disadvantage that due to the small amount of very long iteration steps, it was much harder to tweak the iteration steps for the method to be of reasonable time.

6.4.1 SPSA

Simultaneous pertubation stochastic approach (SPSA) is a stochastic gradient decent algorithm [38, 39]. Its main feature is that it samples only 2 points at every iteration step to estimate the gradient. Then a step is taken in the opposite direction just like a regular gradient decent algorithm.

SPSA uses a standard iterative step namely

$$t_{k+1} = t_k - a_k g(t_k) \tag{6.14}$$

where t_k is the rotational matrix at iteration step k, the a_k's a scalar series converging to 0. $g(t_k)$ is the gradient of $E(t) = E(q(t))$ at t_k. In more regular gradient decent algorithms it would be calculated by looking at the partial derivative in the direction of every basis vector, but SPSA instead uses a random perturbation over all dimensions Δt_k instead: $\hat{t}_{k,\pm} = t_k \pm c_k \Delta t_k$. In the VQE library the distribution used is the Rademacher distribution: every matrix element has a probability of $\frac{1}{2}$ to be either -1 or 1 [27]. The gradient is estimated by:

$$g(t_k) = \frac{E(\hat{t}_{k,+}) - E(\hat{t}_{k,-})}{2c_k} \Delta t_k. \tag{6.15}$$

The factors a_k and c_k are given by:

$$a_k = \frac{a}{k^A}, \tag{6.16}$$

$$c_k = \frac{c}{k^\Gamma}. \tag{6.17}$$

The constants c, A and Γ are set to $0.01, 0.602$ and 0.101 respectively in the library [27]. These are optimized values according to literature [40]. The constant a is a bit of a special case and has to be calibrated using the gradient around the initial point. It is given by:

$$a = \frac{2\pi}{5} \frac{c}{\langle E(\hat{t}_{1,+}) - E(\hat{t}_{1,-}) \rangle_{\Delta t}} \tag{6.18}$$

where the bottom term is the expectation value of the gradient over the distribution of Δt.

SPSA has been successfully used on small VQE systems [8, 28]. While good results were replicated on small systems, SPSA did not perform well on bigger systems, getting stuck in local minima very often. More details about the SPSA results can be found in section 9.3.

6.4.2 Nelder-Mead

The Nelder-Mead simplex method [41], uses $n+1$ function values $f(x_1) \leq f(x_2) \leq \ldots \leq f(x_{n+1})$ to minimize an n-dimensional function $f(\mathbf{x}), \mathbf{x} \in \mathbb{R}^n$. These function values form the simplex and the algorithm stops based on some criteria on their values and returns $f(x_1)$. The original criteria was the sample standard deviation over the current simplex. This leads to bad results for very flat functions as a very large simplex can have low standard deviation in this case. Therefore additional requirements on the size of the simplex can be used as well. Until termination the algorithm keeps improving its worst point (highest function value) and in this sense it does not go towards a minimum but rather away from maxima. The steps to improve on $f(x_{n+1})$ are not very strict and differences exist between implementations. The one implemented in MATLAB as *fminsearch* is the method from [42]:

1. Order the $n+1$ vertices to satisfy $f(x_1) \leq f(x_2) \leq \ldots \leq f(x_{n+1})$.
2. Calculate the centroid $\bar{x}$ of the points better than x_{n+1}: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
3. Attempt to improve with reflection: $x_r = \bar{x} + \rho(\bar{x} - x_{n+1})$. Here $\rho > 0$ is a parameter. If this point is better than x_n but not x_1 replace $f(x_{n+1})$ with $f(x_r)$ and return to step 1. If it is better than x_1 go to step 4. If it is worse than x_n go to step 5.
4. Since $f(x_r) < f(x_1)$ a possible improvement can be found by expanding in the same direction further: $x_e = \bar{x} + \chi(x_r - \bar{x})$. Where $\chi > 1, \chi > \rho$ is a parameter. Now replace $f(x_{n+1})$ with the smaller one out of $f(x_r)$ and $f(x_e)$ and return to step 1.
5. Since x_r would still be the worst point in the simplex it is attempted to improve it by going closer to the center of the simplex. Take the best value x^* out of $f(x_r)$ and $f(x_{n+1})$ and contract it towards the center: $x_c = \bar{x} + \gamma(x^* - \bar{x})$. Here $0 < \gamma < 1$ is a parameter. If x_c is better than x^* replace $f(x_{n+1})$ with $f(x_c)$ and return to step 1. If not continue to step 6.
6. Since improving $f(x_{n+1})$ has failed at this stage the entire simplex is improved instead. At this point all x_i except x_1 are improved by shrinking the simplex: $x_i = x_1 + \sigma(x_i - x_1)$. Where $0 < \sigma < 1$ is a parameter. After this return to step 1.

The MATLAB implementation uses the standard values for the parameters $\rho = 1, \quad \chi = 2, \quad \gamma = \frac{1}{2}, \quad \sigma = \frac{1}{2}$.

The method was successfully used in VQE before on a 6 qubit H_2O system [10]. It is also recommended for VQE in literature [35, 43]. However it has been shown that for large dimensions the Nelder-Mead method "deteriorates" [44]. It makes sense conceptually since the simplex changes less per iteration step due to its dimensionality. This "deterioration" is exactly what happened in the few simulations attempted with the method. Even for the smaller systems considered it required an obscene amount of iteration steps to converge since at times when $f(x_1)$ was improved on it was only an improvement of order 10^{-7} Hartree. Recall that it can take many steps to improve $f(x_1)$ even once. After almost a day it had not converged on a system that the optimizer below could be done with within 3 hours, this can be seen in Figure 6.5. Because of this the method will not be discussed further in this report.

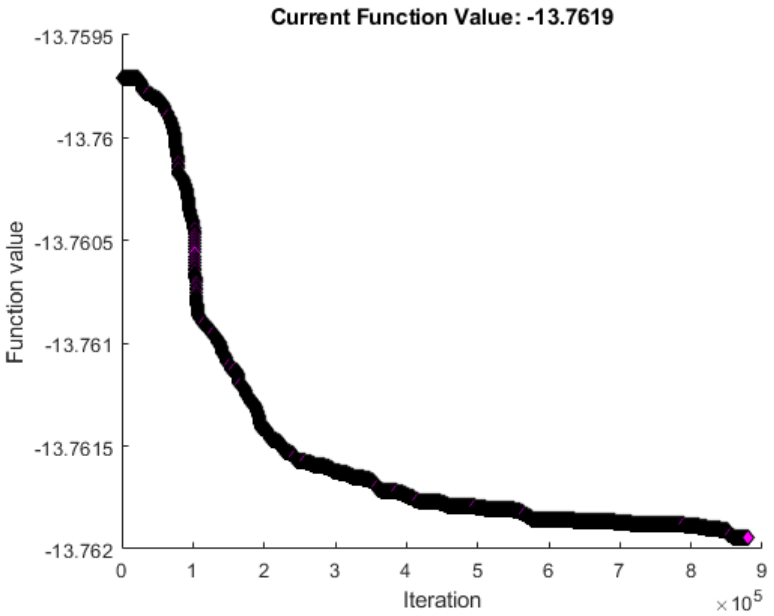


Figure 6.5: Example Nalder Mead simulation. System size of 10 orbitals. Exact solution of this problem is -13.8119. At almost one million iteration steps an improvement of less than 0.0025 was made. The exact solution was still more than 0.0500 away at the time of termination.

6.4.3 fmincon

The fmincon algorithm is MATLAB's non-linear constrained multivariable solver. When looking which solvers did well on the problem, it was chosen to be tested and performed the best, even though the problem is not constrained. There were not many alternatives (one being Nelder-Mead and the other BFGS Quasi-Newton). The first was discussed in the previous section and the BFSG Quasi-Newton did not perform well, most likely because it is not well suited for high dimensions.

It is not clear how the method would work in the first place since it is a barrier method (which requires constraints to set up), while it was used in this project with empty constraint variables. Presumably some very loose constraints are applied automatically, but MATLAB obscures the code of many functions including the code that is relevant here. Regardless a very brief explanation of the constrained version of the algorithm is given. Technical details are omitted, but can be found in [45, 46, 47]. Consider the following minimization problem:

$$\min_x f(x), \quad \text{s.t.} \quad h(x) = 0, \quad g(x) \leq 0 \tag{6.19}$$

A barrier function is used to define approximate problems. Here $\mu > 0$ is the parameter. Letting μ go to zero lets the approximate function go to $f(x)$.

$$\min_{x,s} f(x) - \mu \sum_i \ln s_i, \quad \text{s.t.} \quad s \geq 0, \quad h(x) = 0, \quad g(x) + s = 0 \tag{6.20}$$

These problems do not have to be solved exactly to get convergence and therefore this method is an approximate global optimizer. Solving for these $f_\mu(x, s)$ is done with iterative steps in x and s, reducing some merit function. First an direct step is attempted, using the Hessian of the Lagrangian of $f_\mu(x, s)$ (see [47]). In specific cases where this does not work a trust region method is used, which is guaranteed to work in those situations (see [45, 46, 47]).

Although not well understood and used as a black box, fmincon performed best out of all attempted optimizers and is therefore used in almost all of the results of this project. Why it works better than the alternatives is not clear, but presumably the combination of direct steps and trust region methods is very resilient against hard/high dimensional Hessians compared to the other methods that use the Hessian, such as BFSG Quasi-Newton.

Chapter 7

Reducing Running Times

In this chapter techniques are discussed that allow one to reduce the H_2O system to be simulatable. At the start of the project the MATLAB library could not simulate a full 14 orbital H_2O system, first because the 14 qubit Hamiltonian took to long to generate and after that was fixed, because the 14 qubit Hamiltonian was too big to be workable. A few tricks were attempted to reduce the amount of qubits without too much approximation. These methods are described in this section. The projection method mentioned in subsection 6.1.1 is one of these tricks, but will not be discussed again. In the next section the orbitals/qubit convention discussed before will be dropped for clarity. Since only eigenvalues are considered and they are preserved by the reduction, system size will be in qubits instead of orbitals. This is because of confusion that may arise due to both spatial and spin orbitals being referred to as "orbitals" in the section otherwise. After this section system size will be once again referred to with spin orbitals and reduced qubit numbers.

7.1 Removing Orbitals/Adding electrons to the Core

A method used by the group before and implemented in the library was one where after getting the molecular spatial orbitals they would be removed based on their energy. Negative energy orbitals would be fixed and added to the core as a parameter, while positive energy orbitals would be ignored completely. The orbitals with highest absolute value would be removed from the simulation first as they are most likely to be either occupied in case of a negative energy or unoccupied in case of a positive energy. For each orbital removed two qubits are removed from the simulation. The STO-3G H_2O molecular orbitals considered here will be sorted based on their energy as earlier, with $1a_1$ being the lowest energy and $2b_1$ the highest (see Figure 6.1 for a reminder of the ordering). Orbitals $1a_1$ through $1b_2$ have negative energy and $4a_1$ and $2b_1$ have positive energy. Orbitals added to the core will be called occupied orbitals, while orbitals that are considered are called active. Orbitals that are ignored will be called inactive. Removing any orbital will (obviously) change the electron integrals. In the case of inactive orbitals the result is trivial, as their indices are just removed from the sums. Occupied orbitals are not trivial. Let orbital a of atom k be occupied by 2 electrons. The idea is to now approximate the nucleus K as a point particle with charge $Z_K - 2$. The electronic Hamiltonian gains an additional term:

$$\hat{\mathcal{H}}_2 = \hat{V}_{CC} + \hat{\mathcal{H}} = \hat{V}_{CC} + \sum_{ij} \langle i|\hat{h}|j\rangle a_i^\dagger a_j + \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle a_i^\dagger a_j^\dagger a_l a_k. \tag{7.1}$$

This term will be called the core coefficient and is the (constant) correction of the nuclear repulsion combined with the energies of the fixed electrons. The latter are determined using Koopmans' theorem and are given by $-\epsilon_{a1}$ and $-\epsilon_{a2}$. $\hat{V}_{CC}$ is given by:

$$\hat{V}_{CC} = \sum_{I}^{M}\left[\frac{Z_I(Z_K-2)}{R_{IK}} - \frac{Z_I Z_K}{R_I K}\right] - \epsilon_{a1} - \epsilon_{a2} \tag{7.2}$$

Additionally the second quantization Hamiltonian has changed too, as h is influenced by the change. The operator h_i is now given by:

$$\hat{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_{I=1}^{M}\frac{Z_I - 2\delta_{IK}}{R_{iI}} \tag{7.3}$$

and the new one electron integrals are given by:

$$\langle i|\hat{h}|j\rangle = \int dx_1 \chi_i^*(x_1)\hat{h}_1\chi_j(x_1). \tag{7.4}$$

Note that just as before this is still the electronic Hamiltonian and the Nuclear Repulsion term has to be added to the energy gotten from this Hamiltonian to get the actual total energy of the system. The total energy will be written as three seperate terms:

$$E = V_{NR} + \hat{V}_{CC} + \langle\Phi_0|\hat{\mathcal{H}}|\Phi_0\rangle \tag{7.5}$$

Of these terms only the last one is determined by VQE. The other two are precalculated constants. If no occupied orbitals are considered $\hat{V}_{CC}$ is zero and $\hat{\mathcal{H}} = \mathcal{H}$ and the system returns to the regular form.

Tests were done to see how badly these orbital removals affected the simulation results. Both LiH and H_2O were tested with 1 up to all molecular orbitals active. Instead of using a VQE simulation the exact solution was calculated by taking the smallest eigenvalue of the resulting Hamiltonian directly. The results for water are shown in Figure 7.1. The other plots (LiH and absolute errors) can be found in Appendix B. Some of the differences between geometries could be explained by the fact that the absolute values of the orbital energies shift, while the orbital removal order was kept constant for all geometries for each system size. When a certain orbital that is important in one geometry and not so much in another is removed sooner or later it can change the outcome in favor of one of the geometries.

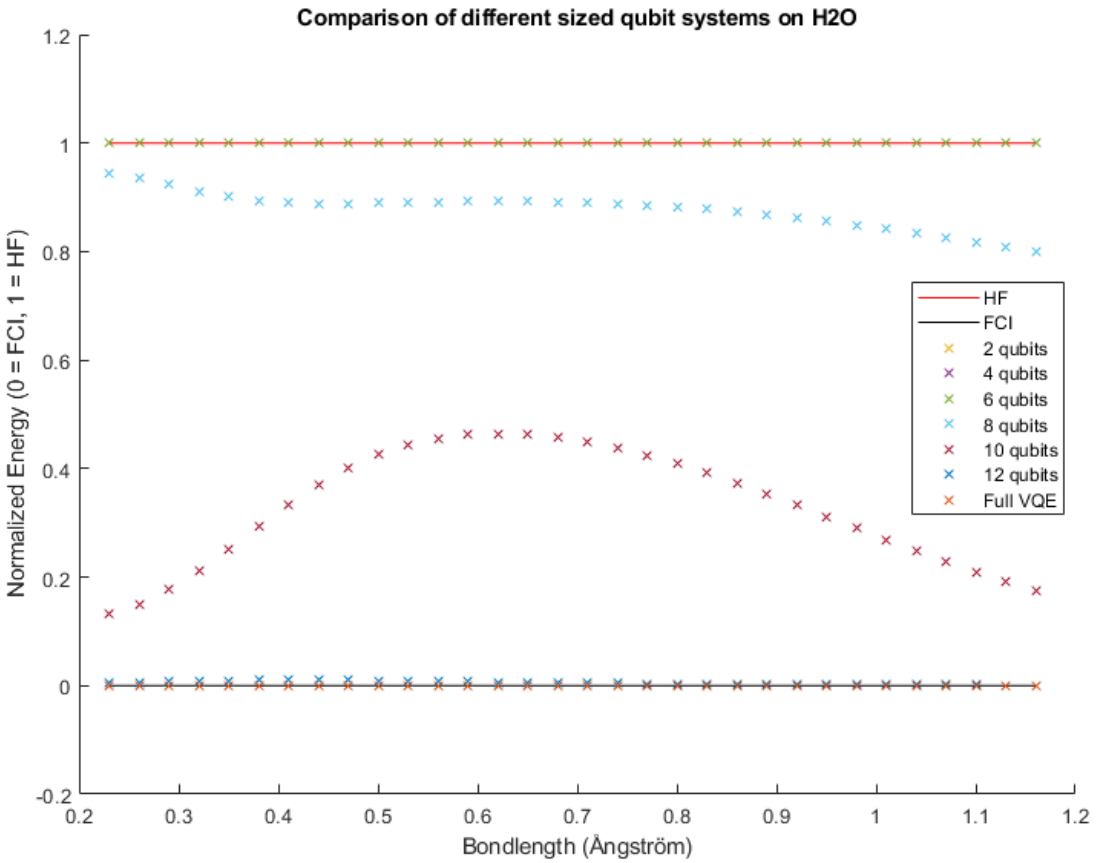


Figure 7.1: Plot showing the difference in simulation result based on the amount of orbitals used for H_2O. The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately. Note that the 2,4 and 6 qubit VQE solutions are on the HF line. Additionally note that while the 12 qubit solution is very close to the FCI solution it is in fact above it for all bond lengths while full (14 qubit) VQE is on the FCI line.

