1 2 9 0

UNIVERSIDADE Ð COIMBRA

Miguel Eduardo de Vasconcelos Morais Murça

MOLECULAR GEOMETRY CALCULATIONS USING A NOVEL QUANTUM VARIATIONAL APPROACH

Julho de 2020

A sincere thank you to Professor José Ignácio Latorre, for accepting me for an internship in QUANTIC, for agreeing to orient my master's thesis, and without whom this work would not be possible.

My deep thanks also to Professor Helena Vieira Alberto, not only for also agreeing to orient my master's thesis, but also for her immense personal support, willingness and availability throughout the master's course. Her technical guidance during the master's course was fundamental in helping me getting started with quantum computing.

Thank you to Adrián Pérez, Sergi Ramos, Carlos Prieto and Diego García of the QUANTIC team for their technical advice, discussions, patience and making me feel welcome at QUANTIC and in Barcelona.

Thank you to Professor Orlando Oliveira and Professor Fernando Nogueira for their availability to help, and support on some theoretical aspects of this thesis.

Thank you to my colleagues at the University of Coimbra for enduring questions about quantum computing and computational electronic structure, even when directed at no one in particular. A special thanks to Diogo Almeida, who always responded with thoughtful feedback and relevant material, and probably endured more questions than average.

Thank you to BiF, it is rare to find a group of people that would find a fake scientific poster funny.

Thank you to Chanise, who made Barcelona a lot more fun.

Thank you to my family and Mariana, for everything else.

Contents

Summary

The work in this thesis bridges two separate lines of previous work: on one hand, it stems from the work of the QUANTIC group, of the University of Barcelona[1], who actively research on quantum computing, and with whom part of the work developed in the context of this thesis was done, under an Erasmus internship. The recent work of the group includes Quantum Variational Algorithms (QVAs) and applications of QVAs and generally quantum computing to different problems. On the other hand, the 2008 publication by Bravyi, DiVincenzo, Loss and Terhal was a starting point for the development of this thesis's work; the methodology therein presented is fundamental in achieving the goal for this thesis: to develop a new, quantum device oriented, method for obtaining the geometric parameters of a molecule (or otherwise physical parameters) that result in the lowest possible energy, and that has low quantum computational requirements (in qubit quality and number).

In the context of this goal, a significant part of the thesis's work effort was dedicated to building a quantum circuit simulator from scratch, to explore theoretical and technical bottlenecks in a "full-stack" approach to simulating Quantum Variational Algorithms.

In section 1 we present material which constitutes the background of the thesis's work, namely the Born-Oppenheimer approximation and second-quantization of a molecular Hamiltonian (section 1.1), and how we may then translate such a second-quantized Hamiltonian into a form that can be evaluated using a quantum computer (section 1.2); Quantum Variational Algorithms and their advantages in a Noisy Intermediate-Scale Quantum (NISQ) regime (section 1.3), and finally the Schrieffer-Wolff transformation (section 1.4).

Sections 3 and 4 form the core of this thesis's work, corresponding to the development "from scratch" of a C/Python library to simulate quantum circuits and quantum variational algorithms (named QOP; section 3) and the elaboration of an original method for molecular geometric parameter calculation, or generally minimal energy parameter determination for some parameterized Hamiltonian (section 4). In section 5, the issue of imposing constraints when only addressing parts of a larger system is discussed, in the context of the quantum variational aspect of the technique proposed in section 4, showing that ignoring electron count constrains on ground state candidates establishes a lower bound energy to the ground state energy.

Finally, we apply the developed work to a few selected systems (H_2, HLi, O_2), presenting and discussing the obtained results in, respectively, sections 6 and 7, where we show that the technique proposed to calculate energy minimizing bond lengths is successful in some test cases, but may fail due to either the way in which the locality of the Hamiltonian is reduced or due to the quantum variational process. Despite these shortcomings, we obtain, using the technique, bond lengths comparable to those obtained using a Hartree-Fock approach.

---

[1] `http://quantic.bsc.es`

Sumário

O trabalho apresentado nesta tese surge como uma extensão de duas linhas separadas de trabalho: por um lado, advém do trabalho prévio do grupo QUANTIC, da Universidade de Barcelona[1], onde parte do trabalho de tese aqui apresentado foi desenvolvido, sob um estágio Erasmus. O grupo QUANTIC tem como área principal de investigação a computação quântica; dos seus trabalhos recentes constam aplicações de Algoritmos Quânticos Variacionais (*Quantum Variational Algorithms*, QVAs) e geralmente computação quântica a diferentes problemas. Por outro lado, a publicação de 2008 por Bravyi, DiVincenzo, Loss e Terhal foi instrumental na elaboração do trabalho aqui apresentado, permitindo concretizar o objectivo proposto: construir um método para o cálculo de geometria molecular que minimize a energia, com requerimentos quânticos computacionais mínimos, tanto em termos da qualidade dos *qubits* utilizados, como em termos do número de *qubits* necessários.

No contexto desse objectivo, uma parte substancial do trabalho desenvolvido nesta tese foi dedicado à construção de um simulador de circuitos quânticos, a fim de explorar as dificuldades teóricas e técnicas inerentes ao desenvolvimento de uma simulação total de um algoritmo quântico variacional.

Na secção 1 é apresentado o material teórico de base ao trabalho desenvolvido nesta tese, nomeadamente: a aproximação de Born-Oppenheimer e a segunda quantização de um Hamiltoniano molecular (secção 1.1), a transformação deste Hamiltoniano para uma forma favorável à utilização de um computador quântico (secção 1.2), Algoritmos Quânticos Variacionais e as suas vantagens no regime atual de computação quântica (*Noisy Intermediate-Scale Quantum*, NISQ; secção 1.3), e finalmente a transformada de Schrieffer-Wolff (secção 1.4).

As secções 3 e 4 formam o trabalho nuclear desta tese: na secção 3 apresenta-se o desenvolvimento "de raíz" de uma biblioteca em C/Python de simulação de circuitos quânticos e algoritmos quânticos variacionais. A biblioteca foi denominada QOP, como acrónimo de **Quantum OP**timizer. Na secção 4 apresenta-se um método original para o cálculo de geometria molecular que minimiza a energia (ou, mais genericamente, dos parâmetros que minimizam a energia de um Hamiltoniano parametrizado). Na secção 5 são discutidas as implicações de constrangimentos no número de electrões impostos sobre o estado electrónico total, aquando de um tratamento apenas parcial do estado electrónico. Verifica-se que ignorar estes constrangimentos poderá produzir estados candidatos ao estado fundamental não-físicos, mas que a energia a eles associada estabelece um mínimo para a "verdadeira" energia de estado fundamental.

Finalmente, o trabalho desenvolvido é aplicado a sistemas moleculares de teste, nomeadamente H_2, HLi e O_2, sendo os resultados obtidos para estes sistemas apresentados e discutidos, respetivamente, nas secções 6 e 7. Observa-se que a técnica proposta na seccção 4 é bem sucedida para alguns dos sistemas considerados. Verifica-se também que não é por vezes possível obter um comprimento de ligação molecular, podendo-se isso relacionar com o processo quântico variacional, mas também com o processo pelo qual se reduz a localidade dos Hamiltonianos considerados. Ainda assim, obtêm-se, pelo processo original proposto, comprimentos de ligação comparáveis aos obtidos com um tratamento Hartree-Fock para vários sistemas.

---

[1] http://quantic.bsc.es

1 Introduction

John Preskill, a leading figure in the field of Quantum Computing, coined both the terms "Quantum Supremacy" [28] and "Noisy Intermediate-Scale Quantum" [41]. These two terms succinctly represent two of the main aspects of present-day quantum computing: on the one hand, the promise of quantum speedups and efficient representation of an exponentially large Hilbert space motivate efforts towards building practical implementations of quantum computers. It is to the milestone of building a quantum computer beyond classical simulation[2] that the name 'quantum supremacy' was given. On the other hand, this supremacy is not to be achieved easily, as quantum computers effectively operate on delicate physical phenomena, in particular entanglement. A careful balance between isolation and control of the system must be maintained, as external perturbations may collapse the quantum state — which is undesirable — but the state must be 'readable', so that we should be able to collapse it when intended. By virtue of these technical difficulties, the number and quality of qubits in current physical realizations of quantum computers is still a concern, and must be taken into account; but there is a sufficient number of qubits to perform quantum computation beyond a simple proof-of-concept. It is because of this that Preskill suggests we are in the 'Noisy Intermediate-Scale' era of quantum computing. In this thesis, we propose and test a novel method of quantum computing applied to computational physics that seeks to take into account both of these aspects of current-era quantum computing: to exploit the natural applications of quantum computers to the treatment of quantum systems, while maintaining quantum computational requirements at an extreme minimum — not only in terms of the quality requirements of the qubits, but also in the number of qubits.

1.1 Molecular Hamiltonian: Second Quantization

In order to determine, for a molecular system, the geometric parameters for which the system's energy is minimal (which we shall refer to, henceforth, as the "at-rest" parameters), we must first describe the interactions of the system. Considering $\vec{r}_i$ the positions of the electrons ($i = 1, 2, \ldots, n_e$) and $\vec{R}_j$ the positions of the nuclei ($j = 1, 2, \ldots, n_R$), then such a Hamiltonian

---

[2]'Beyond simulation' is arguably loosely defined. However, the exponential growth of the space encoded by quantum computers of polynomially more qubits allows us to draw a line in terms of memory: once parameterizing the state space requires more classical values than atoms in the universe, we may be confident that the memory requirements are physically intractable.

$$H = -\sum_i \frac{\nabla^2_{R_i}}{2M_i} - \sum_i \frac{\nabla^2_{r_i}}{2} - \sum_{ij} \frac{Z_i}{|R_i - \hat{r}_j|} + \sum_{i,j>i} \frac{Z_i Z_j}{|R_i - R_j|} + \sum_{i,j>i} \frac{1}{|\hat{r}_i - \hat{r}_j|}$$

accounts for the kinetic energy of the electrons and the nuclei, and the Coulomb interactions between each other (in natural units). An analytical solution for the ground state (energy) of such a Hamiltonian is rarely available [4], and so we seek to tackle the problem with various approximations and/or numerical schemes. In particular, consider first the Born-Oppenheimer approximation [8]: we consider that, because the nuclei are much more massive than the electrons, the latter will rapidly "relax" to the lowest energy configuration for a given set of nucleic positions $\{\vec{R}_i\}$, and so we may separate the nuclear-only related terms from the rest of the Hamiltonian...

$$H = H_{\text{Nuclear}}(\vec{R}_j) + H_{\text{Electronic}}(\vec{r}_i, \vec{R}_j)$$

...with the nuclear positions now appearing in the remaining $H_{\text{Electronic}}$ operator as parameters, and likewise factor the wavefunction...

$$\Psi = \psi_{\text{Nuclear}} \psi_{\text{Electronic}}$$

...such that within a small first order correction, we may obtain Ψ by solving

$$H_{\text{Electronic}}(\vec{R}_j)\psi_{\text{Electronic}} = E_e(\vec{R}_j)\psi_{\text{Electronic}}$$

$$\left(H_{\text{Nuclear}}(\vec{R}_j) + E_e(\vec{R}_j)\right)\psi_{\text{Nuclear}} = E\psi_{\text{Nuclear}}$$

Once $E_e(\vec{R}_j)$ is known for any $\{\vec{R}_j\}$, an energy surface of dimensions $3 \times n_R$ is well defined, with which the motion of the nuclei may be approximated [3]. We will disregard nuclear dynamics, and focus on the electronic energy $E_e(\{\vec{R}_j\})$ (with the associated nucleus-nucleus Coulomb interaction included); the minimizing $\{\vec{R}_j\}$ are regarded as the at-rest parameters within this approximation. As such, H henceforth denotes the Hamiltonian except for the nuclear kinetic energy term.

If we are now to treat this Hamiltonian numerically, we must choose a basis. Consider a complete set of orthonormal spin-orbitals $\{|\chi_i\rangle\}$ and single-electron spin-orbital coordinates $\sigma = \vec{r} \otimes s$, with $\langle \sigma | \chi_i \rangle = \chi_i(\sigma)$ known. Then,

$$H = \sum_{pqrs} \int d\sigma_1 \, d\sigma_2 \, d\sigma_3 \, d\sigma_4 \, |\chi_p \chi_q\rangle \, h_{pqrs}(\sigma_{1...4}) \, \langle \chi_r \chi_s|$$

$$h_{pqrs}(\sigma_{1...4}) = \chi_p^*(\sigma_4)\chi_q^*(\sigma_2) \langle \sigma_4 \sigma_2 | H | \sigma_1 \sigma_3 \rangle \chi_r(\sigma_1)\chi_s(\sigma_3)$$

which, when taking into account the orthogonality of $\{\chi_i\}, \{\sigma\}$, and separating one- and two-body terms[3], becomes

---

[3]For example, for the electron-nucleus term $\hat{h}_{ij} = \frac{Z_i}{|R_i - \hat{r}_j|}$:

$$\hat{h}_{ij} = \sum_{pqrs} |\chi_p\rangle |\chi_q\rangle \langle \chi_q \chi_p | \hat{h}_{ij} | \chi_r \chi_s \rangle \langle \chi_s | \langle \chi_r | =$$

$$\sum_{pqrs} \int d\sigma_1 \, d\sigma_2 \, |\chi_p\rangle |\chi_q\rangle \langle \chi_q | \langle \chi_p | \sigma_1 \rangle \langle \sigma_1 | \hat{h}_{ij} | \sigma_2 \rangle \langle \sigma_2 | \chi_r \rangle |\chi_s\rangle \langle \chi_s | \langle \chi_r | =$$

$$\sum_{pr} \int d\sigma_1 \, d\sigma_2 \, |\chi_p\rangle \langle \chi_p | \sigma_1 \rangle \langle \sigma_1 | \sigma_2 \rangle \frac{Z_i}{|R_i - r|} \langle \sigma_2 | \chi_r \rangle \langle \chi_r |$$

$$H = H_{\text{e.static}} + \sum_{pq} |\chi_p\rangle h_{pq} \langle\chi_q| + \frac{1}{2} \sum_{pqrs} |\chi_p\rangle |\chi_q\rangle h_{pqrs} \langle\chi_r| \langle\chi_s| \tag{1}$$

$$h_{pq} = \int d\sigma\, \chi_p^*(\sigma) \left(-\frac{1}{2}\nabla_r^2 - \sum_j \frac{Z_j}{|r - R_j|} \right) \chi_q(\sigma) \tag{2}$$

$$h_{pqrs} = \int d\sigma_1\, d\sigma_2\, \chi_p^*(\sigma_1)\chi_q^*(\sigma_2) \frac{1}{|r_1 - r_2|} \chi_r(\sigma_1)\chi_s(\sigma_2) \tag{3}$$

with $H_{\text{e.static}}$ the nuclei Coulomb repulsion terms, that can be calculated directly from the nuclei position parameters (and so act as a scalar term).

We seek to construct the many-electron wavefunction, but we cannot simply construct any n_e-body wavefunction from the χ orbitals, as this would not satisfy the fermionic nature of electrons, by which the total wavefunction must be antisymmetric. Rather, we may employ Slater determinants and superpositions thereof to represent different electronic configurations and construct a valid wavefunction:

$$|\Psi\rangle = \sum_{\substack{i \\ \text{in configs.}}} c_i |\chi_{i_1} \cdots \chi_{i_{n_e}}\rangle \qquad , c_i \in \mathbb{C} \tag{4}$$

$$\psi_i(\sigma_1 \cdots \sigma_{n_e}) = \langle\sigma_1 \cdots \sigma_{n_e}|\chi_{i_1} \cdots \chi_{i_{n_e}}\rangle = \begin{vmatrix} \chi_{i_1}(\sigma_1) & \chi_{i_2}(\sigma_1) & \cdots & \chi_{i_{n_e}}(\sigma_1) \\ \chi_{i_1}(\sigma_2) & \chi_{i_2}(\sigma_2) & \cdots & \chi_{i_{n_e}}(\sigma_2) \\ \vdots & \vdots & & \vdots \\ \chi_{i_1}(\sigma_{n_e}) & \chi_{i_2}(\sigma_{n_e}) & \cdots & \chi_{i_{n_e}}(\sigma_{n_e}) \end{vmatrix}$$

If we consider every possible configuration (obeying the Exclusion Principle), we refer to "Full Configuration Interaction" (FCI). If we consider only the Slater determinant of lowest energy orbitals, we refer to "Hartree-Fock" (HF).

Calculating the expected value (i.e., the energy) of such a state "undoes" the projection of H over the orbitals, such that we must calculate the integral (for a general Configuration Interaction state)

$$E = \sum_{ij} \iiint \mathcal{D}\sigma\, \mathcal{D}\sigma'\, c_i c_j^* \psi_i(\sigma_1, \sigma_2, \ldots)\psi_j^*(\sigma_1, \sigma_2, \ldots)H(\sigma_1, \sigma_1')$$

$$(\mathcal{D}\sigma = d\sigma_1\, d\sigma_2 \cdots d\sigma_{n_e})$$

Note, however, the form of eq. (1). There is a projection along a set or orbitals, resulting in a component along a new set of orbitals. This intuitively motivates a rephrasing in terms of a second quantized language:

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s \tag{5}$$

(Note that $\hat{r}_j$ is evaluated as in the space component of σ)

$$\hat{h}_{ij} = \sum_{pq} |\chi_p\rangle \tilde{h}_{pq} \langle\chi_q| \; ; \quad \tilde{h}_{pq} = \int d\sigma\, \chi_p^*(\sigma) \frac{Z_i}{|R_i - r|} \chi_r(\sigma)$$

where integrals eqs. (2) and (3) are unchanged. We introduced the fermionic creation/annihilation operators for each orbital...

$$\{a_i, a_j\} = \left\{a_k^\dagger, a_l^\dagger\right\} = 0$$

$$\left\{a_i, a_j^\dagger\right\} = \delta_{ij}$$

...and define a vacuum $|\rangle = |00\cdots\rangle$ in a Fock Space that describes the occupation of each orbital $|[n_\alpha]\rangle = |n_1 n_2 \cdots\rangle$

$$a_\alpha |\cdots n_\alpha \cdots\rangle = \begin{cases} 0 & \text{if} \quad n_\alpha = 0 \\ |\cdots 0 \cdots\rangle & \text{if} \quad n_\alpha = 1 \end{cases}$$

$$a_\alpha^\dagger |\cdots n_\alpha \cdots\rangle = \begin{cases} |\cdots 1 \cdots\rangle & \text{if} \quad n_\alpha = 0 \\ 0 & \text{if} \quad n_\alpha = 1 \end{cases}$$

1.2 Evaluating a Hamiltonian in a Quantum Computer

Any second quantized Hamiltonian can be systematically mapped to a Pauli operator based form, using a Jordan-Wigner transformation [6]:

$$
\begin{aligned}
a_j^\dagger &\longmapsto \prod_{k=1}^{j-1}(-\sigma_k^z)(\sigma_j^x + i\sigma_j^y)/2 \\
a_j &\longmapsto \prod_{k=1}^{j-1}(-\sigma_k^z)(\sigma_j^x - i\sigma_j^y)/2
\end{aligned}
\tag{6}
$$

The new space of spins very directly relates to the original Fock space, as can be seen from the (spin) ladder operators' inverse mapping...

$$\sigma_j^+ = e^{\left(-i\pi \sum_{k=1}^{j-1} a_k^\dagger a_k\right)} a_j^\dagger$$

$$\sigma_j^- = e^{\left(+i\pi \sum_{k=1}^{j-1} a_k^\dagger a_k\right)} a_j$$

...such that if we employ the computational basis, i.e., take the eigenvectors of the Pauli Z operator as the reference basis with $\{|\downarrow\rangle, |\uparrow\rangle\} \rightsquigarrow \{|0\rangle, |1\rangle\}$, the collective spin state may once again be regarded as a Fock space of orbital occupations.

On the other hand, Pauli operators are "natural" operations on qubits, such that a Hamiltonian composed of Pauli operators can be efficiently evaluated with respect to the expectation value over a previously encoded state. Consider, as an example, the following example of a 1st-neighbour-interacting XXZ Heisenberg system of N bodies:

$$H = \sum_{j=0}^{N-1} \left\{ \sigma_j^x \sigma_{(j+1)\%N}^x + \sigma_j^y \sigma_{(j+1)\%N}^y + \frac{1}{2}\sigma_j^z \sigma_{(j+1)\%N}^z \right\}$$

(where $i\%j$ denotes i modulus j.)

4
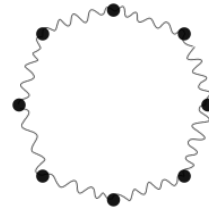
Provided there is an oracle U_ϕ that prepares $|\phi\rangle$, whose energy we are interested in calculating:

$$|0\rangle^{\otimes N} \quad \underline{}\!/^N\underline{}\boxed{U_\phi}\underline{} \quad |\phi\rangle$$

we evaluate the terms in H, by performing measurements in the following circuits until the expectation value of each term is known within error ϵ. [4]

Once $\langle\phi|\sigma_j^u \sigma_{(j+1)\%N}^u|\phi\rangle$ is known, $\langle\phi|H|\phi\rangle$ may be calculated by taking

$$\sum_{j=0}^{N-1}\left\{ \langle\phi|\sigma_j^x \sigma_{(j+1)\%N}^x|\phi\rangle + \langle\phi|\sigma_j^y \sigma_{(j+1)\%N}^y|\phi\rangle + \frac{1}{2}\langle\phi|\sigma_j^z \sigma_{(j+1)\%N}^z|\phi\rangle \right\}$$

Multiple circuits and shots (i.e. reruns of a circuit) are needed to determine the expected value of the Hamiltonian. In order to quantify this overhead, consider a single circuit to be repeatedly ran. For each run of the circuit, N z-basis measurements are performe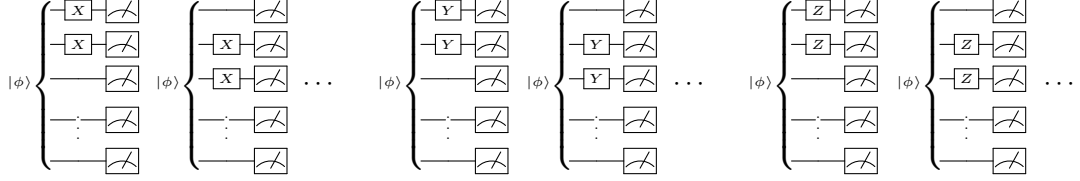d simultaneously; we are interested in the product of all measurements, which has a probability distribution we denote by Z.

Z has support $S_Z = \{-1, +1\}$. A remapping of S_Z into $\{0, 1\}$ yields the Bernoulli distribution Y,

$$Y = (Z+1)/2 \sim \mathcal{B}(p)$$

with p the odds of observing $Z = +1$. Because Y is a Bernoulli distribution, p is also the expected value of the Hamiltonian under the encoded state.

Under n shots, we can also define the stochastic variable W describing how many times $(+1)$ was observed:

$$W = \sum_{i}^{n} Y_i$$

This is a binomial distribution with support $S_W = \{0, 1, \dots, n\}$,

$$W \sim \mathcal{B}(n, p)$$

$$P(W = w) = \binom{n}{k} p^w (1-p)^{w-1}$$

and as such, the mean of W is np and its variance is $np(1-p)$ [21, pp. 48-53]. From here we may extract p within some confidence interval, and although more sophisticated approaches

---

[4] A better choice of circuits to measure can be made, namely by considering commuting terms in the Hamiltonian and otherwise employing techniques that "condense" the number of circuits.

exist[5], we will assume that n is large enough to take the Central Limit Theorem, approximating the distribution of errors with a normal distribution. In this case, the estimator of p is

$$\widehat{p} = x/n$$

<center>(x the number of (+1) measurements)</center>

the confidence interval ($p \in [\widehat{p} - \delta, \widehat{p} + \delta]$) is given by [48]

$$\delta = \frac{z}{n\sqrt{n}}\sqrt{x(n-x)}$$

where z is given by $1 - \alpha/2$, α being the error rate.