The results are what one might expect, each orbital removed brings the solution further from the FCI solution towards the HF solution. However, differences between geometries seemed fairly large (especially for 10 qubits). It is possible that for some of the geometries the result could actually be better if the generalization of using

the same removal order for each of them was removed. For H_2O this resulted in some results being better and some results being unexpectedly worse. Additionally while the energy seemed to be continuous in bond length when all geometries had the same removal order, this was not the case for the case where they had different ones (the graph with discontinuous energies due to different orbital selection can be found in B.4). Because of this exhaustive simulations were done on H_2O with all possible $\binom{7}{k}$ subsets of orbitals for $0 < k < 7$. In Figures 7.2 and 7.3 below, for 4 and 5 orbitals chosen respectively (8 and 10 qubit systems), a subfigure is shown for each of the seven orbitals (ordered left to right top to bottom by energy). With this one can see which orbitals perform well within the possible combinations and which perform worse. The figures for the other system sizes can be found in Appendix B (see Figures B.5 to B.8).

If one takes the time to identify the orbitals that perform well in each case, it is clear that the absolute value of the energy was not a very good indicator for the importance of the orbitals. For 10 qubits, for example, the best result is gotten by removing orbitals $1a_1$ and $1b_2$, not $1a_1$ and $2a_1$ or $1a_1$ and $2b_1$ as one may expect with the absolute value approach. Combining this with the issues of convergence for bigger qubit systems that occurred during the project, these results were used to select the best set of orbitals to use for each amount of qubits in attempts to get better approximate results with a not full system. Additionally the orbitals that perform well in general can be identified, this being the lowest and highest negative orbitals (numbers $1a_1$ and $1b_2$). These results are the same as results found in [11].

One may argue that picking the best molecular orbitals in this way goes against one of the core ideas of the project, where only things deducible with HF are looked at. While it is true that these results cannot be gotten without essentially solving FCI, this way only one of the orbital combinations would have to be simulated, instead of simulating all of them and picking the best one. In a real life application, where the behavior of the orbitals is not known, one would most likely have to try different combinations of orbitals and take the minimal value across all of them. Simulating everything without using this information to get the same minima was no good use of computation time.

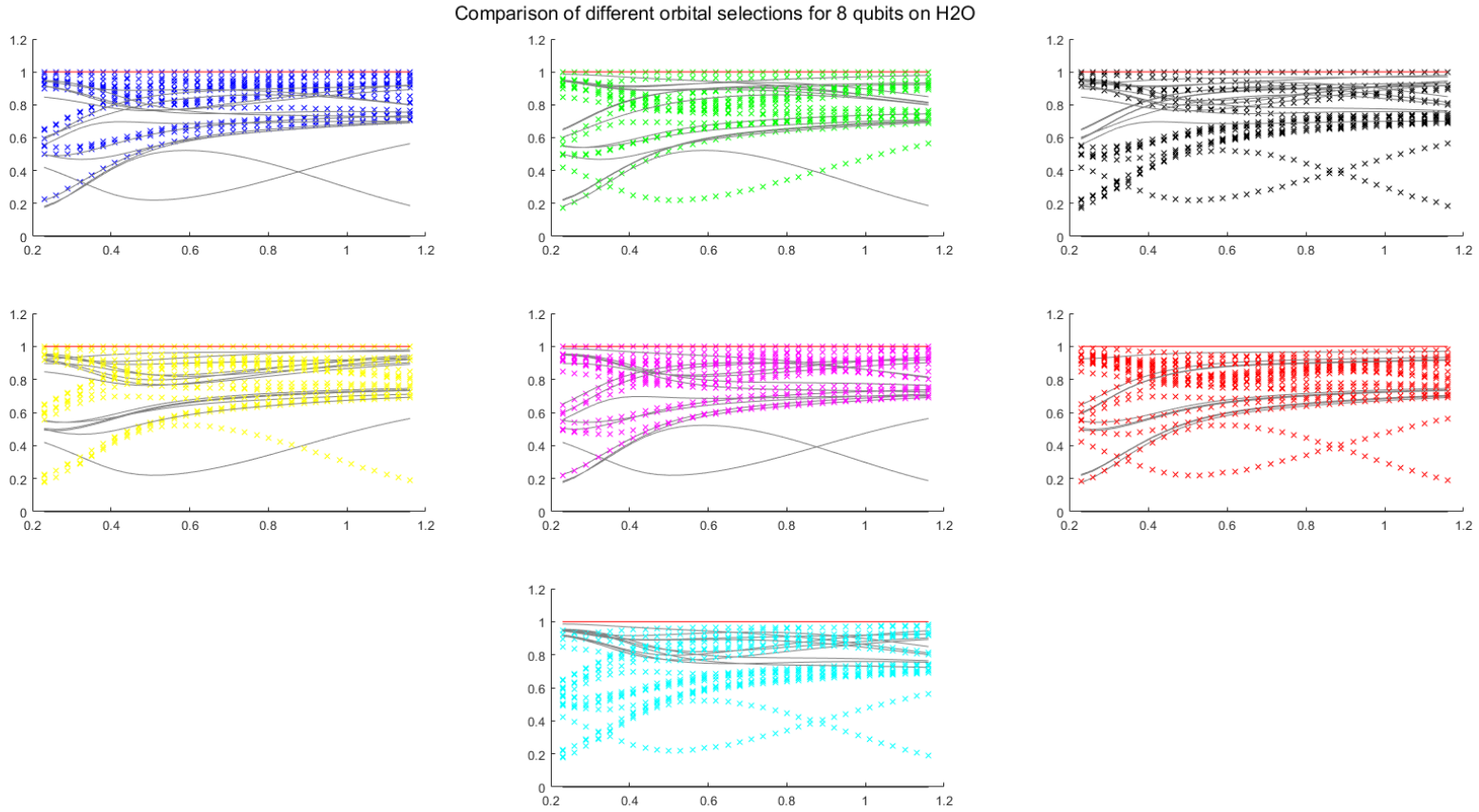


Figure 7.2: Plot showing for each combination of four out of the seven spatial orbitals the minimal eigenvalue of the Hamiltonian, whether made with a combination of orbitals that specific orbital is in (marked and colored) or with a combination of orbitals it is not in (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.

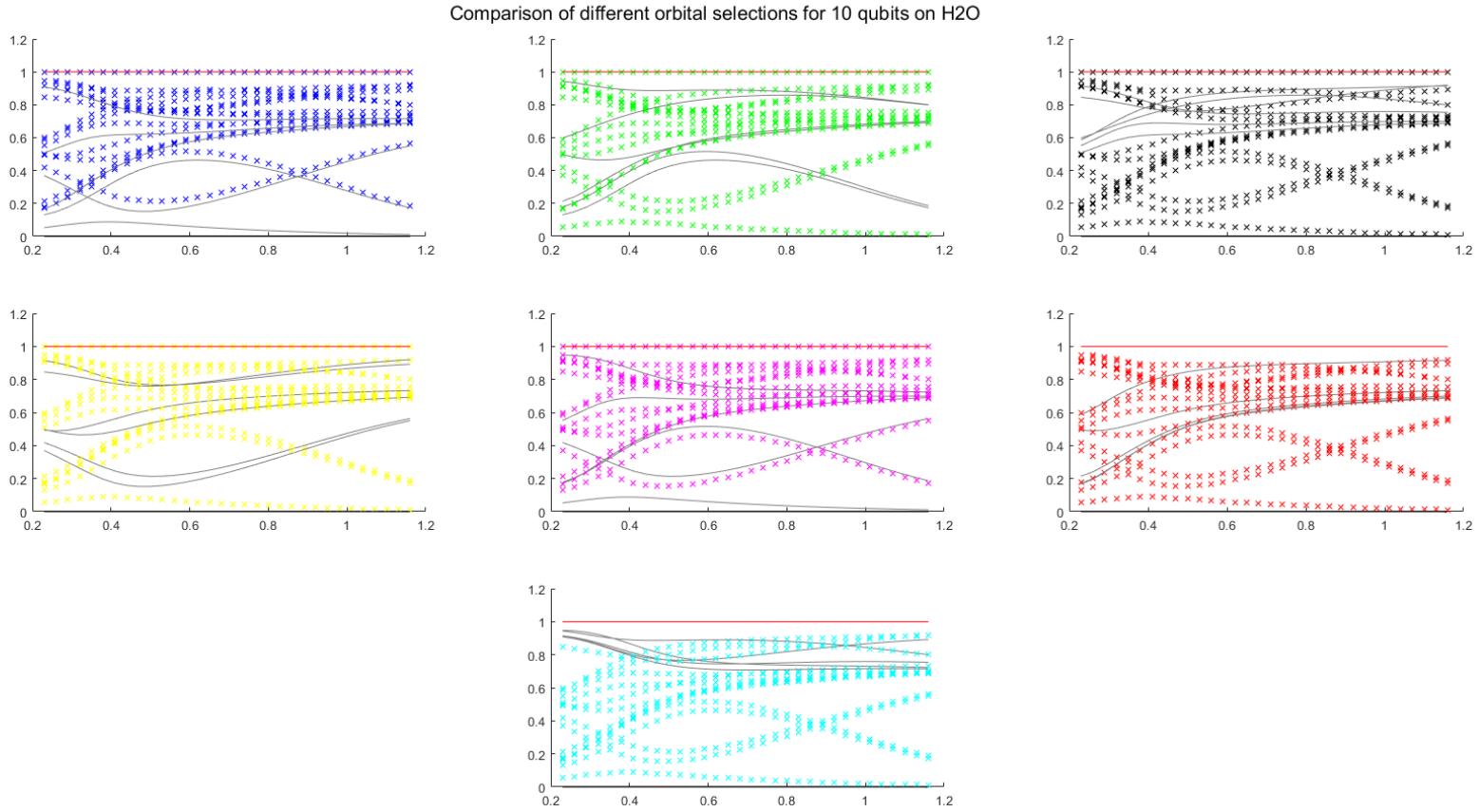


Figure 7.3: Plot showing for each combination of five out of the seven spatial orbitals the minimal eigenvalue of the Hamiltonian, whether made with a combination of orbitals that specific orbital is in (marked) or with a combination of orbitals it is not in (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.

7.2 Closed Shell Hamiltonian

Google's paper on VQE [9] suggests that they only simulate spin-up electrons explicitly, by looking at singlet states. This means every spatial orbital either has two or zero electrons in it. Indeed, this is the Closed shell approximation that was also used in the derivation of the Hartree-Fock method. However what is needed to use this in the VQE library is a second quantization type Closed Shell Hamiltonian. Specifically this means a Closed Shell Hamiltonian that can be transformed by the Jordan-Wigner transformation into a qubit Hamiltonian. No previous work on this was found, most likely because there are better methods to introduce the approximation. However those would require a significant overhaul of the VQE library, which was not a valid option for this project. Needed is a Hamiltonian $\hat{\mathcal{H}}$ in second quantization, such that $\langle\hat{\Phi}|\hat{\mathcal{H}}|\hat{\Phi}\rangle = \langle\Phi|\mathcal{H}|\Phi\rangle$ when $|\Phi\rangle$ is a Closed Shell state. Here $\mathcal{H}$ is the normal second quantization Hamiltonian and $|\hat{\Phi}\rangle$ is the singlet state $|\Phi\rangle$ with the spin-down electrons ignored. For example the state $|\Phi\rangle = |11001111\rangle$ would have $|\hat{\Phi}\rangle = |1011\rangle$. To start the operator that takes a state from closed shell to unrestricted is defined. Take closed shell state $|a_1 a_2 \dots a_N\rangle$ ($a_i \in \{0, 1\}$) now the operator $\mathcal{T}$ takes this to its corresponding electron occupation representation state:

$$\mathcal{T}|a_1 a_2 \dots a_N\rangle = |a_1 a_1 a_2 a_2 \dots a_N a_N\rangle \tag{7.6}$$

where, of course, the first state is some spatial orbital occupation form, while the second state is a normal occupation representation made out of spin orbitals with alternating spin and pairs of the same spatial orbitals as before. Now we need pair-wise creation and annihilation operators $\hat{a}_i^\dagger$ and $\hat{a}_i$. Their definitions follow quite simply from $\mathcal{T}$:

$$\hat{a}_i^\dagger = a_{i\uparrow}^\dagger a_{i\downarrow}^\dagger \mathcal{T} \tag{7.7}$$

$$\hat{a}_i = a_{i\downarrow} a_{i\uparrow} \mathcal{T} \tag{7.8}$$

Next the commutation relations are derived, and the result immediately shows that the idea of implementing this using the Jordan-Wigner transform is not possible. Both the commutator and anticommutator are calculated at the same time by sampling them against some state $|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle$ in normal occupation representation such that the normal operators can be used as $\mathcal{T}$ is applied to the state already:

$$(\hat{a}_i \hat{a}_j^\dagger \pm \hat{a}_j^\dagger \hat{a}_i)|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \tag{7.9}$$

There are now four cases: $i = j$ with i occupied or i unoccupied, $i \neq j$ with i occupied j not occupied and the remaining case is $i \neq j$ and either i unoccupied, j occupied or both. This last case is trivial as it is always zero since $\hat{a}_i$ will try to annihilate $\chi_{i\uparrow}$ and $\chi_{i\downarrow}$, which are not there, or $\hat{a}_j^\dagger$ will try to create $\chi_{j\uparrow}$ and $\chi_{j\downarrow}$, when they are already there. Consider first the case where $i = j$ and i is occupied

$$(\hat{a}_i \hat{a}_i^\dagger \pm \hat{a}_i^\dagger \hat{a}_i)|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle = 0 \pm \hat{a}_i^\dagger \hat{a}_i|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle = \pm|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \tag{7.10}$$

where the first term is zero due to $\hat{a}_i^\dagger$ trying to create $\chi_{i\uparrow}$ and $\chi_{i\downarrow}$. The sign does not change at the end because moving 2 orbitals to the front together (and back for that matter) is an even amount of switches and thus has no sign change due to the antisymmetry principle. Similarly an expression for i unoccupied can be derived:

$$(\hat{a}_i \hat{a}_i^\dagger \pm \hat{a}_i^\dagger \hat{a}_i)|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle = \hat{a}_i \hat{a}_i^\dagger|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \pm 0 = |\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \tag{7.11}$$

Lastly the case where $i \neq j$ and i occupied j unoccupied is considered. Again, since moving pairs gives no sign change due to antisymmetry it quickly follows that:

$$(\hat{a}_i \hat{a}_j^\dagger \pm \hat{a}_j^\dagger \hat{a}_i)|\dots \chi_{i\uparrow}\chi_{i\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle = \hat{a}_i \hat{a}_j^\dagger|\dots \chi_{i\uparrow}\chi_{i\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \pm \hat{a}_j^\dagger \hat{a}_i|\dots \chi_{i\uparrow}\chi_{i\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle$$

$$\hat{a}_i|\chi_{j\uparrow}\chi_{j\downarrow} \dots \chi_{i\uparrow}\chi_{i\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \pm \hat{a}_j^\dagger|\dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle = |\chi_{j\uparrow}\chi_{j\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \pm |\chi_{j\uparrow}\chi_{j\downarrow} \dots \chi_{k\uparrow}\chi_{k\downarrow} \dots\rangle \tag{7.12}$$

From this last derivation it is clear that the anticommutation relation is a complete mess: for $i \neq j$, i occupied, j unoccupied it is $2\hat{a}_i \hat{a}_j^\dagger$. There is still hope for the commutation relation since for $i \neq j$ it is always 0, just like the normal creation and annihilation operators have. The $i = j$ case however is not very workable since the commutator is $\delta_{ij}(1-2n_i)$. This means the fermionic anticommuation relation does not hold and neither does the bosonic commutation relation. The latter does not really matter as the Jordan-Wigner transform is specifically defined to map fermion systems to half spin systems (qubit systems) forcing the fermionic anticommutation relation to hold on the half spin system. Since the goal here is to map the electron pair system to a qubit system it immediately follows that this is not possible because it is shown that the anticommutator is different to the fermionic one. A different transformation that forces the specific commutation/anticommutation relations of this system would have to be developed to implement Closed Shell VQE as envisioned. Recall that there are smarter methods that can already implement Closed Shell VQE and that this result is just for the implementation used in the VQE library of the CQT group.

7.3 HF superposition trick for convergence speedup

Because the entire Hamiltonian is calculated, the qubit representation of the FCI solution can be found easily. This is the lowest eigenvalue of the Hamiltonian and by looking at it, one may expect to get more insight into the solution. There are two observations that are important. Firstly, the Hartree-Fock state (which is a basis state of the qubit statespace), has a very big coefficient (absolute value > 0.96) for all investigated geometries. Additionally, the vector was very sparse. Only 133 out of 512 terms had any contribution bigger than 10^{-10}. Indeed replacing all entries smaller than the 10^{-10} threshold with 0 results in an error with relation to the FCI solution of order 10^{-13} Hartree. This is a good indication that the entanglement depth does not have to increase the rotation variable to an 512 dimensional variable. It is clear that many more basis states (next to the trivial ones left over from the projections) have no component in the global minimum. Since the Hartree-Fock state is very prominent in the final solution, a way to restrict the search space could be to add more of the Hartree-Fock state to the sample states. This will be called a Hartree-Fock superposition (HFS). HFS is implemented in simulations by adding some (variable) constant R to the coefficient that belongs to the Hartree-Fock state and then normalizing the state vector again. Define $q(\vec{\theta})$ as the coefficient vector of $|q(\vec{\theta})\rangle$ created by rotation matrix $\vec{\theta}$ then the scheme is as follows:

$$q(\vec{\theta}) = \begin{pmatrix} a_1 \\ \vdots \\ a_{HF} \\ \vdots \\ a_{2^N} \end{pmatrix} \rightarrow \begin{pmatrix} a_1 \\ \vdots \\ a_{HF} + R \\ \vdots \\ a_{2^N} \end{pmatrix} \rightarrow \frac{1}{\sqrt{\sum a_i^2}} \begin{pmatrix} a_1 \\ \vdots \\ a_{HF} + R \\ \vdots \\ a_{2^N} \end{pmatrix} = \hat{q}_R(\vec{\theta}) \tag{7.13}$$

Where in the normalization factor on the right side a_{HF}^2 is of course replaced by the new coefficient, namely $(a_{HF} + R)^2$.

These new states can be measured efficiently with a system similar to the already existing quantum circuit, requiring only a single additional ancilla qubit [48]. Consider the state $|q_R\rangle$ (dropping the $\vec{\theta}$ here and in the rest of this section).

$$|q_R\rangle = \frac{1}{\alpha}\left(|q\rangle + R|\phi_0\rangle\right) \tag{7.14}$$

Here $\alpha = \langle q_R|q_R\rangle$ is the normalization constant. The energy of this state can be written out as follows:

$$\langle q_R|H|q_R\rangle = \left(\frac{1}{\alpha}\right)^* \frac{1}{\alpha}\left(\langle q|H|q\rangle + R^2\langle\phi_0|H|\phi_0\rangle + R\langle q|H|\phi_0\rangle + R\langle\phi_0|H|q\rangle\right). \tag{7.15}$$

Here R is known, $\langle q|H|q\rangle$ can be calculated using the quantum system as normal, and $\langle\phi_0|H|\phi_0\rangle$ is the Hartree-Fock energy, which is known. There are some remaining unknowns: $\langle q|H|\phi_0\rangle + \langle\phi_0|H|q\rangle$ and α, which can be written as:

$$\alpha = \langle q_R|q_R\rangle = \langle q|q\rangle + R^2\langle\phi_0|\phi_0\rangle + R\langle q|\phi_0\rangle + R\langle\phi_0|q\rangle = 1 + R^2 + R\left(\langle q|\phi_0\rangle + \langle\phi_0|q\rangle\right). \tag{7.16}$$

The unknown terms are:

$$\langle q|H|\phi_0\rangle + \langle\phi_0|H|q\rangle \quad \text{and} \quad \langle q|\phi_0\rangle + \langle\phi_0|q\rangle \tag{7.17}$$

It is known that $|q\rangle = \mathbf{U}_d|\phi_0\rangle$ and with this the terms are written with ϕ_0 in the bra and the ket. Additionally identities are added because they make the step after clearer:

$$\langle\phi_0|\mathbf{U}_d^\dagger HI|\phi_0\rangle + \langle\phi_0|IH\mathbf{U}_d|\phi_0\rangle \quad \text{and} \quad \langle\phi_0|\mathbf{U}_d^\dagger I|\phi_0\rangle + \langle\phi_0|I\mathbf{U}_d|\phi_0\rangle \tag{7.18}$$

For any Hermitian observable $\hat{Q}$, any two operators B, C and a valid state $|\psi\rangle$ ($\hat{Q}$ can act on $B|\psi\rangle$ and $C|\psi\rangle$) it holds that [48]:

$$\langle\psi|B^\dagger\hat{Q}C|\psi\rangle + \langle\psi|C^\dagger\hat{Q}B|\psi\rangle = \frac{1}{2}\left(\langle\psi|(B+C)^\dagger\hat{Q}(B+C)\psi\rangle - \langle\psi|(B-C)^\dagger\hat{Q}(B-C)\psi\rangle\right). \tag{7.19}$$

Since I and H are Hermitian and $I^\dagger = I$, this equality can be used for both unknowns. It turns out the α term is actually needed to determine $\langle\phi_0|H|q\rangle + \langle q|H|\phi_0\rangle$ so that is the starting point. First add an ancilla qubit to the system in the following way:

Here the H gate is the Hadamard gate given by

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{7.20}$$

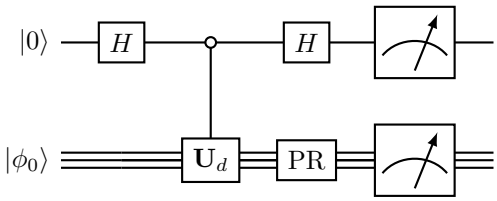


Figure 7.4: Quantum circuit diagram of the HFS circuit. The top line is the ancilla qubit and the bottom bundle are the qubits of the regular quantum circuit. Note that except for the conditioning on the ancilla qubit the circuit has stayed the same for the regular qubits. Recall that the post-rotations and measurements on the original circuit are per qubit unlike what the figure may suggest.

and the top qubit is the ancilla qubit. The normal state preparation circuit is (except for the conditioning on the ancilla qubit) still the same on the rest of the qubits. Now, after the first Hadamard gate the system has state

$$\frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \otimes |\phi_0\rangle \tag{7.21}$$

Next the conditional gate is applied. In the general case there should be two [48]: one conditioned on the ancilla being $|0\rangle$ and one conditioned on the ancilla being $|1\rangle$, however this is avoided since in this case the second operator is not a general one, but the identity. This appears automatically and after just the one conditional gate the state of the system is:

$$\frac{1}{\sqrt{2}}\left[|0\rangle\mathbf{U}_d|\phi_0\rangle + |1\rangle|\phi_0\rangle\right] = \frac{1}{\sqrt{2}}\left[|0\rangle\mathbf{U}_d|\phi_0\rangle + |1\rangle I|\phi_0\rangle\right] \tag{7.22}$$

After the second Hadamard the final state is reached:

$$\frac{1}{\sqrt{2}}\left[|0\rangle(\mathbf{U}_d + I)|\phi_0\rangle + |1\rangle(\mathbf{U}_d - I)|\phi_0\rangle\right] \tag{7.23}$$

Measuring the ancilla now collapses the rest of the system into one of the two branches:

$$|\psi_0'\rangle = \frac{1}{2\sqrt{p_0}}(\mathbf{U}_d + I)|\phi_0\rangle, \quad p_0 = \frac{1}{4}\left\langle\phi_0\right|(\mathbf{U}_d + I)^\dagger(\mathbf{U}_d + I)\left|\phi_0\right\rangle \tag{7.24}$$

$$|\psi_1'\rangle = \frac{1}{2\sqrt{p_1}}(\mathbf{U}_d - I)|\phi_0\rangle, \quad p_1 = \frac{1}{4}\left\langle\phi_0\right|(\mathbf{U}_d - I)^\dagger(\mathbf{U}_d - I)\left|\phi_0\right\rangle \tag{7.25}$$

Where p_0 and p_1 are the probabilities of either state. Measuring the ancilla many times allows one to estimate p_0 and p_1, since they are the probabilities of the ancilla being in $|0\rangle$ and $|1\rangle$ respectively. By using Equation 7.19 with $\hat{Q} = I$ the term $\langle q|\phi_0\rangle + \langle\phi_0|q\rangle = \langle\phi_0|\mathbf{U}_d^\dagger I|\phi_0\rangle + \langle\phi_0|I\mathbf{U}_d|\phi_0\rangle$ is fully determined by p_0 and p_1:

$$\langle\phi_0|\mathbf{U}_d^\dagger I|\phi_0\rangle + \langle\phi_0|I\mathbf{U}_d|\phi_0\rangle = \frac{1}{2}\left(\langle\psi|(\mathbf{U}_d + I)^\dagger(\mathbf{U}_d + I)\psi\rangle - \langle\psi|(\mathbf{U}_d - I)^\dagger(\mathbf{U}_d - I)\psi\rangle\right) = 2p_0 - 2p_1 \tag{7.26}$$

Similarly by measuring H on the rest of the system like normal the conditioned expectations of the Hamiltonian can be found:

$$\hat{E}_0 = \langle\psi_0'|H|\psi_0'\rangle = \frac{1}{4p_0}\left\langle\phi_0\right|(\mathbf{U}_d + I)^\dagger H(\mathbf{U}_d + I)\left|\phi_0\right\rangle \tag{7.27}$$

$$\hat{E}_1 = \langle\psi_1'|H|\psi_1'\rangle = \frac{1}{4p_1}\left\langle\phi_0\right|(\mathbf{U}_d - I)^\dagger H(\mathbf{U}_d - I)\left|\phi_0\right\rangle \tag{7.28}$$

Like before, using Equation 7.19 now with $\hat{Q} = H$ the term of interest ($\langle q|H|\phi_0\rangle + \langle\phi_0|H|q\rangle = \langle\phi_0|\mathbf{U}_d^\dagger HI|\phi_0\rangle + \langle\phi_0|IH\mathbf{U}_d|\phi_0\rangle$) can be written in terms of only $\hat{E}_0, \hat{E}_1, p_0$ and p_1:

$$\begin{aligned}\langle\phi_0|\mathbf{U}_d^\dagger HI|\phi_0\rangle + \langle\phi_0|IH\mathbf{U}_d|\phi_0\rangle = \frac{1}{2}\left(\langle\psi|(\mathbf{U}_d + I)^\dagger H(\mathbf{U}_d + I)\psi\rangle - \langle\psi|(\mathbf{U}_d - I)^\dagger H(\mathbf{U}_d - I)\psi\rangle\right)\\ = 2p_0\hat{E}_0 - 2p_1\hat{E}_1\end{aligned} \tag{7.29}$$

With this all terms of $\langle q_R|H|q_R\rangle$ can be determined with 3 times more measurements than a normal $|q\rangle$ (now H has to be measured on the normal circuit and twice as much on the conditional circuit (for $\hat{E}_0$ and $\hat{E}_1$)). The ancilla measurements to determine p_0 and p_1 can be done at the same time as the measurements of H on the rest of the system.

7.4 Improvements in running times

During the project running time was saved as much as possible. Many of the optimizations are obvious in hindsight, but they are included anyway for completeness. The most obvious time save (which was already in use) was saving Hamiltonians. This allows one to calculate them a single time and then be done. This is not very important for the smaller Hamiltonians, but as the full 14 qubit Hamiltonian was too big to calculate at the start of the project and still took in the order of hours after optimizations, it is very important for the bigger ones. Similarly, when looking at multiple geometries using the same orbital structure for their Hamiltonians their projection matrices are the same. Calculating these individually while calculating the Hamiltonians is not an issue for small Hamiltonians, but for the bigger ones calculating these projection matrices took in the order of 20 minutes. Precalculating them and loading them instead of calculating them each time saved a lot of time. The big time save for the Hamiltonian calculation was due to a switch to sparse matrices. The problem was that the system could not be changed to sparse matrices because some of the operations used did not work on sparse matrices. However in an attempt to save time in a different way all possible Jordan-Wigner operator transform matrices were precalculated. These precalculated matrices could be saved as sparse matrices, resulting in a workaround to the first problem. Now the Hamiltonian could be computed using only sparse matrix multiplication saving enough time to reduce the running speed from months (estimated) to hours. At this point the only thing that had to be worried about was the state preparation. The Kronecker products that build the rotation matrices take up the largest part (almost the entire part) of the simulation. The Z N-qubit rotation however determined by a vector of rotations $\vec{\phi}$ (in practice a column of the rotation matrix $\vec{\theta}$), can be calculated without Kronecker products, due to the fact that the Z-rotation matrix $R_Z(\phi) = Z_\phi$ is diagonal, with terms only differentiating in sign. This means the Kronecker product is also diagonal. The following derivation shows where the trick is headed:

$$Z_{\vec{\phi}} = \bigotimes_j^N Z_{\phi_j} = \bigotimes_j^N \begin{pmatrix} e^{-i\phi_j/2} & 0 \\ 0 & e^{i\phi_j/2} \end{pmatrix} = \begin{pmatrix} \prod_i^N e^{\pm i\phi_j/2} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & & \prod_j^N e^{\pm i\phi_j/2} & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \prod_j^N e^{\pm i\phi_j/2} \end{pmatrix}$$

$$= \begin{pmatrix} e^{(\sum_j^N \pm i\phi_j/2)} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & & e^{(\sum_j^N \pm i\phi_j/2)} & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & \dots & 0 & \dots & e^{(\sum_j^N \pm i\phi_j/2)} \end{pmatrix} \tag{7.30}$$

Here the $\pm$ depends on both j and the row of the matrix. This is due to the structure of the Kronecker product. For example for $j = 1$ the $\pm$ will be $+$ for the first $\frac{2^N}{2}$ rows and $-$ for the rest, while for $j = N$ the $+$ and $-$ will alternate. Important to note is that, when disregarding the sign each term gets reduced to $i\phi_j/2$ for each j. This is what was used to speed up the Kronecker product. Arrays of length 2^N were generated with 1 and -1 in the right places for each $1 \leq j \leq N$. These take very little time to create and therefore dont have to be saved. They should be precalculated in the setup however, since else they would be recalculated each function evaluation. Now each one is multiplied by $i\phi_j$ to get for each j the terms for each diagonal element. Now all these arrays are added together and the result is an array with exactly the terms from the Z rotation, however they still have to be taken to the power e element-wise. Note that this is a time save compared to raising each element to the power e and then doing element-wise multiplications on the arrays. The result is an array holding the diagonal terms of $Z_{\vec{\phi}}$. A last bit of time was saved by instead of transforming the array back to the full matrix, element-wise multiplication was used to rotate the qubit state q. This is of course the same as $Z_{\vec{\phi}}$ is diagonal.