We wish to perform enough shots so that δ is bounded by some ϵ, so

$$\delta < \epsilon \Rightarrow \epsilon^2 > \frac{z^2}{n^3}x(n-x)$$

and the right-hand term of the inequation is maximized by $x = n/2$. Thus, in the worst-case scenario, we are looking for n such that

$$\epsilon^2 > \frac{z^2}{n^3}\frac{n}{2}\left(n - \frac{n}{2}\right) = \frac{z^2}{4}$$

Therefore, with $n = \lceil \frac{z^2}{4\epsilon^2}\rceil + 1$ we ensure that $\delta < \epsilon$.

The main takeaway from this is that with a polynomial (on the error bound) number of runs we may obtain a result which is "precise" ($\delta < \epsilon$). On the other hand we will at most want to evaluate the expected value for k circuits, with k being the number of terms in the Hamiltonian, and this corresponds to a fixed factor on the number of runs. So, provided each circuit is "efficient" at constructing $|\phi\rangle$, we see that the overhead of evaluating $\langle\phi|H|\phi\rangle$ is still polynomial and does not increase the algorithm's complexity.

One should also note that, in practical terms, this overhead does not necessarily imply loss of quantum advantage, as "native" quantum operations, i.e. the applications of quantum gates in quantum circuits, are extremely fast (in the order of tens of picoseconds), even in the current realizations of quantum computers. See, for a concrete example of this, reference [44].

1.3 Quantum Variational Algorithms

We are, thus, guaranteed to be able to evaluate the expectation value of H over a state previously encoded into the quantum computer. However, the circuit must be run many times, and so the state must be repeatedly encoded into it. If we are to perform the energy evaluation efficiently, then this repeated encoding must also be efficient. Moreover, this repeated encoding is necessary to evaluate the energy of a single state, but we are interested in *determining* the ground state.

If one had access to arbitrary quantum resources, one could employ the Phase Estimation Algorithm (PEA); the PEA routine evaluates the eigenvalues of an operator, and thus permits bypassing energy evaluation of different states and rather directly arriving at the eigenvalues of the relevant Hamiltonian. A visual summary of this algorithm is presented in fig. 1, and a more complete discussion of the algorithm is presented in appendix A.

However, the current Noisy Intermediate-Scale Quantum (NISQ) regime of quantum computation typically forbids the use of QPE due to limited relaxation and coherence times (see, e.g., [34]). Thus, we may turn to Quantum Variational Algorithms (QVAs) [40], by which we

---

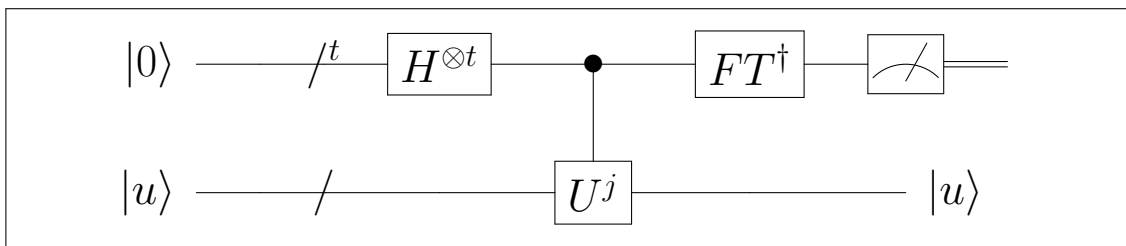[5]By taking into account unobserved measurements using a Bayesian approach.[31]

Figure 1: Circuit schematic reflecting the Quantum Phase Estimation algorithm for an operator U. The first register is a bundle of t wires, and the bottom register has as many wires as dictated by U. H denotes the Hadamard gate. U^j denotes 2^j applications of U controlled by the $(t-j)^{\text{th}}$ wire in the first register (as indexed top-to-bottom), with an implicit run of $j = 0 \ldots t - 1$. $FT^\dagger$ denotes the inverse Quantum Fourier Transform. $|u\rangle$ is generally taken to be an eigenvector of U but need not be, the consequences of which are discussed in appendix A.

define some oracle to prepare a candidate ground state, and then (classically) evolve the oracle's parameters based on the energy evaluation of the candidate.

This requires a definition for such an oracle, referred to as an *Ansatz*, as the oracle effectively "proposes" states for the ground state (fig. 2).
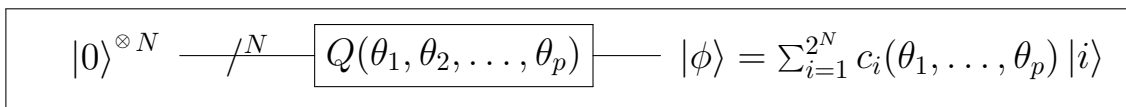
Figure 2: Quantum circuit diagram of an *Ansatz*, to be used in a Quantum Variational Algorithm scheme. The gate depends on p classical parameters, and produces an N-qubit quantum state.

Note that per the Jordan-Wigner transformation, the state of each qubit as expressed in the computational basis $\{|0\rangle, |1\rangle\}$ can be straightforwardly related to the Fock space; recall that

$$a_j^\dagger \longmapsto \prod_{k=1}^{j-1} (-\sigma_k^z)(\sigma_j^x + i\sigma_j^y)/2$$

$$a_j \longmapsto \prod_{k=1}^{j-1} (-\sigma_k^z)(\sigma_j^x - i\sigma_j^y)/2$$

As we are considering a system of fermions, any state of N qubits maps to a valid state of the Fock space of N orbitals. Furthermore, the Hilbert space of N qubits fully spans the possible Jordan-Wigner mappings of any state in the Fock state of N orbitals, as the occupation of each orbital can only be 0 or 1 for each orbital. Further physical restrictions may be imposed, so that not all of the Fock states are valid: this is for now ignored, but discussed in section 5.

We may then consider examples of *Ansätze* for a generic spin system.

A good *Ansatz* will, of course, cover the ground state, i.e., have at least a set of parameters that produces the ground state. If the *Ansatz* does *not* cover the ground state, by minimizing the expected value of the Hamiltonian for the given *Ansatz*

$$\langle H \rangle_{\vec{\theta}} = \langle 0 |^{\otimes N} Q^\dagger(\vec{\theta}) H Q(\vec{\theta}) | 0 \rangle^{\otimes N}$$

7

we obtain the state the maximally overlaps the ground state of H, within the span of the *Ansatz*. To see this, consider the spectral decomposition of H

$$H = \sum_{\lambda_i} \lambda_i \, |\lambda_i\rangle\!\langle\lambda_i|$$

$$H \, |\lambda_i\rangle = \lambda_i \, |\lambda_i\rangle$$

$$\lambda_1 \leq \lambda_2 \leq \cdots$$

and the Variational Principle, which tells us that any state $|\mu\rangle$ that is not the ground state of H must be such that

$$\langle\mu|H|\mu\rangle = \tilde{\lambda} > \lambda_1$$

(or it would otherwise be the ground state of H.)

H is also self-adjoint (for the reason that it is a Hamiltonian), so by the Spectral Theorem its eigenvectors span the space;

$$|\mu\rangle = \sum_{\lambda_i} c_i \, |\lambda_i\rangle$$

$$\langle\mu|H|\mu\rangle = \sum_{\lambda_i \lambda_j} \langle\lambda_i| \, c_i^* H c_j \, |\lambda_j\rangle = \sum_{\lambda_i} |c_i|^2 \lambda_i = a\lambda_1 + (1-a)\bar{\lambda} = \tilde{\lambda}$$

$$a = |c_1|^2 \in [0, x], \quad \bar{\lambda} = \sum_{\lambda_i > \lambda_1} \frac{|c_i|^2}{1-a}\lambda_i > \lambda_1$$

with $x < 1$ some cutoff that is imposed by the *Ansatz* not spanning the ground state. The expectation value $\langle\mu|H|\mu\rangle$ is minimized by maximizing a, thus producing the maximum overlap with $|\lambda_1\rangle$ possible (aside from a phase).

All of the above discussion can be summarized by the schematic in fig. 3.

As the ground state is *a priori* unknown, we can maximize the chances that it is covered by considering an *Ansatz* that spans as most of the N-qubit Hilbert space as possible (again, not considering constraints in the state space). For example, [9, 38, 27] employ a generic structure of Euler rotations and entangling (controlled Pauli Z) gates, presented in fig. 4.

Once an *Ansatz* is chosen, some method of minimization must be employed to minimize the energy with respect to the parameters of the *Ansatz*. This is a generic problem of objective function minimization, which is a rich area of investigation in its own right[6], and so we will simply describe a few relevant optimization algorithms: ADAM, ADADELTA and L-BFGS. ADAM and ADADELTA are machine learning (ML) algorithms and so should be able to optimize under a very large space of parameters. L-BFGS cannot scale as greatly. However, L-BGFS generally converges much faster to the correct extreme value (see section 6 for a practical comparison), but requires complete knowledge of the gradient. ADAM and ADADELTA are designed to be robust to stochastic evaluation of the gradient. There is, therefore, a tradeoff between scaling-capability and performance. It is relevant to consider ML optimization algorithms, as we can expect that sufficiently complex *Ansätze* can depend on a number of parameters comparable to that in a neural network, and so we should characterize how optimization would occur under such circumstances.

---

[6]See, for example, reference [2] for an example of a collection of optimization methods.
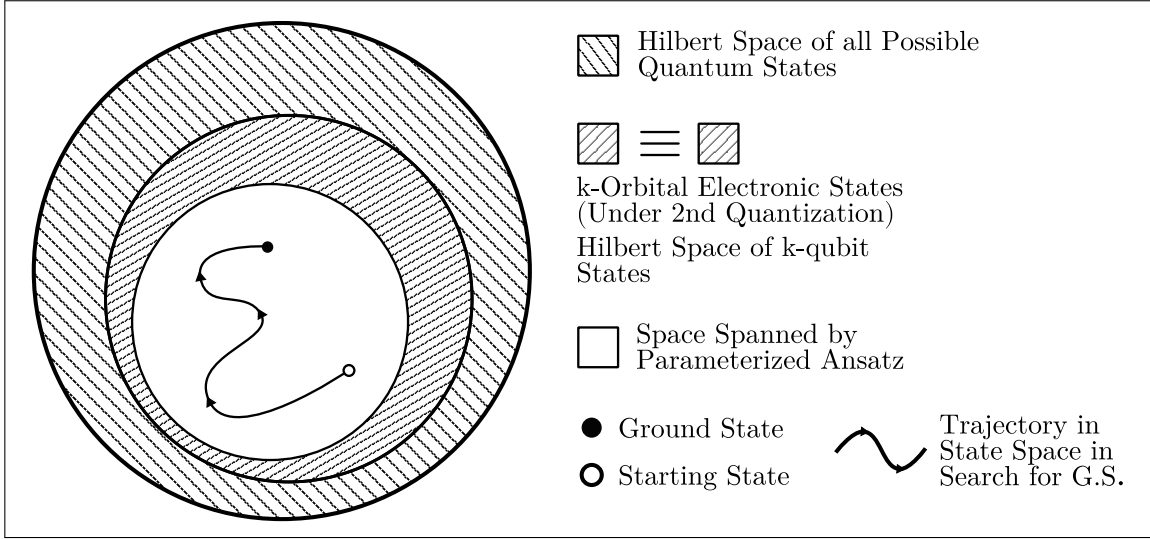
Figure 3: Representation of the search of the ground state of an electronic Hamiltonian using a quantum computer. There are several unspecified details, such as the form of the *Ansatz*, how much of the possible Hilbert space does it span and how the optimization itself, i.e., the trajectory in Hilbert space, is done.
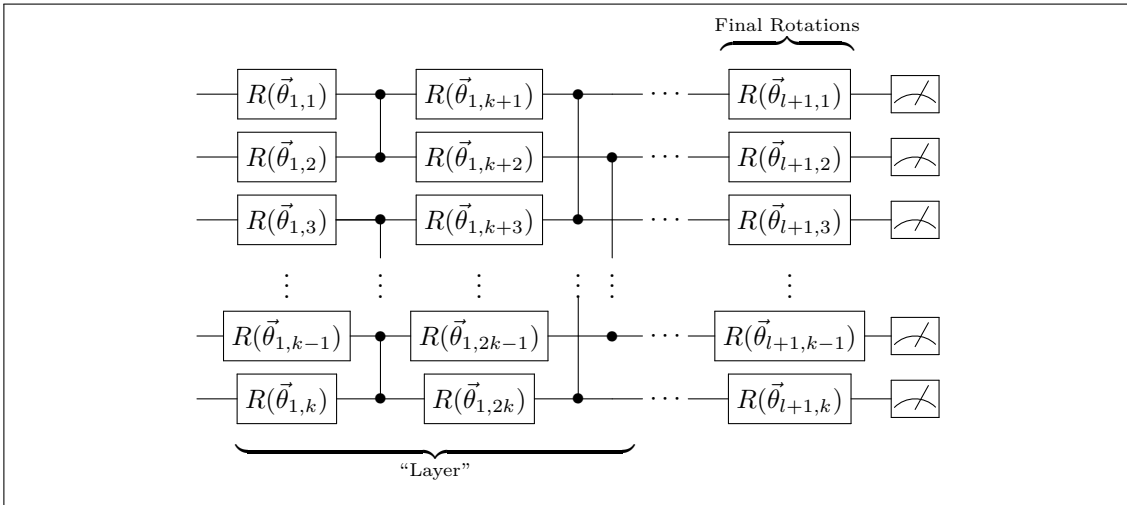
Figure 4: An *Ansatz* that seeks to parameterize as much of the Hilbert space of the corresponding number of qubits as possible. This is done by performing generic Euler rotations on each of the qubits, and then entangling them via controlled Pauli Z operations. The $R(\vec{\theta})$ gate may be written in terms of Pauli operations as $R(\vec{\theta}) = R_x(\theta^{(1)})R_y(\theta^{(2)})R_z(\theta^{(3)}) = \exp\left\{-i\frac{\theta^{(1)}}{2}\sigma^x\right\}\exp\left\{-i\frac{\theta^{(2)}}{2}\sigma^y\right\}\exp\left\{-i\frac{\theta^{(3)}}{2}\sigma^z\right\}$. See [27] for further discussion.