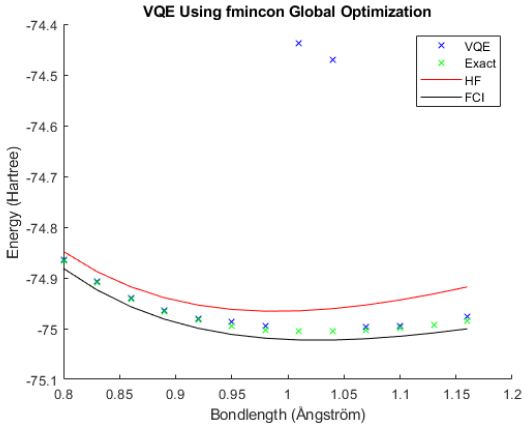
Chapter 8

VQE Simulations of H_2O

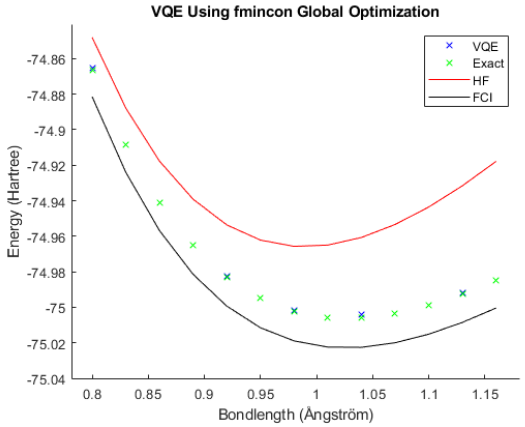
With a workable Full Hamiltonian (9 qubits after projection) attempts were made to reach the FCI solution. The main goal of course is to improve on HF, but getting close to the global minimum would of course be even better. As one may expect finding the global minimum in a 512 (2^9) dimensional space is hard, especially when the variable to optimize is not the vector itself but some rotation matrix that creates the vector. Naive initial simulations resulted in bad results, almost no geometries were optimized to be even near the HF energy, even when the starting point was very close to the HF solution. Initially SPSA was used, as this was already part of the CQT VQE library. It did not perform well and even when cheating and starting at the exact solution the algorithm would walk away from it and not find it back. This behavior specifically could be due to the initial steps of these algorithms being big in relation to the search space, or because of some other reason. Initial tests with the nonlinear global optimizer from MATLAB were very good. That is the reason fmincon was used as optimizer throughout this section, and not SPSA. The SPSA results are discussed in section 9.3. At the start of simulations the results from the 14 orbital system were not good enough to outperform the theoretical minimum of the smaller systems. Additionally the running times were very long so smaller systems were of interest to save time as well. Starting with 8 orbitals (6 qubits) and 10 orbitals (7 qubits), which both came significantly closer to their theoretical minimum (as one would expect), the 12 orbital (8 qubit) system was considered in depth as well. Afterwards the HFS trick was thought off and implemented first in the 12 and 14 orbital systems. The 10 orbital system was then used to look at the method under convergence of the optimizer. Lastly simulations on the convergence of the optimizer on the 12 orbital system were ran. All simulations used either a fixed amount of iteration steps (2000 was chosen after some experimental simulations), or stopped after the global optimizer determined convergence. The word convergence may be used loosely in this section, as not all systems converged to their exact solutions (get within chemical accuracy), even though the optimizer did converge. It should be clear from the context which type of convergence is talked about.

8.1 8 orbitals system

To get to 8 orbitals from 14 a total of 3 spatial orbitals have to be removed ($14 - 3 * 2 = 8$ spin-orbitals). Looking at the relevant part in Figure 7.2 one can see that the best performing combination around a bond length of 1 angstrom is the one with orbitals $1b_1$,$3a_1$,$4a_1$ and $2b_1$ (i.e. $1a_1$,$2a_2$ and $1b_2$ removed). Note that this is not a combination that would have been predicted as the best one before. Going with the lowest absolute value orbitals would of course not remove $1b_2$ (which is the lowest absolute value energy orbital). Experiments were done to look at the convergence of the 8 qubit system as function of the entanglement depth. Since the 8 orbital system is 6 qubits after reduction one would expect that a depth of 3 would be enough. At a depth of 3 the rotation matrix has dimensions $3 * 3 + 2 = 11$ by 6 which is 66 and thus bigger than the dimension of the qubit state q which is $2^6 = 64$. However at entanglement depth 3 the classical optimizer sometimes reaches termination criteria without good results. This is shown in Figure 8.1a. There is no clear explanation, but possibilities are either that the space is not linearly independent and thus lower dimensional than the aforementioned 66 or that the classical optimizer for some reason sometimes fails to find the global minimum at low entanglement depth. It is not clear which one of these (if either) is the correct reason: for the latter it is unclear if the optimizer should perform worse (e.g. get stuck in local minima more often) in higher dimensional problems, because there is more space, or perform better, because there are more clear paths towards minima. For very large entanglement depths the result gets bad, but here it gets better at for example entanglement



(a) Entanglement depth 3



(b) Entanglement depth 5

Figure 8.1: Plots of the energy of H_2O at different bond lengths for the 8 orbital system. Entanglement depths 3 (a) and 5 (b). Blue are the VQE simulation results, green are the exact results for the 8 orbital Hamtilonian, red is the HF solution and black the FCI solution. When no blue mark is visible it is under the green mark. Of special interest are the two outliers at entanglement depth 3 in (a). All errors in (b) are within chemical accuracy of the exact solution of the 8 orbital Hamiltonian.

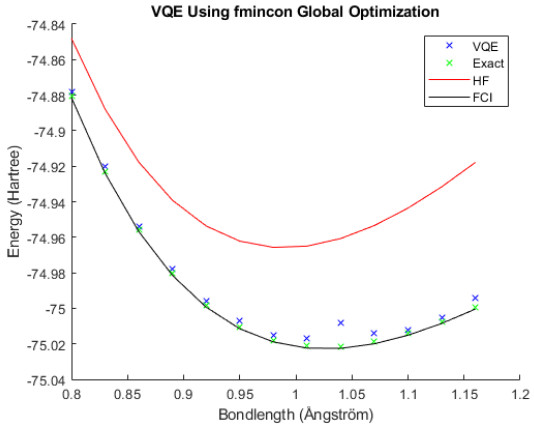
depth 5, this can be seen in Figure 8.1b. Additionally, different geometries still perform well and the difference is very big between the converging and non-converging points, which indicates that low dimensionality may not be the issue. Note that almost all the results in Figure 8.1 improve on Hartree-Fock and are within chemical accuracy of the exact value for the approximated 8 orbital system. Except for illustrating this weird behavior the 10 orbital system easily outperformed the 8 qubit one without unreasonable running times. Therefore the 8 orbital system was not used further.

8.2 10 orbital system

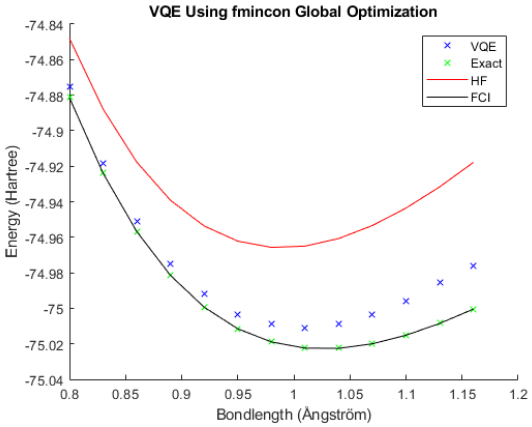
The 10 orbital system turned out to be somewhat of a sweet spot for the project. It was the biggest system small enough to run a few dozen simulations with no iteration cap and was therefore used for the main results on the HFS scheme. The removed molecular orbitals for this system are, as discussed in section 7.1, $1a_1$ and $1b_2$ as suggested in [11]. Once again this is not a orbital set that would have been picked using the absolute value criteria that was used before. It can be predicted theoretically however [11]. An important note is that the exact solution of the 10 orbital system is within chemical accuracy of the FCI solution for bond lengths around 1Å, for the angle considered in this project. This means that it is possible to reach the goal of FCI using this system (and the bigger ones of course).

8.2.1 Entanglement Depth

This 7 qubit system was considered using entanglement depth 7 and 10. Entanglement depth 7 was chosen as the lowest entanglement depth before the results worsened even though there was a slight outlier present at entanglement depth 7. It was concluded that this was due to inconsistent convergence and a stopping criteria of a fixed amount of iterations rather than due to the entanglement depth. The difference between depth 7 and 8 can be seen in Figure 8.2. The difference was initially attributed to the entanglement depth but it is not certain that for more iterations the 7 depth would still outperform the 8 depth. Because of this and for fairer comparison the 10 orbital system was also looked at with entanglement depth 10, which is the depth chosen for the 12 and 14 orbital systems. There was not enough time to let the system converge without iteration limit for a range of entanglement depths, which would be a better way to determine the correct choice for the simulations without iteration limit that follow.



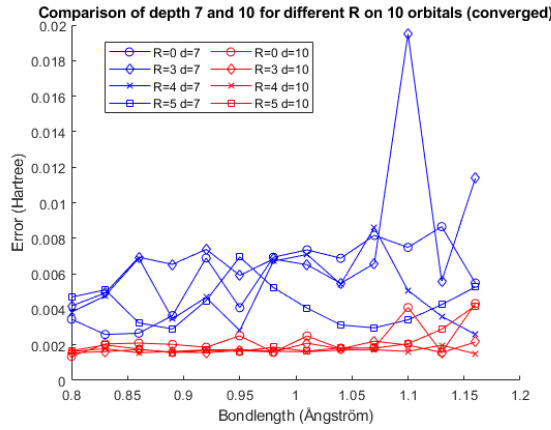
(a) Entanglement depth 7



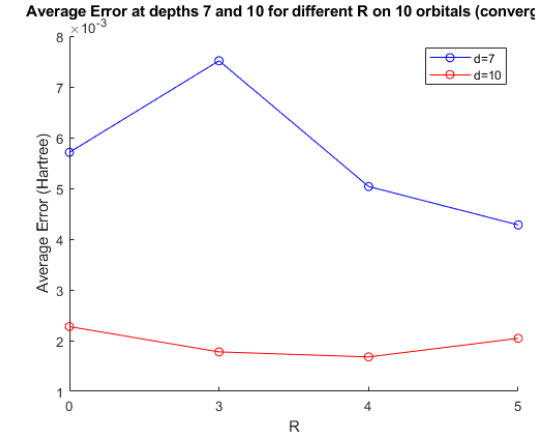
(b) Entanglement depth 8

Figure 8.2: Plots of the energy of H_2O at different bond lengths for the 10 orbital system. Entanglement depths 7 (a) and 8 (b). Blue are the VQE simulation results, green are the exact results for the 10 orbital Hamtilonian, red is the HF solution and black the FCI solution. Of special interest is the height difference between (a) and (b).

Comparing the 7 and 10 entanglement depth was fairly simple. Convergence of the global optimizer took longer for entanglement depth 7 than for entanglement depth 10, sometimes even timing out before convergence on the very high boundaries still in place (for all other systems attempted the optimizer converged within these boundaries). This seems similar to what happened for 8 orbitals and a possible explanation is again that even though the search space is larger for $d = 10$, finding paths towards a minimum is also easier due to the dimensionality. As a result of this only a few simulations of both depths were compared to see which one performed better. The outcome was pretty clear with $d = 10$ outperforming $d = 7$ by a significant margin when the global optimizer was ran until convergence. Not only when using the HFS trick, but also without it. In Figure 8.3 the errors are shown for various parameters R of the HFS. These specific values of R where chosen because they were integers around the minimum (more about that later) and the other values in the interval $[0, 7]$ had convergence issues for $d = 7$. Because of these results $d = 7$ was not considered further.



(a) Error per bond length

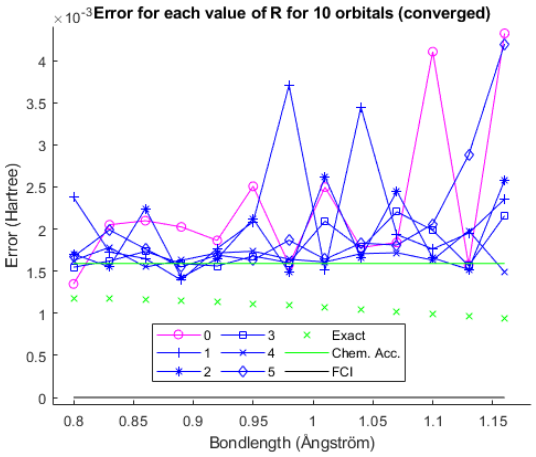


(b) Average error per R value

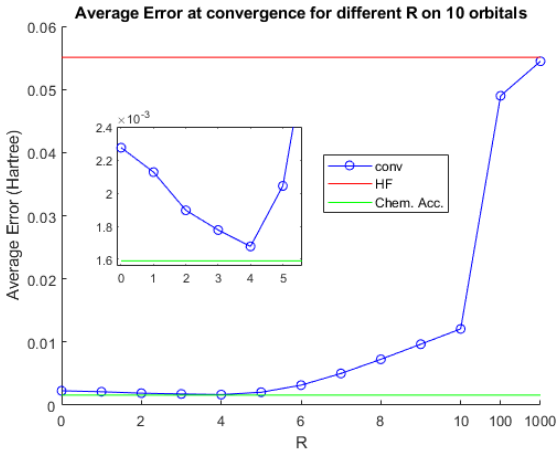
Figure 8.3: Comparison of entanglement depths 7 and 10 on H_2O with 10 orbitals considered. Red is depth 10 and blue is depth 7. In (a) the error per bond length is shown for different values of R of the HFS. In (b) the average error over the interval of bond lengths is shown per value of R.

8.2.2 Hartree-Fock superposition

When investigating the HFS scheme three key questions were considered. Does the scheme indeed improve convergence in the sense that it performs well at low total iteration steps? Does the scheme worsen results when the amount of iteration steps is enough for convergence without the trick? And lastly: What is the best value for R? On the 10 orbital system all three of these questions were looked at. On the 12 orbital system all three were looked at without much depth (the simulations took very long). On the full system the second question was not looked at, as it could not converge in reasonable time. To answer these questions a range of R values were used at different convergence criteria. First consider the case where the optimizer is allowed to converge (i.e. terminates not due to a maximum amount of steps or iterations, but due to internal stopping criteria). This case is the best one to determine the best R value, because one would expect that for other stopping criteria the energies obtained would be equal at best. At the same time, this would show how much worse the HFS would make it for the optimizer to find a good minimum, as the pinching of the space could very well restrict it. This turned out to be not the case and R-values below 6 results were, on average, better than without the HFS ($R = 0$). This can be seen in Figure 8.4. Indeed in Figure 8.4a it can be seen that for some bond lengths the HFS allows the optimizer to find energies within chemical accuracy of FCI. Additionally this plot is a good reason to expect that the optimizer is outmatched by the problem. One would expect more consistency when allowed to converge. This is not the case and even the difference between the R-values of interest is not clear when looking at Figure 8.4a. When the average error per R-value is plotted it becomes more clear (see Figure 8.4b). The R-value with smallest average error is $R = 4$, almost reaching chemical accuracy on average. This plot also served as a sanity check as it is clear that theoretically for $R \to \infty$ the HFS should keep the system locked in the HF ground state resulting in the Hartree-Fock energy as minimum, which can be seen in the figure.



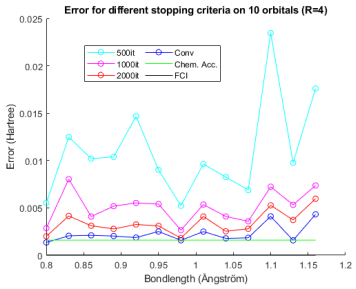
(a) Error per bond length for R-values of interest ($R < 6$). $R = 0$ marked in magenta for clarity.



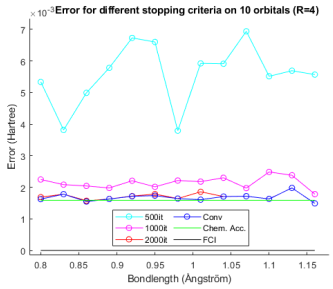
(b) Average error per R value, values of interest are zoomed in on.

Figure 8.4: Comparison of different R-values on H_2O with 10 orbitals considered. Iterations are stopped by the optimizer's convergence criteria. The green line is chemical accuracy w.r.t. FCI. In (a) the error per bond length is shown for different values of R of the HFS and the exact solution of the system is indicated with green marks. In (b) the average error over the interval of bond lengths is shown per value of R. Note that in (a) there is no clear better R-value, but in (b) it is shown that $R = 4$ has the lowest average error per bond length, performing only slightly below the chemical accuracy.

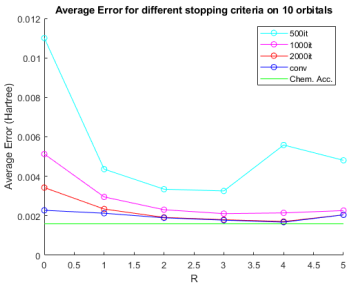
Looking at different stopping criteria it is expected that for low iteration steps the results will be very bad and then they will gradually increase up to the convergence results. The results in Figure 8.5 show that this was indeed what happened. In Figure 8.5a and Figure 8.5b below only the $R = 0$ and $R = 4$ results are shown for all convergence criteria as all other simulated R-values had similar results but are less interesting. Indeed for all simulated R-values the criteria are nicely ordered above each other. In Figure 8.5c the average errors per R-value for all stopping criterion are shown. Here the expected behavior appears as well, as the averages are also nicely ordered. Observations of interest are that as expected, the difference between iteration steps becomes smaller for larger R. This is an indication that convergence requires less steps for bigger R. Additionally while 500 iterations is significantly worse, the 1000 iterations perform quite decently and the difference between 2000 iterations and convergence of the optimizer has almost disappeared at $R = 4$.



(a) Error per bond length for different stopping criteria with $R = 0$.



(b) Error per bond length for different stopping criteria with $R = 4$.



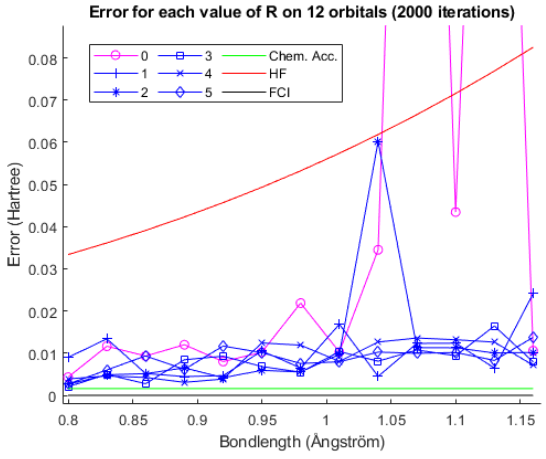
(c) Average error per R value for different stopping criteria.

Figure 8.5: Comparison of different stopping criteria on H_2O with 10 orbitals considered. In (a) and (b) the error per bond length is shown for different values of R (0 and 4 respectively) of the HFS. In (c) the average error over the interval of bond lengths is shown per value of R. Note that in all three figures the convergence criteria nicely keep their hierarchy.

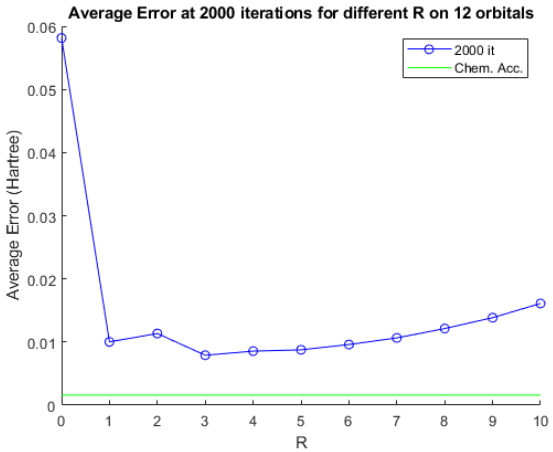
8.3 12 orbital system

To get the 12 orbital system only $1a_1$ was taken as occupied. The exact ground state energy of the 12 orbital system is very close to FCI (see Figure B.8). After picking entanglement depth 10 for the Full Hamiltonian, the same was done with the 12 orbital system. The main reason was fair comparison of the 12 and 14 orbital systems. Obviously the 9 qubit system would require more qubits and gates than the 8 qubit one, but certain attributes of the circuit scale with depth and not with size. A maximum of 2000 iteration steps was taken as convergence criteria, also the same as the full Hamiltonian. These simulations took approximately 1.5 days to complete. Later it was attempted to let the global optimizer converge on the 12 orbital system using HFS and these simulations took approximately 3 days for $R = 3$ and $R = 5$, but more than a week for $R = 0$.

The HFS on 12 orbitals was looked at for 2000 and unlimited iteration steps. First the 2000 iteration step stopping criterion was considered. A similar plot as with the 10 orbital system was made for the 12 orbital system (with less R-values). This is shown in Figure 8.6. Again it is not clear which of the R-values of interest are performing best in Figure 8.6a, but it can be seen that without the HFS trick the results are sometimes even worse than HF at 2000 iteration steps. These outliers are not visible in the figure, but are at 0.311 and 0.268. In Figure 8.6b the average errors are shown. Here it can be seen that the best R-value at 2000 iteration steps is likely in the interval $[2, 4]$ assuming the relation to be continuous.



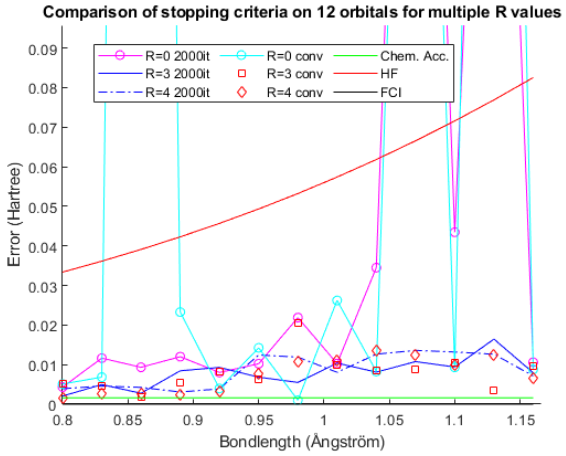
(a) Error per bond length for R-values of interest ($R < 6$). $R = 0$ marked in magenta for clarity.



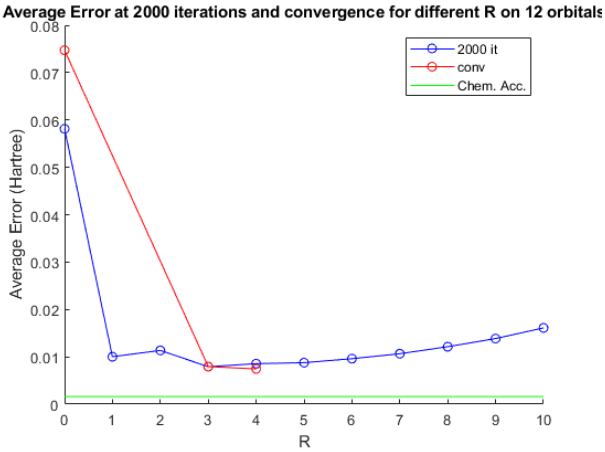
(b) Average error per R value.

Figure 8.6: Comparison of different R-values on H_2O with 12 orbitals considered. Simulations are stopped after 2000 iteration steps. The green line is chemical accuracy w.r.t. FCI. In (a) the error per bond length is shown for different values of R of the HFS. In (b) the average error over the interval of bond lengths is shown per value of R. Note that again in (a) there is no clear better R-value (although $R = 0$ does poorly for a few bond lengths). In (b) the difference between $R = 0$ and the other R-values becomes clearer.

Afterwards the 12 qubit system was allowed to converge, but due to this taking so much time, only a few R-values were considered: $0, 3$ and 4. The results are shown in Figure 8.7. For $R = 0$ an incredible outlier was found. This outlier was not taken into consideration for the average errors, as it was almost 3 times larger than the outliers for $R = 0$ at 2000 iteration steps. It can be seen in Figure 8.7a ,where the line goes out and back into the figure very fast around bond length 0.85Å. It is completely unclear what happened at this specific bond length. Not only this but also the outliers that were already present were equal or worse for the convergence stopping criterion. It is unexpected that the converged result does not outperform the 2000 iterations result every time, especially since the 10 orbital system did not show this behavior. This is likely another indicator that the optimizer can not handle the system size and gets stuck in local minima. It is a bit strange however how this can happen for systems with the same R-value, as the optimizer should be deterministic. The previous plot of the average errors is shown again, this time with the added points for $R = 0, 3, 4$ with different stopping criteria. For $R = 3$ and $R = 4$ it is fine to compare stopping criteria, but for $R = 0$ some caution is required. Because of the outlier the mean would not be very representative. Therefore the outlier was removed from the average in Figure 8.7b, making the comparison between the 2000 iteration steps and optimizer convergence fairer. Figure 8.7b shows that the converged results are better than the 2000 iteration ones on average, as one would expect except for $R = 0$. Removing all outliers from both averages changes this result in favor of the convergence.



(a) Error per bond length all R-values simulated with both stopping criteria. $R = 0$'s are marked in magenta and cyan for clarity.

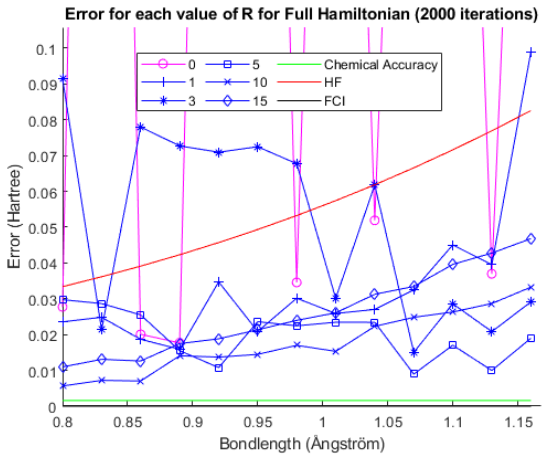


(b) Average error per R value.

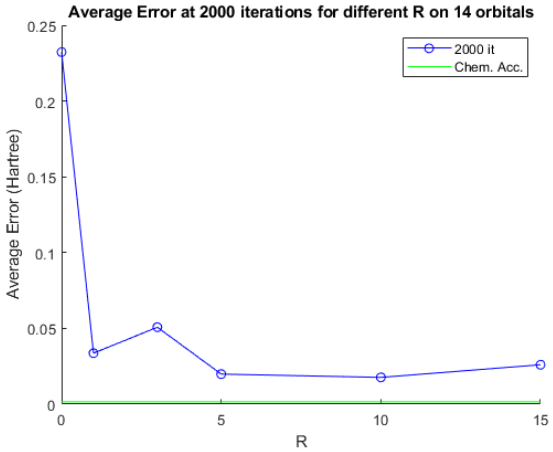
Figure 8.7: Comparison of different R-values on H_2O with 12 orbitals considered. Simulations are for different stopping criteria: 2000 iterations and optimizer convergence. The green line is chemical accuracy w.r.t. FCI. In (a) the error per bond length is shown for different values of R of the HFS. In (b) the average error over the interval of bond lengths is shown per value of R. In blue 2000 iteration steps and in red optimizer convergence. Note that again in (a) there is no clear better R-value. Additionally note that optimizer convergence is not strictly outperforming. In (b) the difference between 2000 iteration steps and optimizer convergence is more visible. Note that the average of $R = 0$ with convergence had the outlier removed.

8.4 Full Hamiltonian system

As explained before, just like the 12 orbital system the entanglement depth was taken as 10, knowing that it may very well be too low. Additionally the convergence criteria was 2000 iteration steps, which is no where close to convergence without tricks. The results did however confirm the convergence speedup (in terms of required iteration steps) of the HFS trick. The results without the trick are very bad even in comparison to R values that the other simulations showed to be too high, most likely due to the size of the system. This is shown in Figure 8.8. This time the superiority of the HFS is very clear. For low values of R there are many values above the HF solution (see Figure 8.8a). The optimal R-value is less clear due to bigger step size between values of R, but the best performing on average is $R = 10$ (not clear in Figure 8.8b, but the point at $R = 10$ is slightly lower than the one at $R = 5$).



(a) Error per bond length all simulated R-values of interest ($R < 6$). $R = 0$ is marked in magenta for clarity.



(b) Average error per R value.

Figure 8.8: Comparison of different R-values on H_2O with all 14 orbitals considered. Simulations are stopped after 2000 iteration steps. The green line is chemical accuracy w.r.t. FCI. In (a) the error per bond length is shown for different values of R of the HFS. In (b) the average error over the interval of bond lengths is shown per value of R. Note that this time in (a) there is some clear difference between R-values. In (b) the best R-value is not that clear as the slope is very flat.

8.5 Comparison of the different systems

Since no new results are posted in this section, it is clear that the chemical accuracy w.r.t. FCI has not been reached with a single method for all bond lengths. However for completeness some of the different methods and system sizes, are compared directly here. The comparison is shown in Figure 8.9. It is clear that the 10 orbital system performed best.

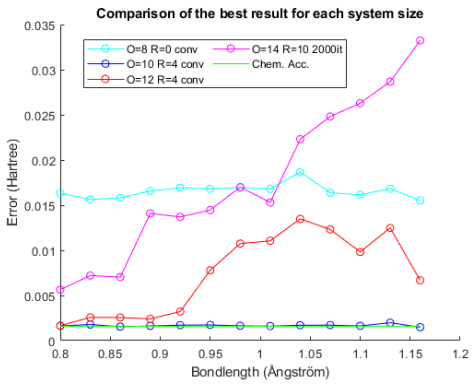
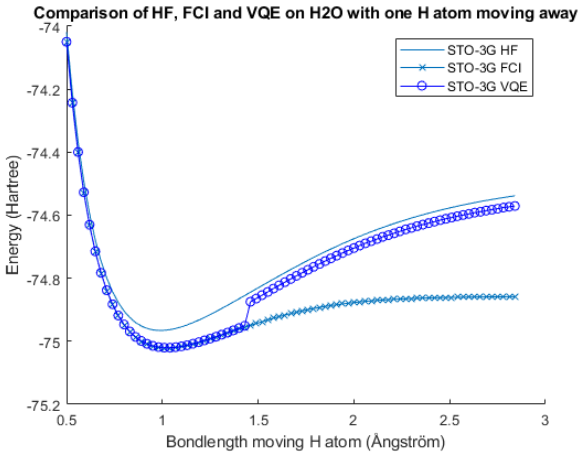
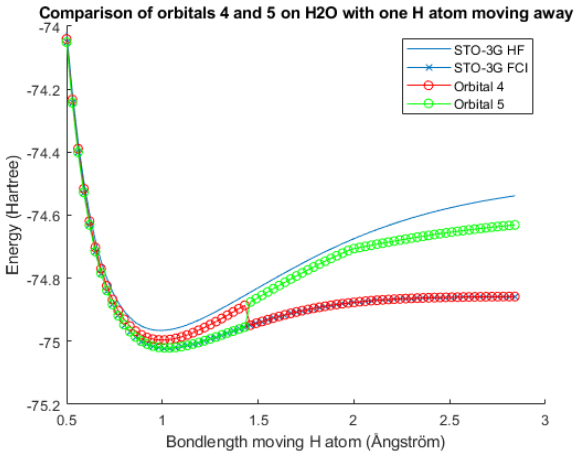


Figure 8.9: Comparison of the best methods for 8, 10, 12 and 14 orbitals. These are: for 8 orbitals $R = 0$ converged, for 10 orbitals $R = 4$ converged, for 12 orbitals $R = 4$ converged and for 14 orbitals $R = 10$ with 2000 iteration steps.

Now the problem from subsection 4.4.2 is reconsidered, but now using VQE instead of classical methods. The 10 orbital, $R = 4$ method with optimizer convergence was used. The results are shown in Figure 8.10. The sudden jump in Figure 8.10a seemed weird and the hypothesis was that the $1b2$ orbital dropped in energy below the next highest one ($3a1$). As said before: in the simulations the orbitals are ordered by energy and the position in this ordering is picked instead of the orbital itself. For all previous simulations $3a1$ was orbital 4 and $1b2$ was orbital 5 for all bond lengths. For these bond lengths this was most likely not the case, as indicated by Figure 8.10b. Here it can be seen that the lines suddenly exchange position. Because of this a second VQE simulation was done, switching the number of the orbital at the correct bond length, such that the $1b2$ orbital was filled for all bond lengths.

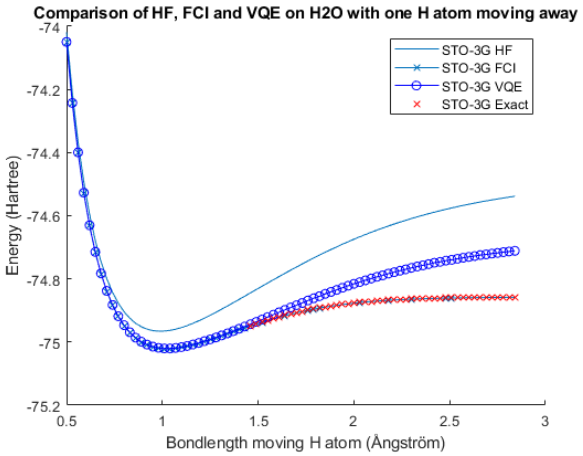


(a) Error per bond length of the moving atom. Shown for STO-3G HF, STO-3G FCI and STO-3G VQE on 10 orbitals with $R = 4$. Filled orbitals are 1 and 5 in terms of energy.

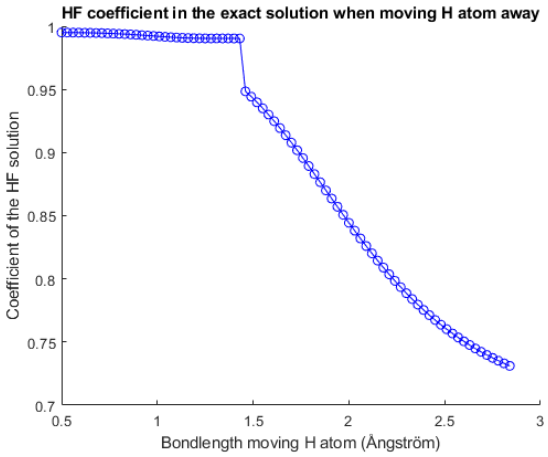


(b) Error per bond length of the moving atom. Shown for STO-3G HF, STO-3G FCI and the exact solutions of the 10 orbital Hamiltonian when filling orbitals 1 and 4 (shown as 4) and orbitals 1 and 5 (shown as 5).