9

1.3.1 ADAM

ADAM is a "first-order gradient-based [algorithm for] optimization of stochastic objective functions, based on adaptative estimates of lower-order moments" [29]. It is built to incorporate the advantages of ADAGRAD [39] and RMSProp [1], i.e., to function with sparse gradients and a moving objective and/or varying parameters. The ADAM algorithm, as presented in [29], is presented in algorithm 1.

Algorithm 1 Computing an ADAM optimization routine.

Require: ϵ ▷ Very small increment (e.g., 10^{-8})
Require: α ▷ Stepsize (e.g., 0.001)
Require: $\beta_1, \beta_2 \in [0, 1)$ ▷ Exponential decay rates for the moment estimates (e.g., $0.9, 0.999$, respectively)
Require: $f(\theta)$ ▷ Stochastic objective function with parameters θ
Require: θ_0 ▷ Initial parameter vector
    $m_0 \leftarrow 0$ ▷ Initialize 1st moment vector
    $v_0 \leftarrow 0$ ▷ Initialize 2nd moment vector
    $t \leftarrow 0$ ▷ Initialize timestep
    while θ_t not converged **do**
        $t \leftarrow t + 1$
        $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ ▷ Get gradients with respect to stochastic objective at timestep t
        $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ ▷ Update biased first moment estimate
        $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ ▷ Update biased second raw moment estimate
        $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ ▷ Compute bias-corrected first moment estimate
        $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ ▷ Compute bias-corrected second raw moment estimate
        $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ ▷ Update parameters
    end while
    return θ_t ▷ Resulting parameters

1.3.2 ADADELTA

Presented in 2012, ADADELTA [51] is a dynamically adapting gradient-based optimization method that builds upon other adaptive methods, like ADAGRAD[7] and ADAM [29]. Its main features are robustness over meta-parameters and to noisy gradients, and minimal memory overhead. The ADADELTA algorithm, as presented in [51], is presented in algorithm 2.

1.3.3 L-BFGS

The L-BFGS (Low-memory—Broyden–Fletcher–Goldfarb–Shanno, sometimes typeset as LM-BFGS) algorithm [11] is a quasi-Newtonian optimization algorithm based on a low-memory representation of an estimate of the inverse Hessian matrix. It requires a "backlog" of m iterations of the parameter-space position and associated gradients. The L-BFGS algorithm, ignoring the variant where sufficiently low curvature information is discarded to prevent over-estimation of the optimization step, is presented in algorithm 3. Modifying the algorithm to check whether local curvature is above a certain threshold before updating the memory queues is trivial. This curvature-checking variant of L-BFGS was implemented in the QOP optimization tooling as a proof-of-concept. To obtain practical results (lower bound calculations in the proposed geometry

---

[7]which, like ADADELTA, is also unpublished; notes on the ADAGRAD method can be found at [39].

Algorithm 2 Computing an ADADELTA optimization routine.

Require: ρ ▷ Decay rate
Require: ϵ ▷ Very small increment
Require: Θ_1 ▷ Initial parameters
    $\mathbf{RMS}[x_t] \overset{\text{def}}{=} \sqrt{E[x^2]_t + \epsilon}$
    $E[g^2]_0 = 0,\ E[\Delta\Theta^2]_0 = 0$ ▷ Initialize accumulation variables
    loop
        $g_t = \nabla_\Theta C$ ▷ Compute gradient
        $E[g^2]_t = \rho E[g^2]_{t-1} + (1 - \rho)g_t^2$ ▷ Accumulate gradient
        $\Delta\Theta_t = -\frac{\mathbf{RMS}[\Delta\Theta]_{t-1}}{\mathbf{RMS}[g]_t} g_t$ ▷ Compute update
        $E[\Delta\Theta^2]_t = \rho E[\Delta\Theta^2]_{t-1} + (1 - \rho)\Delta\Theta_t^2$ ▷ Accumulate update
        $\Theta_{t+1} = \Theta_t + \Delta\Theta_t$ ▷ Apply update
        if termination condition is verified **then**
            break out of loop
        end if
    end loop

---

calculation technique) we used the Scipy [47] implementation of L-BFGS(-B), which enjoys a much greater level of optimization.

1.4   Schrieffer-Wolff Transform

The locality of a qubit Hamiltonian is defined as the greatest number of qubits affected by a single term of the operator. Generally, a Hamiltonian acting on N qubits will be N-local, as it will involve $(0 \dots N)$-local terms:

$$
\begin{aligned}
H = \ &\langle\text{scalar}\rangle + \sum_{\substack{i\in\{1,\dots,N\} \\ s\in\{x,y,z\}}} a_{i,s}\sigma_i^s && \text{(1-local)} \\[2mm]
&+ \sum_{\substack{i\neq j \\ (i,j)\in\{1,\dots,N\}^2 \\ s_i,s_j\in\{x,y,z\}^2}} b_{i,j,s_i,s_j}\sigma_i^{s_i} \otimes \sigma_j^{s_j} && \text{(2-local)} \\[2mm]
&+ \quad \cdots \\[2mm]
&+ \sum_{s_1,\dots,s_N\in\{x,y,z\}^N} c_s\ \sigma_1^{s_1} \otimes \sigma_2^{s_2} \otimes \cdots \otimes \sigma_k^{s_N} && \text{(N-local)}
\end{aligned}
\tag{7}
$$

Bravyi et al. [10] propose a transform that does not exactly lower the locality of a given Hamiltonian, but rather produces a *different* Hamiltonian, of lower locality and greater qubit count, whose ground state matches that of the original Hamiltonian. They have termed this mapping a Schrieffer-Wolff (SW) transformation, which operates on the principle that by introducing an ancillary qubit we may either map a k-local Hamiltonian to a $(\lceil k/2 \rceil + 1)$-local Hamiltonian, or a 3-local Hamiltonian to a 2-local Hamiltonian. These "sub-mappings" are termed *gadgets*. By recursively applying the gadgets, we may map any k-Hamiltonian to a 2-local Hamiltonian. The proof that these gadgets correctly reproduce the ground state energy of the original Hamiltonian

Algorithm 3 Computing a L-BFGS optimization routine.

Require: m ▷ Number of "memorized" previous states
Require: S ▷ Double ended queue capable of containing m parameter steps
                          ▷ (index 1 corresponds to most recent, index m corresponds to oldest)
Require: Y ▷ Double ended queue capable of containing m gradient differentials
                                                 ▷ (indexed like S)
Require: $\vec{x}_0$ ▷ Initial parameters
Require: $f(\vec{x})$ ▷ Cost function
    $\vec{x} \leftarrow \vec{x}_0$
    $g_{k-1} \leftarrow \vec{0}$
    while halting condition is not satisfied **do**
        $g_k \leftarrow \nabla_{\vec{x}} f(\vec{x})$
        $q \leftarrow g_k$
        for i **in** $1 \ldots \text{SIZE}(S, Y)$ **do**
            $\rho_i \leftarrow 1/(Y_i^T S_i)$
            $\alpha_i \leftarrow \rho_i S_i^T q$
            $q \leftarrow q - \alpha_i Y_i$
        end for
        $\gamma_k \leftarrow (S_1^T Y_i)/(Y_i^T Y_i)$
        $H_k^0 \leftarrow \gamma_k I$
        $z \leftarrow H_k^0 q$
        for i **in** $\text{SIZE}(S, Y) \ldots 1$ **do**
            $\beta_i \leftarrow \rho_i Y_i^T z$
            $z \leftarrow z + S_i(\alpha_i - \beta_i)$
        end for
        $z \leftarrow -z$
        $\vec{x} \leftarrow \vec{x} + z$
        push $g_k - g_{k-1}$ into Y
        push z into S
        $g_{k-1} \leftarrow g_k$
    end while

---

is non-trivial and can be found in [10], along with the formal requirements for the transform to be applicable. Therefore, we will not reproduce the proof here, but will note that the operating principle of the proof is decoupling the high- and low-energy subspaces of the gadget Hamiltonians and showing that a projection onto the low-energy subspace matches the target (i.e., original) Hamiltonian, up to a bounded error.

1.4.1 Subdivision Gadget $(k \longmapsto \lceil k/2 \rceil + 1)$

Let the original Hamiltonian H be written as

$$H = J\, A \otimes B$$

with $J > 0$ the *coupling* of the Hamiltonian, and A and B operators acting on disjoint subspaces of the subspace in which H acts. An ancillary qubit u is introduced to produce the following mapping

$$\boxed{H = J\, A \otimes B \quad \longmapsto \quad H' = H_0 + V_{\text{pert.}} + V_{\text{corr.}}}$$ (8)

$$H_0 = \Delta \cdot I_A \otimes I_B \otimes |1\rangle\langle 1|_u$$

$$V_{\text{pert.}} = \sqrt{\frac{\Delta J}{2}}\left(-A \otimes I_B + I_A \otimes B\right) \otimes X_u$$

$$V_{\text{corr.}} = \frac{J}{2}\left(A^2 \otimes I_B + I_A \otimes B^2\right) \otimes I_u$$

(where I denotes the identity operator, K_A, K_B denote the operator K over the subspace of operators A, B, K_u denotes the operator K acting on the ancillary qubit u, Δ a free constant associated to the mapping, X the Pauli σ_x operator)

It is proven in [10] that, for a given Δ,

$$\mathbf{GSE}\,[H'] = \mathbf{GSE}\,[H] + \mathcal{O}(J^{3/2}\Delta^{(-1/2)})$$

($\mathbf{GSE}\,[H]$ denoting the ground state energy of Hamiltonian H)

Therefore, by choosing arbitrarily small ϵ and letting

$$\Delta = J\epsilon^{-2}$$ (9)

the error is made $\mathcal{O}(J\epsilon)$.

1.4.2 3-to-2-local Gadget

Let the Hamiltonian be of the form

$$H = J\, A \otimes B \otimes C$$

with $J > 0$ the coupling, A, B, C single qubit operators acting on disjoint spaces. Introducing ancillary qubit u, the lower locality Hamiltonian is produced using the following mapping:

$$\boxed{H = J\, A \otimes B \otimes C \quad \longmapsto \quad H' = H_0 + V_{\text{pert.}}^{(1)} + V_{\text{pert.}}^{(2)} + V_{\text{corr.}}^{(1)} + V_{\text{corr.}}^{(2)}}$$ (10)

$$V_{\text{pert.}}^{(1)} = -\Delta^{2/3} J^{1/3}\, |1\rangle\langle 1|_u \otimes I_A \otimes I_B \otimes C$$

$$V_{\text{pert.}}^{(2)} = \frac{\Delta^{2/3} J^{1/3}}{\sqrt{2}} X_u \otimes (-A \otimes I_B + I_A \otimes B) \otimes I_C$$

$$V_{\text{corr.}}^{(1)} = \frac{\Delta^{1/3} J^{2/3}}{2} I_u \otimes (-A \otimes I_B + I_A \otimes B)^2 \otimes I_C$$

$$V_{\text{corr.}}^{(2)} = \frac{J}{2} I_u \otimes (A^2 \otimes I_B + I_A \otimes B^2) \otimes C$$

(where I is the identity operator, and K_W denotes the action of operator K over the qubit associated with operator W, Δ a free constant associated to the mapping, X the Pauli σ_x operator)

Again, it is proven in [10] that

$$\mathbf{GSE}\left[H'\right] = \mathbf{GSE}\left[H\right] + \mathcal{O}(J^{4/3}\Delta^{-1/3})$$

So that, again for some ϵ,

$$\Delta = J\epsilon^{-3} \Rightarrow \mathbf{GSE}\left[H'\right] = \mathbf{GSE}\left[H\right] + \mathcal{O}(\epsilon J) \tag{11}$$

1.4.3 Cross-Gadget Contributions

It is also proven in [10] that successively employing gadgets introduces a bounded error: if n gadgets are used on a Hamiltonian whose coupling is J, then the "cross-gadget error" introduced is bounded by $\mathcal{O}(\epsilon n J)$.

2 Objectives

The goal of this thesis is to develop and test a new, quantum device oriented, method for obtaining the geometric parameters of a molecule (or otherwise physical parameters) that result in the lowest possible energy, within the space of these parameters. In a diatomic molecular system, for example, this would correspond to the bond length. The method utilizes a quantum variational algorithm (QVA) approach, and is designed to make use of the strengths of quantum computers, in particular the fact that the expectation value of an operator composed of Pauli operators can be efficiently evaluated over a previously encoded state. On the other hand, it is designed to be applicable in extreme cases of the current framework of quantum computation, the Noisy Intermediate-Scale Quantum (NISQ) regime, namely when the number of available qubits is too small to encode a trial state. We explore the bottlenecks of such a simulation task with high control by implementing a quantum circuit simulator down to a low-level of implementation. We apply it on testing the method in simple molecular systems.

3 Development of a Quantum Simulator: 'QOP'

In order to maintain full control over the simulation and investigate possible bottlenecks, as well as implement the discussed ideas straightforwardly while preserving speed of simulation, an important part of the thesis work was devoted to the construction of our own quantum simulator in C99 [25]. It was named QOP, standing for **Q**uantum **Op**timizer. C, as a low level language, provides high control over the computation, but on the other hand interfaces natively with Python [18], and so we had the possibility of exposing a high level, flexible API.

The quantum simulator was scoped so that the C subset could operate independently of the Python interface layer; we refer to this pure C subset as C-QOP. The C subset itself was
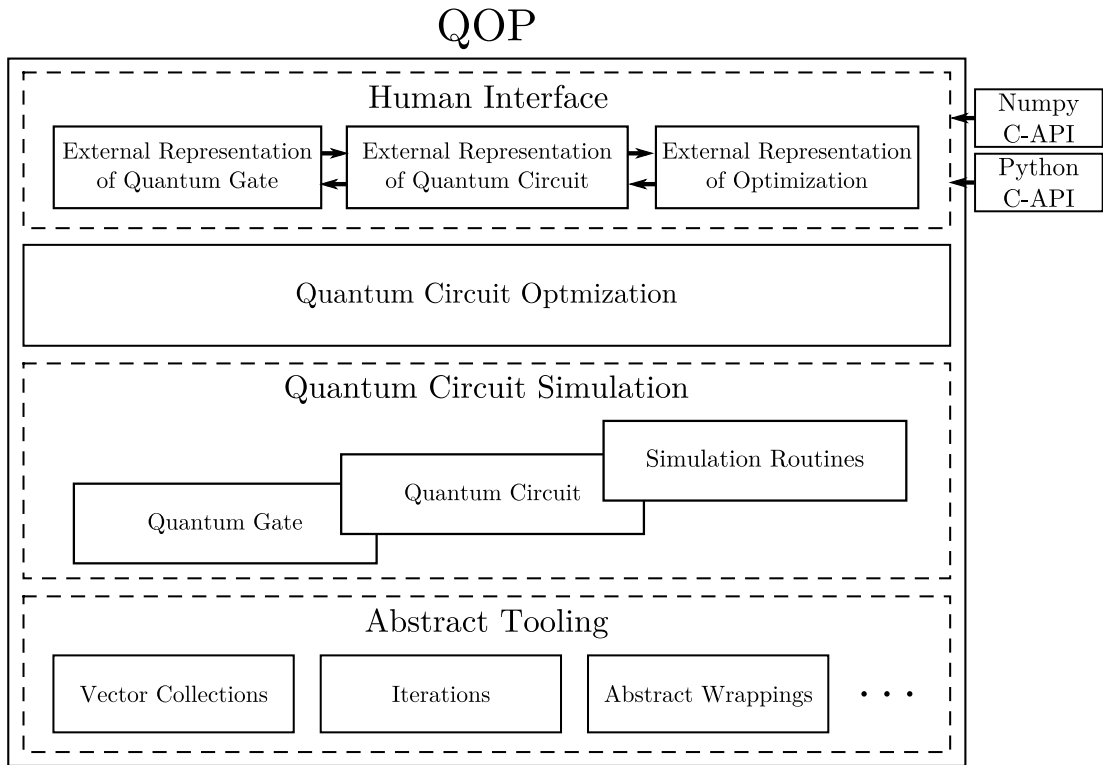
Figure 5: Different scopes of the developed quantum simulator, organized in terms of their interdependence. Items at the bottom are self-contained, and each element above depends on the ones below. Note that at the "Human Interface" level, all items are interdependent in implementation. The simulator was named QOP (as in *Q*uantum *Op*timizer), and C-QOP is used to describe the stack from "Quantum Circuit optimization" down, as that subset of the simulator is self-contained and can be used with C.

constructed with a "bottom-up" approach, with each layer of implementation depending strictly on the layers below. The top-most layer is responsible for the Python interface, and relies also on the NumPy API [36]. The different scopes and dependencies of the whole simulator are presented in fig. 5.

The bottom-most layer ("Astract Tooling"; cf. fig. 5) provides only a few common programming abstractions; C's standard library is reduced when compared to, for example, C++'s standard library. An example is arrays of variable size, typically called "lists" or "vectors". Other constructs included in this layer are "options", as an idiomatic way of expressing the possibility that a variable's value is not defined, and "iterators" as a fail-safe way of accessing collections of elements in memory.

All of these constructs relate not with the task of simulating a quantum circuit or simulating QVAs, but rather with ways of idiomatically expressing programming ideas inside the paradigm of C. As such, they will not be discussed in detail. Nonetheless, the very atomic nature of every element in the abstract tooling layer, along with an effort to thoroughly document their implementation should make it easy to understand their internals and use.

While it is unavoidable that we must store all of the 2^q amplitudes of a generic q-qubit

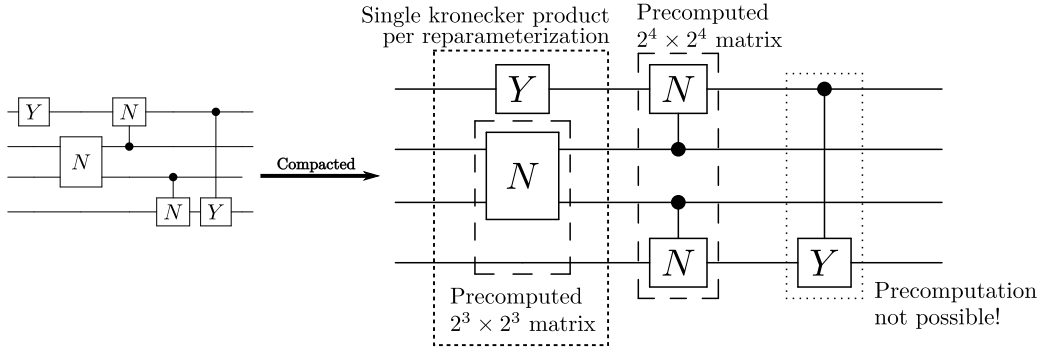quantum state if we wish to describe it[8], the same is not true of a quantum gate — that is, we may store different computational representations of a quantum gate. As an example of this, consider a 4 qubit circuit, with a single NOT (Pauli X) gate acting on the third qubit. This operation may be described by the matrix

$$
\begin{array}{c}
\overset{/2}{\underset{\boxed{X}}{\rule{0pt}{0pt}}}
\end{array}
= X^{(3,4)} = I^{\otimes 2} \otimes X \otimes I =
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
$$

such that the state's amplitudes after the action of the gate are simply given by $\vec{\psi'} = X^{(3,4)}\vec{\psi}$ (with $\vec{\psi}$ the column vector of amplitudes). On the other hand we can consider the action of the gate to be represented by a tensor $\chi^{\nu}{}_{\mu}$ such that, if the 4-qubit state is given by

$$
\sum_{\alpha\beta\gamma\delta} Q^{\alpha\beta\gamma\delta} \left|\alpha\beta\gamma\delta\right\rangle \qquad \text{where} \qquad \alpha,\beta,\gamma,\delta \in \{0,1\}
$$

then the state after the action of the gate is given by

$$
\sum_{\alpha\beta\gamma\delta} \tilde{Q}^{\alpha\beta\gamma\delta} \left|\alpha\beta\gamma\delta\right\rangle \qquad \text{where} \qquad \tilde{Q}^{\alpha\beta\gamma\delta} = \sum_{\mu} \chi^{\gamma}{}_{\mu} Q^{\alpha\beta\mu\delta} \qquad (\chi^{\nu}{}_{\mu}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{12}
$$

Thus we see that the choice of gate representation will impact computation time and memory requirements, but will also be related to the computational representation of the state.

With QOP, we adopted a metadata approach, where each gate is stored by its "local" (non-controlled) matrix representation, along with information regarding whether and by which qubit it is controlled, and to what qubit the gate is applied. Because this representation is less memory intensive and more flexible and human oriented (compare, for example, the metadata description of $X^{(3,4)}$ with the matrix representation above), but less convenient for the actual computation, we designated this representation a *soft gate*. Once a matrix is computed from this metadata representation, we refer to a "hardened" form, where the properties of the gate (such as the controlling qubit or target qubit) cannot be easily changed without recomputing the matrix.

We thus make the following distinctions, extending the above discussion and separation of concept/representation to other objects of quantum computation:

(i) A *gate object* is an internal representation of a quantum gate, uncoupled from its action in a quantum circuit; it typically consists of a single-qubit matrix representation and possibly

---

[8]It would be odd if it were not so, as that would mean amplitudes were correlated in an arbitrary state.

Figure 6: Example of the hardening process in the first iteration of the quantum simulator. Gates that are parameterized are marked with a "Y", while non-parameterized gates are marked with a "N". Notice that "empty" spots in the circuit are equivalent to the identity gate, so that they are also considered.

parameters. It has the ability of mutating its matrix representation in place when given new parameters.