Figure 8.10: In (a) the VQE simulation of the moving H atom system is shown. Of note is the abrupt jump from close to the FCI solution towards the HF solution. In (b) the same plot but now with the exact solutions of the Hamiltonians with 10 orbitals gotten by occupying orbitals 1 and 4 (red) and orbitals 1 and 5 (green).



(a) Error per bond length of the moving atom. Shown for STO-3G HF, STO-3G FCI and STO-3G VQE on 10 orbitals with $R = 4$. Additionally the exact solution is shown where the VQE solution worsens. Filled orbital is $1b2$ for all bond lengths.

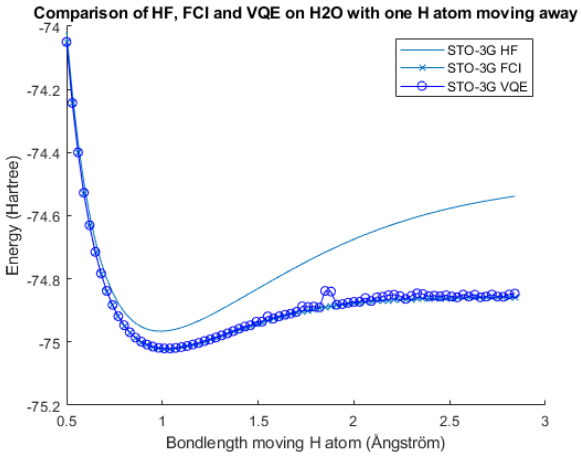


(b) Coefficients of the HF solution basis state in the exact solution of the Hamiltonian per bond length of the moving H atom.

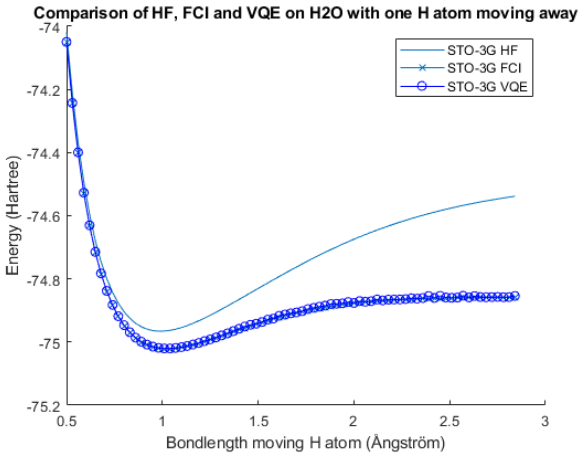
Figure 8.11: The results of the VQE simulations on H_2O with a moving H atom with 10 orbitals considered. In (a) the energy per bond length of the moving H atom is shown for $R = 4$. In (b) the coefficients of the HF solution in the exact solution of the Hamiltonian are shown per bond length of the moving H atom.

The second VQE simulation had some unexpected results, which led to some more simulations. The result of these are found in Figure 8.11. It is important to note that this VQE simulation was also done using $R = 4$ on 10 orbitals and the results from this specific simulation are shown in Figure 8.11a. It is clear that the results are not good. The VQE solution is not even close to the FCI one, while one would expect the method to perform very good based on previous results. It even has the same asymptotic shape as HF. The proposed explanation for this was that the HF solution was not as important anymore in the exact solution and that the HFS would therefore destroy the answer at a lower R-value. To investigate this the coefficients of the HF solution in the exact solution were calculated and plotted. This can be seen in Figure 8.11b. As expected a significant drop off in the importance (or considered differently: a drop off in accuracy) of HF was found.

To make sure that this was indeed the issue (and not for example that the FCI solution was harder to find in the asymptote), the simulation was repeated with $R = 0$ and $R = 1$. This is shown in Figure 8.12. In Figure 8.12a the asymptotic result is shown for $R = 0$ and it is clear that it (as expected) behaves as FCI. In Figure 8.12b it can be seen that $R = 1$ outperforms $R = 0$ in a similar way to before. Indeed, just as before convergence speed was faster as well. This also makes sense as the importance of the HF solution is still very significant, just not as big as before. These results show that the optimal value of R is variable depending on how good the HF solution approximates the FCI one.



(a) Error per bond length of the moving atom. Shown for STO-3G HF, STO-3G FCI and STO-3G VQE on 10 orbitals with $R = 0$.



(b) Error per bond length of the moving atom. Shown for STO-3G HF, STO-3G FCI and STO-3G VQE on 10 orbitals with $R = 1$.

Figure 8.12: The results of the VQE simulations on H_2O with a moving H atom with 10 orbitals considered. In (a) the energy per bond length of the moving H atom is shown for $R = 0$. In (b) the energy per bond length of the moving H atom is shown for $R = 1$. Values for bond lengths below 1.46 Åwere not simulated again to save time, so those are from VQE with $R = 4$.

Chapter 9

Miscellaneous results

The majority of this project have been failed attempts to make the full qubit system converge or to speed up running time. This section covers a non-exhaustive list of some of the more interesting minor results that did not really fit anywhere.

9.1 Optimize $|q\rangle$ directly instead of rotation matrix

To see whether the size of the system was not the biggest problem, simulations were done where q was the variable to optimize instead of the rotation matrix. The idea was that since the rotations take most of the running time, the running time may improve enough to counter the increase in dimension. Additionally there were concerns about the search space being too scrambled for gradient based optimizers. While this would not be applicable in a realistically sized quantum system, a positive outcome would be a good indication for where the problem lies. This did not happen as convergence was horrible and no results even remotely close to the HF solution were found using this method.

9.2 Precalculate X rotations

Because just about every slow matrix problem in the project was fixed by precalculating matrices, this was also attempted for the X rotation matrix. This is by far the slowest part in the simulation as it currently stands. A similar method to the one used on the Z rotation was attempted, with block matrices instead of the array, but the multiplications needed to combine the results at the end already took longer than the entire Kronecker product so this idea was binned.

9.3 SPSA

SPSA always got stuck in local minima, resulting in very fast convergence, but bad results. A lot of things have been attempted. Pulling towards the Hartree-Fock state does not work for SPSA, since it will converge towards the Hartree-Fock solution and get stuck there. Tweaking the parameter a that determines the convergence speed did not help either. Calibrating a was attempted, but since the rotations are symmetric around the zero-rotation gradient approximation was not easily done. Instead a range of a values greater than 1 were attempted and no better results were found. The other parameters were not investigated as their optimal values were taken from literature in the first place. It is known that SPSA is very dependent on the initial state. However, even when starting close to the exact solution no good results were achieved. The most likely explanation is that due to the scrambling of the search space any sample step will move so far away from the starting state that the starting state does not matter. This combined with good results from fmincon were the reason to stop using SPSA to try to improve on the Hartree-Fock solution for H_2O.

For completeness some tests done on SPSA are mentioned. Firstly due to the randomness of SPSA and the quick

convergence, one can use many runs with a single starting point to improve results (one takes the minimum of all SPSA runs). While it seems like a good idea, for bigger systems it does not help since all runs will be bad anyway. Picking the best out of 10 gives no relevant improvement. This addition however may be useful when good starting locations are used/or SPSA is used in a multi-start framework. The behavior of SPSA as a relation of entanglement depth is very well simulatable. By using the above idea of doing multiple runs with the same starting point one can plot a spread of energies at each entanglement depth. Allowing for a comparison in variance of the results for each entanglement depth.

Chapter 10

Conclusion

10.1 Conclusions and Discussion

VQE is a promising method to improve on the easy to compute, but not amazing, Hartree-Fock solution of a system. Implementing it is still cutting edge, but simulating it can provide knowledge or methods that can be put to use when the implementation becomes accessible. The focus of this thesis was to look at H_2O as a medium sized molecule of interest and look to solve issues that arise when VQE has to look at bigger systems and use VQE to improve on the Hartree-Fock solution of H_2O, getting as close to the FCI solution as possible. Since the problem has many aspects and many parameters some were ignored. The same basic entangler, the same starting state, the same angle of the water molecule and the same discrete set of bond lengths were used in the simulations. This list is not exhaustive and it is not unreasonable that some of the issues arise from a badly chosen parameter that was kept constant throughout. Specifically the starting state and initial rotation are something that should be looked at. Because of the failure of local search algorithms in this report the choice had reduced impact, but it is not unreasonable to assume that a good starting point may allow local optimizers to find good solutions even on more complex systems than H_2O.

A few methods to reduce the amount of qubits were looked at. Firstly an attempt was made to implement a Closed Shell Hamiltonian into the CQT VQE library. This turned out to be incompatible with the Jordan-Wigner transform. Secondly an the exhaustive simulation of all possible orbital selections for H_2O showed that selecting the correct orbitals is not trivial. The previously used scheme of picking the highest absolute value energy orbitals in turn was not great. The best result is gotten when the HF molecular orbitals that are closest to FCI molecular orbitals are chosen. While it is not clear how to determine this without either calculating the FCI solution or generating the Hamiltonian (both of which are no options for the preparation of VQE in reasonable time), it should be clear that results vary significantly with orbital choice, when the choice to remove orbitals is made. At the very least multiple orbital selections can be looked at in a real-life application where VQE is used as a faster alternative to FCI. This way the FCI solution does not have to be calculated, but bad orbital selection can be avoided.

Actual simulations of H_2O were done to see if improvements on the Hartree-Fock solution were achievable. Using optimally chosen orbitals, great results were gotten even for smaller approximated systems (8 and 10 qubits before projection). Both were fast enough to have the global search algorithm converge and ended up very close to the exact energy of the approximated system. An additional result that was taken from the smaller qubit systems was that in fact the amount of parameters of the rotation matrix may have to be higher than the dimension of the system for nice convergence, as very bad outliers were found after convergence of the global optimizer for entanglement depths corresponding with less parameters than the system itself. Even smaller systems can improve on Hartree-Fock if more time wants to be saved (or if an actual quantum processing unit is created with a small amount of qubits). The bigger systems (12 and 14 qubits before projection) required many iteration steps to get below the HF threshold. This could be seen clearly in the results as the differences between geometries was quite large at 2000 iteration steps.

To help convergence the structure of the exact solution (extracted from the Hamiltonian) was used to implement a strategy that restricted the search space of the optimization algorithm. This method added more weight to the Hartree-Fock basis state, without restricting the optimizer too much. This way the HF basis state was always a significant part of the vectors that were looked at, while leaving the rest to the optimizer. Using this

HFS trick, results were improved significantly. While reaching the FCI solution was still not achievable within this project, it is argued that this trick can help for bigger systems in general, as it is hypothesised that the HF-solution basis state will be the major component of the solution for bigger systems in general, not just for H_2O. Additionally the impact of the method was shown to be bigger for bigger systems. It is unclear how much the trick hinders convergence towards FCI, as the optimizer did not converge far enough in general, most likely due to too small of an entanglement depth for 12 and 14 orbitals. Still the lowest energies found with VQE in the project were found using HFS. Finally the limitation of the scheme was illustrated by its failure to get correct asymptotic behavior for the moving one H atom away system. This reemphasises that the assumption that the HF solution is an big part of the exact solution is very important. For decreasing importance of the HF solution the R-values should be adjusted accordingly. It is not clear if the scheme will have the same positive results on running times and solution under these conditions, when the R-value is picked correctly.

This thesis shows that a lot has to be done with respect to the optimization of VQE for bigger systems to reach FCI. While FCI is not a requirement, and the result gotten in this report for the 10 orbital system with $R = 4$ is a very nice result that almost reaches that threshold, no classical optimization method was found that could reach this threshold. While running times are dominated by the part that actual VQE would run on the quantum computer, the error when the optimizer detects convergence was very big a lot of the time. In fact cutting the optimizer off early gave the similar results (and strangely enough sometimes better).

10.2 Future Work

This project raises a lot of questions and there are a quite some areas that require further exploration. Most significantly the assumptions made to make simulating the quantum chip simple and straightforward are frankly unrealistic. These systems have technical issues such as gate errors and decoherence, which require systems to be optimized in regards to gate efficiency and error mitigation as well as the computational optimizations that were considered in this report. Even more basic things such as the required repeated measurements to get a good approximation for a singular trial state are not trivial and were ignored in this project.

Some unexplained results in this report that ask for additional exploration are for example the differences in convergence for the geometries of the 8 qubit system (6 after reduction). At entanglement depth 3 and 4 outliers could still be found far above the Hartree-Fock solution. It is very strange that the results are this bad, especially with the extra added information that for these entanglement depths the rotation matrix has more parameters than the state vectors themselves. An explanation could be looked at by analysing the specific trial states that the global optimizer uses. The issue is probably related to the space coverage of the entanglement matrix, which is already a known area of interest.

The SPSA algorithm was a failure in all the ways it was implemented in this project. It is hesitantly concluded that SPSA does not scale up well beyond the ≈ 30 parameters claimed in [8]. However because of its incredible convergence speed (as it is a local optimizer and not a global one), something that can be looked into is using some form of multi-start SPSA instead. It is not clear this will work on medium to large size systems as more involved selection of starting locations becomes required, combined with the fact that during this project SPSA failed to return to local minima it started in or near to. It seems very likely that the bad SPSA results are due to implementation errors, but none have been found.

Most other optimizers performed even worse and unless increasingly better initial states are found (with respect to system size) they will not perform on medium or large NISQ era systems either. The fmincon method, although used as a black-box, performed better but also showed signs of struggles with the system size. On the 12 orbital system for example, when let to determine by itself when it was converged it would take more than 10000 iteration steps to get a worse result than a previous simulation with only 2000 steps. There could be a good reason for this that is unknown due to the secret code, but the method should be deterministic so this makes little sense. It is therefore argued that for although classical computational chemistry can bail some of these optimizers out when accurate starting locations can be determined (with UCC for example), it is very important to design and research the classical optimizer part of VQE for when the systems get big.

The Hartree-Fock superposition trick seems very promising, but some work will have to be done to make it practical. Firstly a running time cost vs benefit analysis has to be done. While convergence is sped up in the sense that less iterations are needed for the optimizer, it will take a constant factor more computing time on the quantum circuit. The results of the method were good. For all regular systems the best result was gotten using the HFS. Not only did it improve the results for all these systems for all stopping criteria, it was also the

only reason the simulations of the 14 orbital system could be reduced below the HF threshold. The results are only for one optimizer though, so the HFS should be tested further on other optimization methods.

The failings of the method should also be investigated. While, in hindsight, such results could have been expected as the HF solution has the wrong asymptotic behavior, it is clear that the R-value has to be chosen smaller. This is reasonable to derive from the fact that even for the worst system considered the coefficient was still above 0.70, significantly larger than all others. It is interesting to know whether or not the method still performs well under these conditions provided that the correct R-value is chosen. One may expect the performance of the scheme to be worse, however the difference between $R = 0$ and $R = 1$ was quite significant with respect to running time so even small R-values could already yield a positive result.

Bibliography

[1] Peter W Shor. 'Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer'. In: *SIAM review* 41.2 (1999), pp. 303–332.

[2] Lov K Grover. 'Quantum mechanics helps in searching for a needle in a haystack'. In: *Physical review letters* 79.2 (1997), p. 325.

[3] Lov K Grover. 'Quantum computers can search arbitrarily large databases by a single query'. In: *Physical review letters* 79.23 (1997), p. 4709.

[4] Petar Jurcevic et al. 'Demonstration of quantum volume 64 on a superconducting quantum computing system'. In: *Quantum Science and Technology* 6.2 (2021), p. 025020.

[5] John Preskill. 'Quantum Computing in the NISQ era and beyond'. In: *Quantum* 2 (Aug. 2018), p. 79. ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79. URL: https://doi.org/10.22331/q-2018-08-06-79.

[6] Richard P Feynman. 'Simulating physics with computers'. In: *Feynman and computation*. CRC Press, 2018, pp. 133–153.

[7] Alberto Peruzzo et al. 'A variational eigenvalue solver on a photonic quantum processor'. In: *Nature communications* 5.1 (2014), pp. 1–7.

[8] Abhinav Kandala et al. 'Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets'. In: *Nature* 549.7671 (2017), pp. 242–246.

[9] Frank Arute et al. 'Hartree-Fock on a superconducting qubit quantum computer'. In: *Science* 369.6507 (2020), pp. 1084–1089.

[10] Teng Bian et al. 'Quantum computing methods for electronic states of the water molecule'. In: *Molecular Physics* 117.15-16 (2019), pp. 2069–2082. DOI: 10.1080/00268976.2019.1580392. eprint: https://doi.org/10.1080/00268976.2019.1580392. URL: https://doi.org/10.1080/00268976.2019.1580392.