(ii) A *soft gate*, as discussed above, is a metadata-based internal representation of the action of a gate in a circuit. It is composed of a gate object and metadata, such as the qubit the gate acts on, the position of the gate in the circuit, and if and what qubit controls the gate. It is made a distinct entity from a gate object because it contains information regarding not only the quantum gate but the gate's relationship with the circuit.

(iii) A *circuit object* is an internal representation of a quantum circuit. It is essentially a set of soft gates.

(iv) A quantum circuit *slice* is a set of gates acting in parallel and in disjoint subspaces in a quantum circuit.

(v) A *compacted* quantum circuit is one in which no gate can be moved to an earlier slice without commuting with some other gate.

(vi) A *hardened* quantum circuit object is one whose internal representation is optimized towards computing the effect of the circuit on a state, but does not allow the modification of the circuit without recalculation of the representation.

The first, more naïve, approach to circuit action computation was to compact the circuit as much as possible, then compute the matrix corresponding to each slice, avoiding recomputation when possible upon reparameterization. The action of the circuit would then be given by the successive matrix products.

When "hardening" the circuit, vertical sequences of non-parameterized gates would be identified, and for these the full matrix representation would be immediately calculated. Then, when actually performing a simulation with some parameters, the matrix form for the parameterized gates would be determined, and aggregated with the other gates of the slice to form the final matric form of the slice. A visual representation of this hardening process is given in fig. 6.

One issue that arises is determining the matric form of any arbitrary controlled gate. For this, the technique presented by Gidney in [23] was used. Considering a scalar value with a special algebra, μ, that obeys
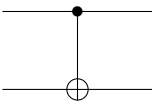
17

$$\mu \otimes U = \mu \cdot I_n$$

$$U \otimes \mu = I_n \cdot \mu$$

where U is an $n \times n$ matrix, and I_n is the $n \times n$ identity matrix, a "control matrix" can be defined as
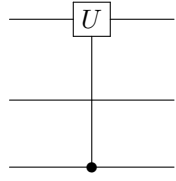
$$C = \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix}$$

With this, the matrix corresponding to a controlled operation can be obtained by taking adequate kronecker products; for example:

 $\equiv \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \mu & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$

where the μ values are cast back to 1 in the final matrix. Another more complex example illustrating the asymmetry of the algebra of μ is[9]

 $\equiv \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix} =$

$$= \begin{pmatrix} u_{11} & 0 & u_{12} & 0 \\ 0 & u_{11} & 0 & u_{12} \\ u_{21} & 0 & u_{22} & 0 \\ 0 & u_{21} & 0 & u_{22} \end{pmatrix} \otimes \begin{pmatrix} \mu & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \mu & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{11} & 0 & 0 & 0 & u_{12} & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{11} & 0 & 0 & 0 & u_{12} \\ 0 & 0 & 0 & 0 & \mu & 0 & 0 & 0 \\ 0 & u_{21} & 0 & 0 & 0 & u_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & u_{22} & 0 & 0 & 0 & u_{22} \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u_{11} & 0 & 0 & 0 & u_{12} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_{11} & 0 & 0 & 0 & u_{12} \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & u_{21} & 0 & 0 & 0 & u_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & u_{22} & 0 & 0 & 0 & u_{22} \end{pmatrix}$$

This construct is useful because it allows to algorithmically construct the matrix of any controlled operation. See appendix B for further discussion of the validity of μ ("tainted numbers").

---

[9]Notice that if the μ appears on the right-hand side of the tensor product, it is the index of the element on the left-hand side that determines whether the product is equal to μ or 0; if the element a on the left has indexes (i, j), the result of $a \cdot \mu = \delta_{ij}\mu$.

The reason this approach was unsuccessful is that it amplified the inefficiency of representing quantum operations as a full-size matrix. Namely, each slice of a circuit ultimately corresponds to a $2^q \times 2^q$ sized matrix, where q is the number of qubits of the circuit. Simultaneously holding a matrix representation of each slice in memory accelerates the problem; then, a q-qubit circuit of horizontal depth s requires $2^{2q \log s}$ values in memory!

Finally, regardless of memory concerns, the usefulness of precomputing vertical sequences of non-parameterized gates is questionable, as it presupposes a circuit structure of a few parameterized gates amidst many parameterized gates in a circuit of small horizontal depth. This is not observed in

- fully non-parameterized circuits,
- circuits where every gate is (or most gates are) parameterized,
- the Quantum Fourier Transform (QFT) circuit, or
- the Quantum Phase Estimation (QPE) circuit,

with the latter two being reference circuits.

For all of these reasons, this approach was abandoned in favor of a second approach: to sidestep working with full-size matrix products by working strictly within the subspace affected by each specific gate (akin to eq. (12)). In this language, a controlled unitary operation would be denoted

$$
\begin{array}{c} i \xrightarrow{\quad\bullet\quad} j \\ \alpha \;\;\fbox{}\;\; \beta \\ k \;\boxed{U}\; l \end{array}
\quad \sim \quad
\delta^j_{\;q^{(i)}} \delta^1_{\;q^{(i)}} U^l_{\;q^{(k)}} + \delta^j_{\;q^{(i)}} \delta^0_{\;q^{(i)}} \delta^l_{\;q^{(k)}} = |1\rangle\langle 1|_i \otimes U_k + |0\rangle\langle 0|_i \otimes I_k
$$

There is, however, an optimization to be made in a tensor-like approach: to efficiently iterate over the tensor indices so as to take advantage of swap-symmetric indices, rather than perform the same set of operations more than once (in the analogy of the drawing above, to "copy" the result of the operation to all possible values of α,β). For this, the relevant index "bit mask" is precomputed, so that iterations over the tensor indices are split into two parts: iterations over the bits of the indices under the mask, which imply a new computation, and iterations over the bits not covered by the mask, for which a previously calculated value can be reused. Iteration of binary numbers under a bit mask was done based on Chase's "twiddle" algorithm [12].

With this in mind, the simulation algorithm presented in algorithm 4 was constructed.

This approach avoids calculating matrix products entirely by breaking down what would be a $2^N \times 2^N$ matrix product into the non-redundant product of scalars in the appropriate single-qubit subspaces. Notice that only the $2 \cdot 2^N$ values of $\vec{\imath}$ and $\vec{o}$ need to be held in memory during the whole routine (excluding the gate matrices — these require $\sharp_s \cdot 2^2$ values to be in memory for each slice, $\sharp_s$ being the number of gates in slice s). By utilizing the gate object's matrix directly, we also avoid the need for a "hardening" stage in the same sense as in the first approach. Instead, this stage is used to cache information about the circuit that does not change with parameterization.

On the other hand, this approach assumes that every gate on the circuit acts on a single qubit, possibly controlled by another qubit. However, this is reasonable; the set of all single qubit gates plus the CNOT gate provide a universal basis set, i.e., "all unitary operations on arbitrarily many qubits... can be expressed as compositions of these gates" [15, 5]. Since the three x, y, z rotation gates are available, and any unitary operation on a single qubit ($SU(2)$) is realizable with three rotation gates [46], any single qubit operation is realizable.

The fact that gates can be controlled also satisfies the CNOT requirement, as a $cR_x(\pi) \equiv$ CNOT (up to a global phase). Finally, many of the currently available physical realizations of quantum computers also similarly only support a universal set of one qubit operations plus the CNOT operation [30].

19

Algorithm 4 Algorithm employed in QOP for the simulation of a N-qubit quantum circuit.

---

    function AND(x,y)
        Bitwise logical And
    end function
    function OR(x,y)
        Bitwise logical Or
    end function
    function GATEBITSTRING(slice)
        return Binary string with a 1 at indices where the slice's matching index wires
                                      have a gate, and a 0 otherwise.
    end function
    function CGATEBITSTRING(slice)
        return Binary string with a 1 at indices where the slice's matching index wires
                                 have a gate or a control, and a 0 otherwise.
    end function
Require: List of slices l
Require: List of gates in each slice g
Require: Input state amplitudes $\vec{i} \leftarrow \langle i_1, i_2, \ldots, i_{2^N} \rangle$
Begin:
    for slice s **in** l **do**
        $cgb \leftarrow$ CGATEBITSTRING(s)
        $\vec{o} \leftarrow \langle 0, 0, 0, \ldots \rangle$
        for x **in** $\left(2^N \text{ AND } cgb\right)$ **do**
            for y **in** $\left(2^N \text{ AND GATEBITSTRING}(s)\right)$ **do**
                $coef \leftarrow 1.$
                for gate g **in** slice s **do**
                    $k \leftarrow$ WIRE(g)
                    $U \leftarrow$ MATRIX(g)
                    $u \leftarrow U_{y_k, x_k}$
                    if g is controlled by index p **and** $x_p = 0$ **then**
                        $u \leftarrow \delta_{x_k, y_k}$
                    end if
                    $coef \leftarrow coef \cdot u$
                end for
                for z **in** $(2^N \text{ AND NOT}(cgb))$ **do**
                    $out \leftarrow (z \text{ OR } x) \text{ AND } y$
                    $o_{out} \leftarrow o_{out} \cdot coef$
                end for
            end for
        end for
        $\vec{i} \leftarrow \vec{o}$
    end for
Done: $\vec{i}$ now contains the output of the circuit.

---

4   Method for Calculating Geometric Parameters

Quantum Variational Algorithms address part of the limitations of a Noisy Intermediate-Scale Quantum regime, namely the reduced relaxation and coherence times. However, in a NISQ regime there is another limitation: the reduced number of qubits available. In particular, the previously presented tools (Born-Oppenheimer approximation, Jordan-Wigner Transformation, Quantum Variational Algorithms) allow one to search for the ground state (energy) of some Hamiltonian provided if and only if some trial state can be encoded; we should not hope to determine the ground state of an N-qubit Hamiltonian (using a quantum computer) if we cannot produce any candidate state at all. We propose, then, a different approach, by noting that when treating a physical system we might not be interested in determining the ground state energy itself, but rather the physical parameters of the system corresponding to the ground state configuration — the "true" parameters of the system at rest:

$$|\phi\rangle \ : \quad \min_{|\phi\rangle} \langle\phi|H(R_i)|\phi\rangle \quad \longrightarrow \quad \langle R_1, \ldots\rangle \ : \quad \min_{\langle R_1, \ldots\rangle} \langle\mathbf{GS}[H(R_i)]|H(R_i)|\mathbf{GS}[H(R_i)]\rangle$$

(where $\mathbf{GS}[H]$ denotes the ground state of Hamiltonian H, and $\{R_i\}$ are the nucleic positions; cf. section 1.1)

The right-hand side is still not computable, as it still requires knowledge of the ground state. We may recall, however, that if some Hamiltonian H can be written as a sum of terms,

$$H = h^{(1)} + h^{(2)} + \cdots + h^{(u)}$$

then we are guaranteed that

$$\mathbf{GSE}\left[H\right] \geq \mathbf{GSE}\left[h^{(1)}\right] + \cdots + \mathbf{GSE}\left[h^{(u)}\right] \tag{13}$$

($\mathbf{GSE}\left[x\right]$ denoting the ground state energy of x)

This is straightforward to derive; H is self-adjoint and so has real eigenvectors. This implies $h^{(1)} \ldots h^{(u)}$ must also be self-adjoint, and so by the Spectral Theorem[10],

$$\langle\phi|H|\phi\rangle \,,\, \langle\phi|h^{(1)}|\phi\rangle \,,\, \ldots,\, \langle\phi|h^{(u)}|\phi\rangle \in \mathbb{R} \qquad \forall\,|\phi\rangle$$

We have thus that $\mathbf{GSE}\left[H\right]$ is a function onto the real numbers given by a sum of functions onto the real numbers, and trivially

$$f(x) = \sum_i f_i(x)$$

$$\min_x f(x) = f(a) \qquad \min_x f_i(x) = f_i(a_i)$$

$$f_i(a) \geq f_i(a_i) \Rightarrow \min_x f(x) = f(a) = \sum_i f_i(a) \geq \sum_i f_i(a_i) = \sum_i \min_x f_i(x)$$

---

[10]Regardless of complex factors in the decomposition of $|\phi\rangle$ in the eigenvectors of H:

$$|\phi\rangle = \sum_k c_k\,|\lambda_k\rangle \,, \qquad c_k \in \mathbb{C}$$

$$\langle\phi|H|\phi\rangle = \sum_k |c_k|^2\,\langle\lambda_k|H|\lambda_k\rangle \,, \qquad |c_k|^2 \in \mathbb{R}$$

We may proceed to exploit this if the post–Jordan-Wigner–transformation Hamiltonian H in study can be separated into terms that involve different qubits:

$$H = H_{123} + H_{45} \quad \longrightarrow \quad \mathbf{GSE}\left[H\right] \geq \mathbf{GSE}\left[H_{123}\right] + \mathbf{GSE}\left[H_{45}\right]$$

The ground state energies of H_{123}, H_{45} *separately* can be calculated with a three qubit quantum computer, rather than requiring five qubits, as does calculating the ground state energy of H.

However, generically and as discussed under section 1.4, an n-qubit Hamiltonian will be n-local, so that even if we are looking to calculate a lower bound to the ground state energy, rather than the ground state energy itself, the n-body term will always require a n-qubit quantum computer. To be able to calculate a lower energy bound per eq. (13) we require as many qubits as the locality of the Hamiltonian.

On the other hand, provided we *could* calculate a lower bound, what information regarding the system's at-rest parameters could we obtain? Defining

$$\mathbf{LB}\left[H(\{R_i\})\right] \stackrel{\text{def}}{=} \sum_i \mathbf{GSE}\left[h^{(i)}(\{R_i\})\right] \quad ; \quad H = \sum_i h^{(i)}$$

(Where, again, $\{R_i\}$ denotes the nucleic positions — see section 1.1)

there is, in fact, no guarantee that

$$\min_{\{R_i\}} \mathbf{GSE}\left[H(\{R_i\})\right] = \min_{\{R_i\}} \mathbf{LB}\left[H(\{R_i\})\right]$$

nor that

$$\min_{\{R_i\}} \mathbf{GSE}\left[H(\{R_i\})\right] \approx \min_{\{R_i\}} \mathbf{LB}\left[H(\{R_i\})\right]$$

nor even that $\min_{\{R_i\}} \mathbf{LB}\left[H(\{R_i\})\right]$ *exists*. It is an objective of this thesis to study the relationship between the left- and right-hand terms of the above, and if and in what conditions we can extract information about the system's at-rest parameters from the lower bound minimum. If the equations above hold, we will refer to this as the *structure* of the original ground state energy curve holding under the chosen lower bound.

We then sketch the full proposed procedure for calculating a lower bound energy curve for a molecule, using a quantum computer that cannot support the full Hamiltonian, making use of the various tools previously introduced:

---

1. Write the 1ˢᵗ-quantized Born-Oppenheimer Hamiltonian for the molecule in study,

2. Choosing an orbital function basis set, convert it into a 2ⁿᵈ-quantized form,

3. Using a Jordan-Wigner transform, obtain a Pauli-operator–based description, and

4. Apply Schrieffer-Wolff transforms as needed to reduce the locality of the Hamiltonian to $\leq$ the number of qubits available,

5. Calculate a lower bound by finding the ground state of separate terms of the Hamiltonian (using the available quantum computer/a Quantum Variational Algorithm approach).

The first three points are covered by sections 1.1 to 1.4. We will now refine the latter two points.

4.1 Decomposition "Styles"

Schrieffer-Wolff transformations are defined exclusively for Hamiltonians that are of either the form $J\,A \otimes B$ or $J\,A \otimes B \otimes C$ (cf. section 1.4); this is not necessarily the case for a general Hamiltonian. A Hamiltonian acting in space $A \otimes B$ will generally be of the form

$$H = \sum_{A_i\ B_j} c_{ij} A_i \otimes B_j$$

with $\{A_i\}$ all the Pauli-operator–based operators possible to define on A, and $\{B_j\}$ the analogous for B. We may seek to determine $\{a_i\}, \{b_j\}$ such that the system could be exactly written as a single decomposable term,

$$H \stackrel{?}{=} \left(\sum_i a_i A_i \right) \otimes \left(\sum_j b_j B_j \right)$$

(Implying the system of $i \cdot j$ equations $\{a_i b_j = c_{ij}\}$ for $i + j$ unknowns; the system is either under- or over-constrained.)

The previous equation may be written in matrical form as

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix} \begin{pmatrix} b_1 & b_2 & b_3 & \cdots \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \cdots \\ c_{21} & c_{22} & & \\ c_{31} & & c_{33} & \\ \vdots & & & \ddots \end{pmatrix} =$$

$$= \begin{pmatrix} a_1 & 0 & 0 & \cdots \\ a_2 & 0 & 0 & \cdots \\ a_3 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} \begin{pmatrix} b_1 & b_2 & b_3 & \cdots \\ 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \cdots \\ c_{21} & c_{22} & & \\ c_{31} & & c_{33} & \\ \vdots & & & \ddots \end{pmatrix} \quad (14)$$

such that the above is in the form of a Singular Value Decomposition[11] of the matrix (c_{ij}), and indicates that (c_{ij}) has only one Singular Value. However, all c_{ij} are arbitrary, so in general may have more than one singular value. As such $\{a_i\}, \{b_j\}$ will not exist.

We propose, then, different applications of the Schrieffer-Wolff transformation to reduce the locality of a given Hamiltonian:

4.1.1 Direct Decomposition

The first approach is simply to apply locality reducing gadgets to every term of the Hamiltonian that has a locality above the intended, separately. This is always possible, since any term can be partitioned into one of the two 'decomposable' forms (up to a minus sign in the coupling, which can be absorbed into the operators). The main benefit to this approach is that little to no preprocessing of the Hamiltonian must be done; a single pass over each term of the Hamiltonian is enough to identify and reduce the locality of each as needed. The down side to this approach, however, is the introduction of many ancillary qubits, possibly unnecessarily.

---

[11]The Singular Value Decomposition (SVD) is a generalization of the eigenvalue decomposition of a matrix. It consists of the decomposition of a matrix A into a product $U\Sigma V^\dagger$, where U and V are unitary matrices, and Σ is a non-negative diagonal matrix. Being unitary, the columns of U and V define the (respectively) left- and right-singular vectors. The diagonal values of Σ are termed the singular values of A. See [45] for a discussion of the history and definition of the Singular Value Decomposition.

4.1.2 Singular Value Decomposition (SVD) Based Decomposition

We may attempt to introduce less ancillary qubits by noting that we can obtain a "minimal decomposable form" of the Hamiltonian by employing a Singular Value Decomposition. We start by choosing subspaces A, B such that any term of the Hamiltonian can be written as

$$\sum_{ij} c_{ij} A_i \otimes B_j \qquad \text{where} \qquad A_i \in A , \ B_j \in B$$

with A_i, B_j being a tensor product of elements from $\{I, \sigma_x, \sigma_y, \sigma_z\}$ (cf. eq. (7)), to which we may assign an arbitrary lexicographic order[12]. We then construct the matrix (c_{ij}) based on the lexicographic order of A_i, B_j and perform a singular value decomposition on it. This yields left- and right- singular vectors, $(a_i^{(\alpha)})$ and $(b_j^{(\alpha)})$ respectively, corresponding to the singular values $s^{(\alpha)}$, such that

$$\begin{pmatrix} & c_{ij} & \end{pmatrix} = \sum_{\alpha}^{S} s^{(\alpha)} \begin{pmatrix} a_i^{(\alpha)} \end{pmatrix} \begin{pmatrix} & b_j^{(\alpha)} & \end{pmatrix}$$

The above may be reinterpreted in the context of the chosen operator basis as

$$H = \sum_{\alpha}^{S} s^{(\alpha)} \left(\sum_i a_i^{(\alpha)} A_i \right) \otimes \left(\sum_j b_j^{(\alpha)} B_j \right) = \sum_{\alpha}^{S} s^{(\alpha)} h^{(\alpha)}$$

so that for each of the $h^{(\alpha)}$ terms that has locality above ideal, we may employ the appropriate locality-reduction gadget, with the guarantee that we have chosen a new operator basis that minimizes the number of terms.

This technique cannot be blindly applied recursively; if, given a Hamiltonian that involves w systems, we consider the "low space" operators $\{\Lambda\}$ to be the operators acting on the first $\lfloor w/2 \rfloor$ systems, and the "high space" operators $\{\Xi\}$ to be the operators acting on systems of index $(\lfloor w/2 \rfloor + 1) \ldots w$, then we have, after applying once the SVD-based Schrieffer-Wolff transformation,

$$H = \sum_{\alpha}^{S} J^{(\alpha)} \Lambda^{(\alpha)} \otimes \Xi^{(\alpha)}$$

$$\Lambda^{(\alpha)} \propto \sum_i a_i^{(\alpha)} A_i \qquad \Xi^{(\alpha)} \propto \sum_j b_j^{(\alpha)} B_j \qquad J > 0$$

$$H \longmapsto H' = \sum_{\alpha} \Delta^{(\alpha)} |1\rangle\langle 1|_u^{(\alpha)} + \sqrt{\frac{\Delta^{(\alpha)} J^{(\alpha)}}{2}} \left(-\Lambda^{(\alpha)} + \Xi^{(\alpha)} \right) \otimes X_u^{(\alpha)} + \frac{J^{(\alpha)}}{2} \left(\Lambda^{(\alpha)2} + \Xi^{(\alpha)2} \right)$$

which is indeed of locality $(\lceil w/2 \rceil + 1)$, as we are "mixing" Λ (and Ξ) with the ancillary qubit's space separately. However, if we now blindly apply the same procedure again, because H' involves all of the original systems plus an ancillary one, the low and high spaces are redefined;

$$\{\Lambda'\} \rightsquigarrow \text{generally operators of systems } 1 \ldots \lfloor \frac{\lceil w/2 \rceil + 1}{2} \rfloor$$

$$\{\Xi'\} \rightsquigarrow \text{generally operators of systems } \left(\lfloor \frac{\lceil w/2 \rceil + 1}{2} \rfloor + 1 \right) \ldots (\lceil w/2 \rceil + 1)$$

---

[12]For example, $(I \otimes I \otimes \cdots \otimes I)$, $(I \otimes \cdots \otimes I \otimes \sigma_x)$, $(I \otimes \cdots \otimes I \otimes \sigma_y)$, $(I \otimes \cdots \otimes I \otimes \sigma_z)$, $(I \otimes \cdots \otimes \sigma_x \otimes I)$, $(I \otimes \cdots \otimes \sigma_x \otimes \sigma_x)$, etc.

With the coupling of these new operators with a new ancillary qubit, we may actually obtain a Hamiltonian of locality higher than H'. As an example, consider a Hamiltonian involving 6 systems, generically denoted as

$$H \sim h_{123456}$$

After one pass of the above presented technique, we obtain

$$H' \sim h_{1237} + h_{4567}$$

where a 7th ancillary system was introduced. However, H' is now a 7-qubit Hamiltonian, and so a SVD of the corresponding (c_{ij}) matrix (as above defined) will generally yield operators that simultaneously involve systems $1\,2\,3$ and $4\,5\,6\,7$, separately. Therefore, after another pass, and in the same notation

$$H'' \sim h_{1238} + h_{45678}$$

and locality has actually increased from H' to H''.

As a solution to this, we recursively apply the technique to each of the subspaces produced. In the context of the example above, this would mean that after the first pass, h_{1237} and h_{4567} are separately considered, yielding

$$H'' \sim h_{128} + h_{378} + h_{459} + h_{679}$$

The tradeoff to this approach is that more than one ancillary body may now be introduced in each pass.

Regardless, the upshot of the SVD-based decomposition scheme is that a minimal number of ancillary qubits are introduced (as opposed to Direct Decomposition). This comes as a direct consequence of performing a Singular Value Decomposition; we effectively choose the basis of operators that produces the minimum number of terms to be decomposed.