[11] Yunseong Nam et al. 'Ground-state energy estimation of the water molecule on a trapped-ion quantum computer'. In: *npj Quantum Information* 6.1 (2020), pp. 1–6.

[12] David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Education Limited, 2014.

[13] Ian Glendinning. *Rotations on the Bloch sphere*. 2010.

[14] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[15] TMP Chem. *Hartree-Fock Theory Playlist*. 2018. URL: https://www.youtube.com/playlist?list=PLm8ZSArAXicIijiVIx0yfk2ZOK-16ycji.

[16] Max Born and Robert Oppenheimer. 'Zur quantentheorie der molekeln'. In: *Annalen der physik* 389.20 (1927), pp. 457–484.

[17] Arvi Rauk. 'Orbital interaction theory of organic chemistry'. In: John Wiley & Sons, 2004. Chap. Appendix A: A Derivation of Hartee-Fock Theory.

[18] S Francis Boys. 'Electronic wave functions-I. A general method of calculation for the stationary states of any molecular system'. In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 200.1063 (1950), pp. 542–554.

[19] Benjamin P Pritchard et al. 'New basis set exchange: An open, up-to-date resource for the molecular sciences community'. In: *Journal of chemical information and modeling* 59.11 (2019), pp. 4814–4820.

[20] L Brillouin. In: *Actualites Sci. et Ind.* 71 (1933).

[21] I Mayer. 'On the generalized Brillouin theorem for spin projected wave functions'. In: *Acta physica Academiae Scientiarum Hungaricae* 34.4 (1973), pp. 305–309.

[22] Péter R. Surján. 'The Brillouin Theorem'. In: *Second Quantized Approach to Quantum Chemistry: An Elementary Introduction*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 87–92. ISBN: 978-3-642-74755-7. DOI: 10.1007/978-3-642-74755-7_11. URL: https://doi.org/10.1007/978-3-642-74755-7_11.

[23] Michael A Nielsen et al. 'The Fermionic canonical commutation relations and the Jordan-Wigner transform'. In: *School of Physical Sciences The University of Queensland* 59 (2005).

[24] Pascual Jordan and Eugene P Wigner. 'About the Pauli exclusion principle'. In: *Z. Phys* 47.631 (1928), pp. 14–75.

[25] Jacob T Seeley, Martin J Richard and Peter J Love. 'The Bravyi-Kitaev transformation for quantum computation of electronic structure'. In: *The Journal of chemical physics* 137.22 (2012), p. 224109.

[26] Andrew Tranter, Peter J Love, Florian Mintert and Peter V Coveney. 'A comparison of the bravyi–kitaev and jordan–wigner transformations for the quantum simulation of quantum chemistry'. In: *Journal of chemical theory and computation* 14.11 (2018), pp. 5617–5630.

[27] Robert de Keijzer. 'Optimization of Variational Quantum Eigensolver'. 2019.

[28] R.J.P.T. de Keijzer, V.E. Colussi, B. Škorić and S.J.J.M.F. Kokkelmans. 'Optimization of the Variational Quantum Eigensolver for Quantum Chemistry Applications'. Submitted, not yet published. 2021.

[29] *Why is BeH2 Linear and H2O Bent?* [Online; accessed 2021-07-22]. 11th May 2021. URL: https://chem.libretexts.org/@go/page/13472.

[30] Saad Yalouz et al. 'A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states'. In: *Quantum Science and Technology* 6.2 (2021), p. 024004.

[31] Jasper Postema. 'Simulating the Hydrogen Molecule and Quantum Magnetism using a Variational Quantum Eigensolver'. 2018.

[32] Sergey Bravyi, Jay M Gambetta, Antonio Mezzacapo and Kristan Temme. 'Tapering off qubits to simulate fermionic Hamiltonians'. In: *arXiv preprint arXiv:1701.08213* (2017).

[33] Kanav Setia et al. 'Reducing qubit requirements for quantum simulations using molecular point group symmetries'. In: *Journal of Chemical Theory and Computation* 16.10 (2020), pp. 6091–6097.

[34] Jordan Cotler and Frank Wilczek. 'Quantum overlapping tomography'. In: *Physical review letters* 124.10 (2020), p. 100401.

[35] Jarrod R McClean, Jonathan Romero, Ryan Babbush and Alán Aspuru-Guzik. 'The theory of variational hybrid quantum-classical algorithms'. In: *New Journal of Physics* 18.2 (2016), p. 023023.

[36] Ikko Hamamura and Takashi Imamichi. 'Efficient evaluation of quantum observables using entangled measurements'. In: *npj Quantum Information* 6.1 (2020), pp. 1–8.

[37] Daochen Wang, Oscar Higgott and Stephen Brierley. 'Accelerated variational quantum eigensolver'. In: *Physical review letters* 122.14 (2019), p. 140504.

[38] James C Spall. 'A stochastic approximation technique for generating maximum likelihood parameter estimates'. In: *1987 American control conference*. IEEE. 1987, pp. 1161–1167.

[39] James C Spall et al. 'Multivariate stochastic approximation using a simultaneous perturbation gradient approximation'. In: *IEEE transactions on automatic control* 37.3 (1992), pp. 332–341.

[40] James C Spall. 'Implementation of the simultaneous perturbation algorithm for stochastic optimization'. In: *IEEE Transactions on aerospace and electronic systems* 34.3 (1998), pp. 817–823.

[41] John A Nelder and Roger Mead. 'A simplex method for function minimization'. In: *The computer journal* 7.4 (1965), pp. 308–313.

[42] Jeffrey C Lagarias, James A Reeds, Margaret H Wright and Paul E Wright. 'Convergence properties of the Nelder–Mead simplex method in low dimensions'. In: *SIAM Journal on optimization* 9.1 (1998), pp. 112–147.

[43] Nikolaj Moll et al. 'Quantum optimization using variational algorithms on near-term quantum devices'. In: *Quantum Science and Technology* 3.3 (2018), p. 030503.

[44] Lixing Han and Michael Neumann. 'Effect of dimensionality on the Nelder–Mead simplex method'. In: *Optimization Methods and Software* 21.1 (2006), pp. 1–16.

[45] Richard H Byrd, Mary E Hribar and Jorge Nocedal. 'An interior point algorithm for large-scale nonlinear programming'. In: *SIAM Journal on Optimization* 9.4 (1999), pp. 877–900.

[46] Richard H Byrd, Jean Charles Gilbert and Jorge Nocedal. 'A trust region method based on interior point techniques for nonlinear programming'. In: *Mathematical programming* 89.1 (2000), pp. 149–185.

[47] Richard A Waltz, José Luis Morales, Jorge Nocedal and Dominique Orban. 'An interior algorithm for nonlinear optimization that combines line search and trust region steps'. In: *Mathematical programming* 107.3 (2006), pp. 391–408.

[48] Maria Schuld et al. 'Evaluating analytic gradients on quantum hardware'. In: *Physical Review A* 99.3 (2019), p. 032331.

Appendices

Appendix A

Matrix Element Proof

Given two Slater determinants ϕ_i and ϕ_j, which have more than two different orbitals, it has to be shown that $\langle\phi_i|H|\phi_j\rangle = 0$. Assume without loss of generality that ϕ_i and ϕ_j differ by three orbitals:

$$|\phi_j\rangle = a_a^\dagger a_b a_c^\dagger a_d a_e^\dagger a_f |\phi_i\rangle \tag{A.1}$$

where a, b, c, d, e and f have to be different indices. Now look at the matrix element of interest in the second quantization:

$$\langle\phi_i|H|\phi_j\rangle = \sum_{ij}\langle i|h|j\rangle\langle\phi_i|a_i^\dagger a_j|\phi_j\rangle + \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle\langle\phi_i|a_i^\dagger a_j^\dagger a_l a_k|\phi_j\rangle \tag{A.2}$$

Next ϕ_j is substituted out:

$$\langle\phi_i|H|\phi_j\rangle = \sum_{ij}\langle i|h|j\rangle\langle\phi_i|a_i^\dagger a_j a_a^\dagger a_b a_c^\dagger a_d a_e^\dagger a_f|\phi_i\rangle + \frac{1}{2}\sum_{ijkl}\langle ij|kl\rangle\langle\phi_i|a_i^\dagger a_j^\dagger a_l a_k a_a^\dagger a_b a_c^\dagger a_d a_e^\dagger a_f|\phi_i\rangle \tag{A.3}$$

Here the problem shows up: there are not enough arbitrary orbitals $(i.j, k, l)$ to get non-zero terms. Indeed, the Hamiltonian only gives one pair in the one-electron term and two pairs in the two-electron term. This is where the more than two requirement comes from. Writing out all possible pairs to proof none work would the most straightforward method but it would also be a waste of time so instead the following argumentation is used. It is clear that the bra-ket terms in the above can only be either 1 or 0 due to orthogonality. Therefore the only way to make the term non-zero is to use the creation and annihilation operators in such a way that that both the bra and the ket are the same. Now here is the problem: there are 6 operators of different orbitals out of a total of 8 for the one-electron term and 10 for the two-electron term. It is impossible to have equivalent operations on both sides as after using the not fixed operators at least 2 operators that must be different from each other will be left behind. Therefore the bra and ket can never be equal and thus every term in the above will be zero. This directly gives the result needed.

Appendix B

Orbital Removal Figures

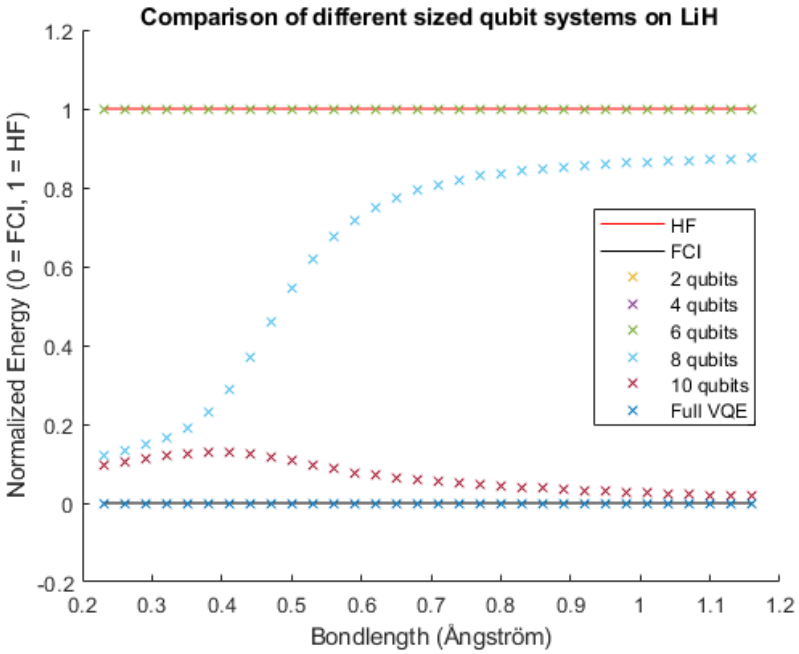


Figure B.1: Plot showing the difference in simulation result based on the amount of qubits used for LiH. The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately. The complete VQE solution is on the FCI line and the 2,4 and 6 qubit VQE solutions are on the HF line.

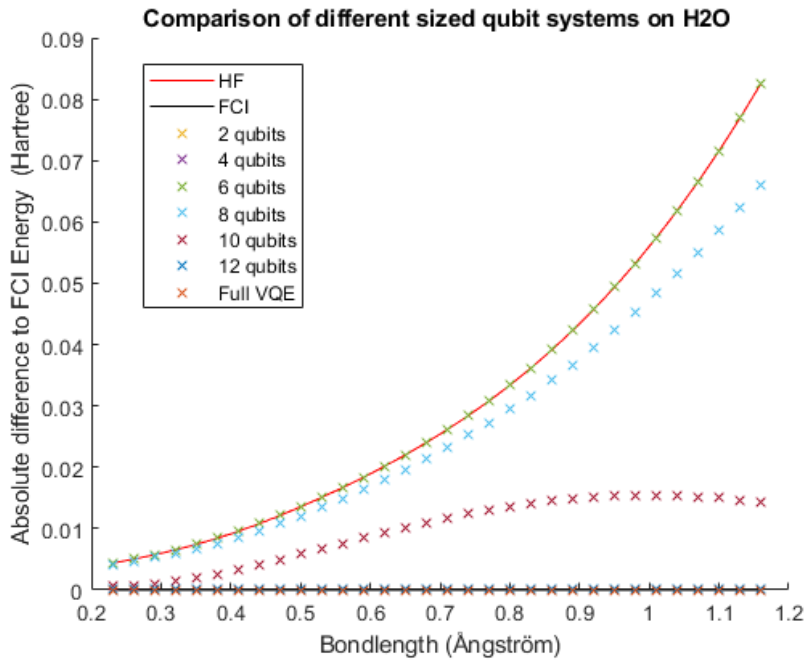


Figure B.2: Plot showing the difference in simulation result based on the amount of qubits used for H_2O. On the y-axis is the difference between the FCI solution and the methods as shown in the legend. Note that the 2,4 and 6 qubit VQE solutions are on the HF line. Additionally note that while 12 qubits is very close to the FCI solution it is in fact above it for all bond lengths, while complete (14 qubit) VQE is on the FCI line.

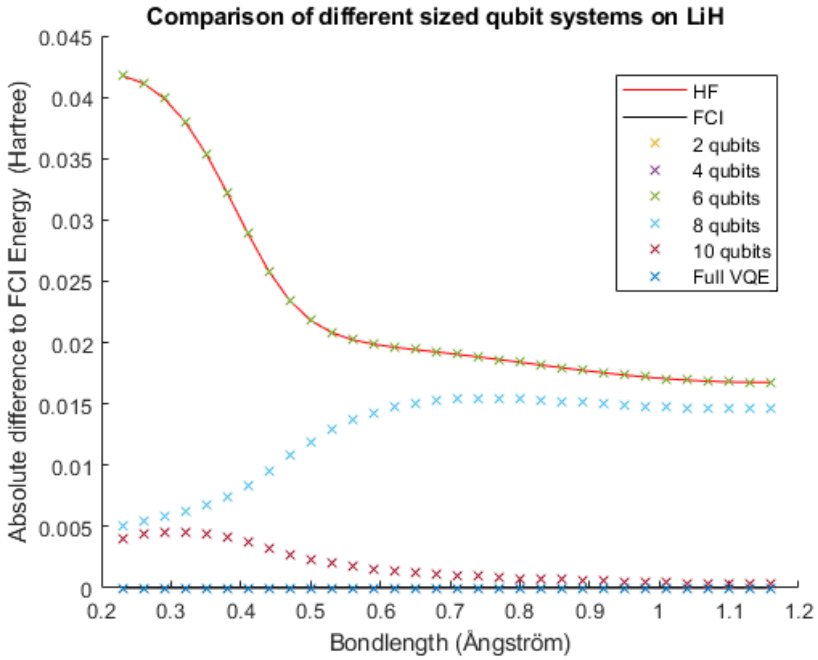


Figure B.3: Plot showing the difference in simulation result based on the amount of qubits used for LiH. On the y-axis is the difference between the FCI solution and the methods as shown in the legend. The complete VQE solution is on the FCI line and the 2,4 and 6 qubit VQE solutions are on the HF line.

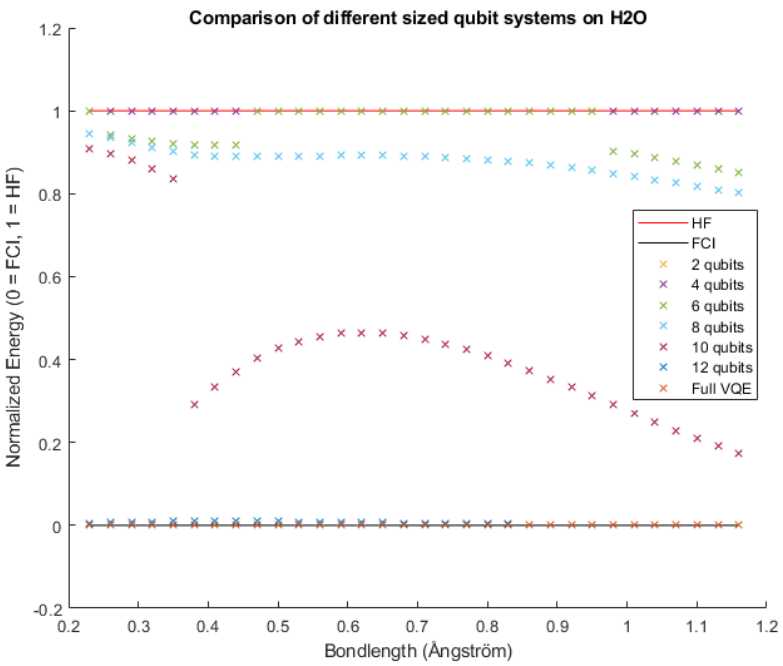


Figure B.4: Plot showing the difference in simulation result based on the amount of qubits used for H_2O. On the y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately. The complete VQE solution is on the FCI line and 2,4 and part of 6 orbital VQE solutions are on the HF line. Of special note is the discontinuous behavior for the systems where different geometries had different orbital selection (for example 6 and 10)

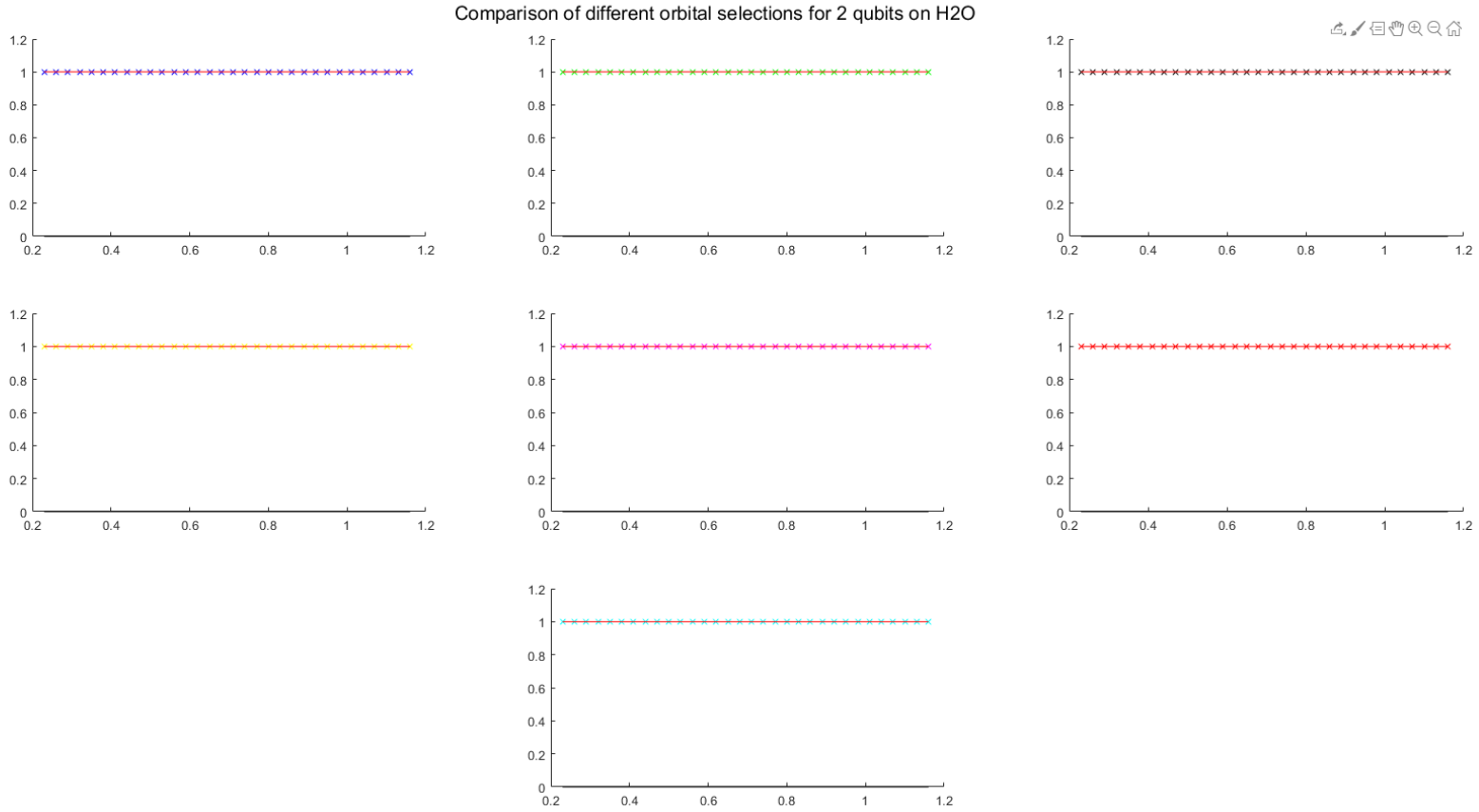


Figure B.5: Plot showing for each out of the seven orbitals the minimal eigenvalue of the Hamiltonian, whether made with that specific orbital (marked) or not (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.

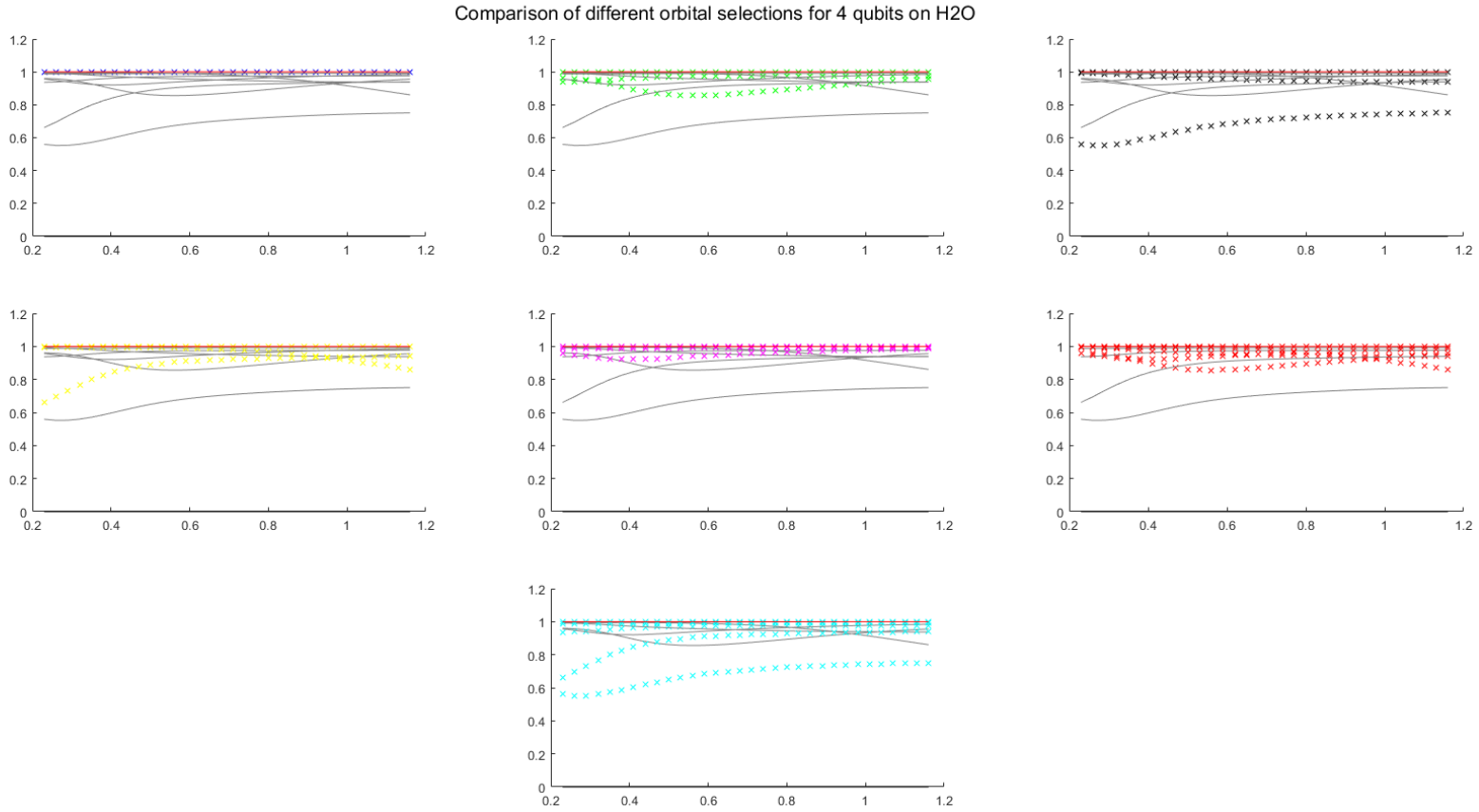


Figure B.6: Plot showing for each combination of two out of the seven orbitals the minimal eigenvalue of the Hamiltonian, whether made with a combination of orbitals that specific orbital is in (marked) or with a combination of orbitals it is not in (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.

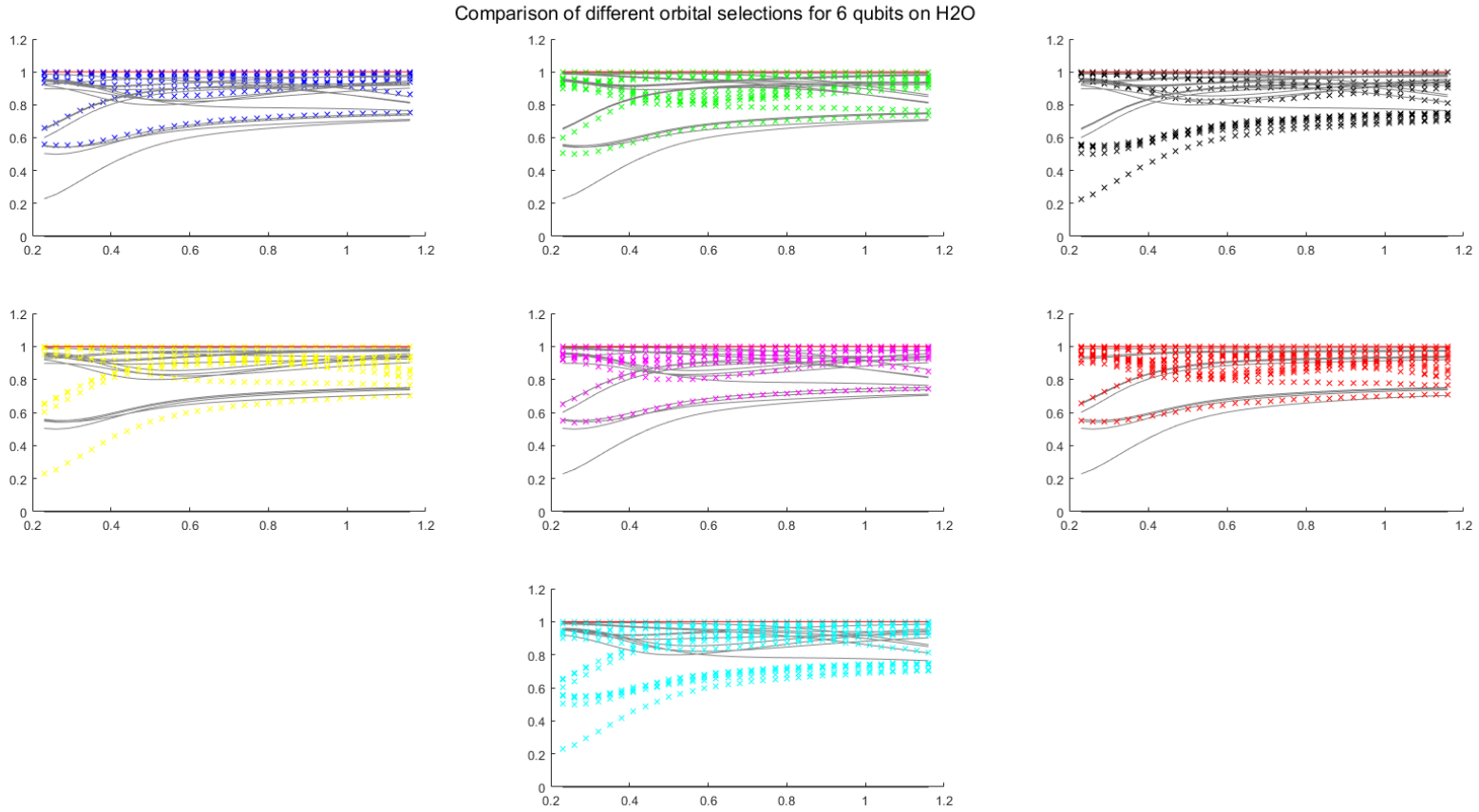


Figure B.7: Plot showing for each combination of three out of the seven orbitals the minimal eigenvalue of the Hamiltonian, whether made with a combination of orbitals that specific orbital is in (marked) or with a combination of orbitals it is not in (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.

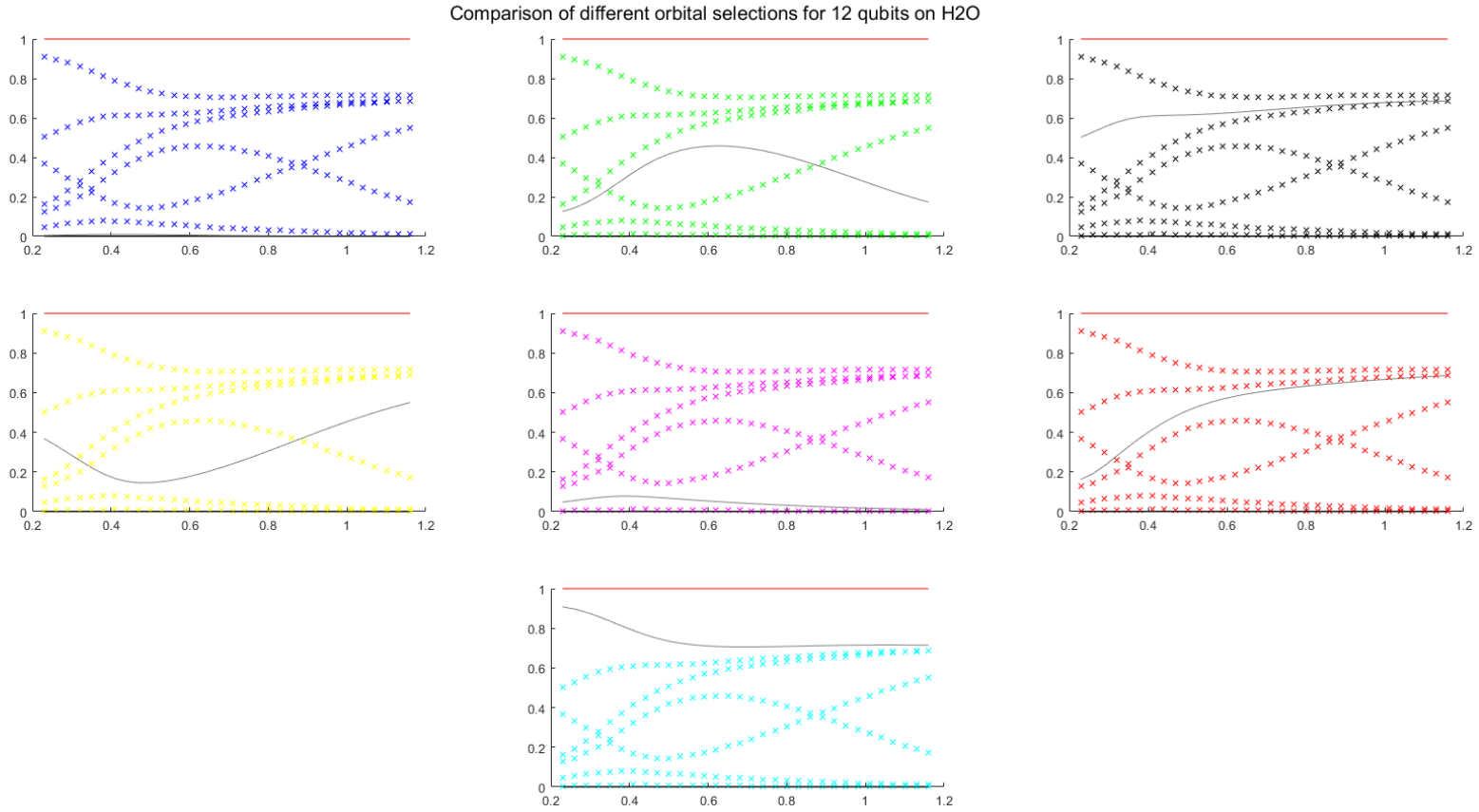


Figure B.8: Plot showing for each combination of six out of the seven orbitals the minimal eigenvalue of the Hamiltonian, whether made with a combination of orbitals that specific orbital is in (marked) or with a combination of orbitals it is not in (grey line). The y-axis is normalized such that the FCI solution equals 0 and the HF solution equals 1, all simulation results have been normalized appropriately.