Finally, we may define the

4.1.3 k-Singular Value Decomposition (k-SVD) Based Decomposition

Singular Value Decomposition can be used for matrix compression (see, e.g., [50] for applications to image compression). This is done by considering only the first k largest (magnitude-wise) singular values and corresponding singular vectors, which results in a lower rank matrix. This lower rank matrix is guaranteed to be the best possible approximation of the original matrix for its rank[13] [16]. Likewise, we may consider only the k largest contributing terms in the Singular Value Decomposition of the Hamiltonian at each pass of the SVD Decomposition scheme (see above). We refer to this as a k-SVD Based Decomposition [14]. This approach further reduces the number of ancillary qubits to a maximum of k in each pass, but at the cost of only approximating (with a variable degree of precision) the Hamiltonian to be mapped.

The success of this technique will, therefore, vary with the original Hamiltonian, and how well it can be approximated with the k-major components. On the other hand, the usefulness of this technique is clearer in practice; see section 6.

---

[13]"Best possible approximation" for a lower rank matrix is defined in terms of the Frobenius norm $\|A\|_F = \sqrt{\sum_{ij} |A_{ij}|^2}$, such that the best approximating matrix B of rank lower than A is $\min_B \|A - B\|_F$.

[14]Note also that we may now refer to the technique presented in section 4.1.2 as ∞-SVD Decomposition, to highlight that there is no cutoff imposed (as opposed to in k-SVD Decompositions).

4.2 Partition "Styles"

We have, thus, specifically addressed how to employ the Schrieffer-Wolff transform to reduce the locality of H. However, we still need to perform a partition of the reduced-locality Hamiltonian, which is to say, choose groups of terms of the Hamiltonian that involve $\leq$ the number of qubits available for computation. For example, if the post-decomposition Hamiltonian is of the form (reusing previous notation)

$$H' = h_{123} + h_{345} + h_{45} + h_{56}$$

and there are three qubits available for computation, then we may calculate a lower bound to H' by any of the following

(i) $\mathbf{LB}\,[H'] = \mathbf{LB}\,[h_{123}] + \mathbf{LB}\,[h_{345}] + \mathbf{LB}\,[h_{45}] + \mathbf{LB}\,[h_{56}]$

(ii) $\mathbf{LB}\,[H'] = \mathbf{LB}\,[h_{123}] + \mathbf{LB}\,[h_{345} + h_{45}] + \mathbf{LB}\,[h_{56}]$

(iii) $\mathbf{LB}\,[H'] = \mathbf{LB}\,[h_{123}] + \mathbf{LB}\,[h_{345}] + \mathbf{LB}\,[h_{45} + h_{56}]$

($\mathbf{LB}\,[H]$ denoting a lower bound to the ground state energy of Hamiltonian H)

We could, for example, calculate *all* of the above lower bounds, and choose the highest energy (as the best approximation in value to the original ground state). However, the number of possible combinations grows quickly; for n k-local terms there are $S(n,k) = \sum_{i=1}^{k} \binom{n}{i}$ possible groupings of those terms such that each group involves at most k systems. The sum of binomial coefficients has no known closed form [24, p. 165], but lower, upper and tight bounds have been proposed (e.g. [49]). An asymptotic tight bound may be obtained (so as to give a concrete idea of the growth) by approximating the sum to a geometric series[15]:

$$\sum_{i=1}^{k} \binom{n}{i} = \Theta\left(\left(1 - \frac{2k}{n}\right)^{-1} \binom{n}{k}\right)$$

On the other hand, finding the partition that maximizes the lower bound to the energy corresponds to optimally partitioning a hypergraph, where each hyperedge corresponds to a term of the Hamiltonian, and each vertex corresponds to a body. Finding graph partitions is generally an NP-hard problem [17, 7, 20]. Therefore, we take a heuristic position: that minimizing the number of "cuts" (i.e., number of separate terms to determine the ground state energy of) maximizes the lower bound energy, and conversely that maximizing the number of cuts minimizes the ground state energy lower bound. The latter case is trivially given by finding the ground state of each of the terms of the Hamiltonian, individually; we will refer to this partition scheme

---

[15]Following the discussion in [26]; it is relatively easy to prove that $\left(1 - \frac{2k}{n}\right)^{-1}\binom{n}{k} - 1$ is an upper bound to $S(n,k)$:

$$\frac{\binom{n}{k}}{\binom{n}{k-1}} = \frac{k}{n-k+1} \overset{\text{def}}{=} x > \frac{k-i}{n-k-i+1} = \frac{\binom{n}{k-i}}{\binom{n}{k}} \qquad x \leq \frac{2k}{n}$$

$$\Rightarrow \sum_{i=0}^{k} \binom{n}{i} \leq \binom{n}{k} \sum_{j=0}^{k} x^j \leq \binom{n}{k} \frac{1}{1-x} \leq \binom{n}{k}\left(1 - \frac{2k}{n}\right)^{-1}$$

That it is an asymptotic lower bound as well will not be proven.

as **Direct Partitioning**. On the other hand, a method to determine a partition with a minimal amount of cuts (**Minimal Partition**) is here proposed. Consider the following example:

A Hamiltonian H is 3-local and involves 4 subsystems, such that, in line with previous notation:

$$H = h_1 + h_2 + h_4 + h_{12} + h_{14} + h_{24} + h_{134}$$

We notice that if one is searching for the 3-sized partition (i.e., the sets of at-most-3 systems to consider together) with the least number of groupings, it is unnecessary to consider groupings with less than the maximum allowed number of elements; in this case, we need only consider the partition

$$\{\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,3,4\},\{1,3,5\},\{1,4,5\},\{2,3,4\},\{2,3,5\},\{2,4,5\},\{3,5,6\}\}$$

...or a subset thereof. Any other partition will contain a grouping of a single subsystem, which we may always ignore in favour of a larger grouping that can include the term (e.g., $\{1\}$ can be ignored in favour of $\{1,2,3\}$, $\{3,4\}$ in favour of $\{1,3,4\}$, etc.). It becomes then a matter of finding which of the groupings in the relevant partition contain no terms (and so are suppressed); we do this by considering the larger-locality terms of H first and taking a 'score' approach, so that we suppress groupings where possible.

We start by sorting the terms of H in descending order of number of involved subsystems...

$$h_{134}, \ h_{24}, \ h_{14}, \ h_{12}, \ h_1, \ h_2, \ h_4$$

...and distribute the terms among the available groupings, taking care to always choose the least empty groupings:

(I)

$h_{134} \longrightarrow \{1,3,4\}$

(II)

$h_{24} \xrightarrow{\frac{1}{3}} \begin{array}{l} \{1,2,4\} \\ \{2,3,4\} \\ \{2,4,5\} \end{array}$

(III)

$h_{14} \xrightarrow{\frac{1}{2}} \begin{array}{l} \{1,2,4\}\,_{h_{24}} \\ \{1,3,4\}\,_{h_{134}} \\ \cancel{\{1,4,5\}} \end{array}$

(IV)

$h_{12} \longrightarrow \begin{array}{l} \{1,2,4\}^{\,h_{14}}_{\,h_{24}} \\ \cancel{\{1,2,3\}} \\ \cancel{\{1,2,5\}} \end{array}$

(V)

$h_1 \longrightarrow \begin{array}{l} \{1,2,4\}^{\,h_{14}}_{\,h_{24}} \\ \cancel{\{1,3,4\}}\,_{h_{134}} \\ \cancel{\{1,2,5\}} \\ \cancel{\{1,2,3\}} \\ \cancel{\{1,3,5\}} \\ \cancel{\{1,4,5\}} \end{array}$

(VI)

$h_2 \longrightarrow \begin{array}{l} \{1,2,4\}^{\,h_{14}}_{\,h_{24}} \\ \cancel{\{2,3,4\}}\,_{h_{24}} \\ \cancel{\{2,4,5\}}\,_{h_{24}} \\ \cancel{\{2,3,5\}} \\ \cancel{\{1,2,3\}} \\ \cancel{\{1,2,5\}} \end{array}$

(VII)

$h_4 \longrightarrow \begin{array}{l} \{1,2,4\}^{\,h_{14}\,h_1}_{\,h_{24}\,h_2} \\ \cancel{\{1,3,4\}}^{\,h_{134}}_{\,h_{14}} \\ \cancel{\{2,3,4\}}\,_{h_{24}} \\ \cancel{\{2,4,5\}}\,_{h_{24}} \\ \cancel{\{1,4,5\}} \end{array}$
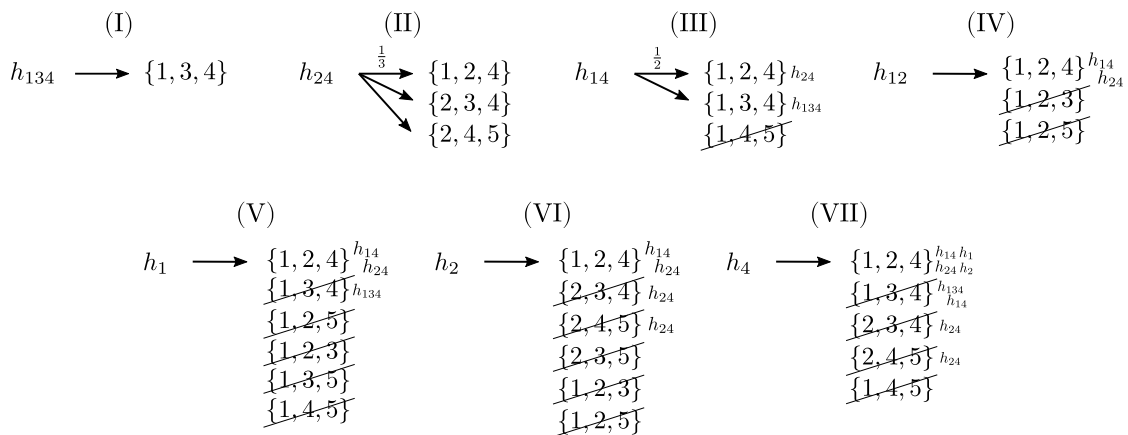
So that, in this case, the considered partitioning is:

$$\{1,2,4\} \ \leftarrow \ \tfrac{h_{24}}{3} + \tfrac{h_{14}}{2} + h_{12} + h_1 + h_2 + h_4$$

$$\{1,3,4\} \ \leftarrow \ h_{134} + \tfrac{h_{14}}{2} \ , \ \ \{2,3,4\} \ \leftarrow \ \tfrac{h_{24}}{3} \ , \ \ \{2,4,5\} \ \leftarrow \ \tfrac{h_{24}}{3}$$

which indicates a lower bound to the ground state energy of H is to be calculated as

27

$$\mathbf{LB}\left[H\right] = \mathbf{GSE}\left[\frac{h_{24}}{3} + \frac{h_{14}}{2} + h_{12} + h_1 + h_2 + h_4\right] + \mathbf{GSE}\left[h_{134} + \frac{h_{14}}{2}\right] + \mathbf{GSE}\left[\frac{h_{24}}{3}\right] + \mathbf{GSE}\left[\frac{h_{24}}{3}\right]$$

(Again using $\mathbf{LB}\left[x\right]$ to denote a lower bound to the ground state energy of x, and $\mathbf{GSE}\left[y\right]$ to denote the ground state energy of y)

The method, as applied to a general N-system Hamiltonian to be partitioned in k-system partitions, can be summarized as follows:

---

1. Sort the terms of the N-system Hamiltonian in descending order of number of systems involved in each term,

2. For each possible k-sized partition, assign it a score of 0,

3. Let the next term of the Hamiltonian (in descending order of number of systems involved) be h_α, where α denotes the subsystems affected by the term,

4. Sort the partitions in descending score,

5. Assign h to all partitions of maximum score that can "fit" h_α; when more than one partition of maximum score can contain the term, consider a term of $\frac{1}{\langle\text{\# of possible partitions}\rangle} \cdot h_\alpha$ for each partition (i.e., "distribute" h_α between the different possible partitions),

6. For each partition that was assigned a term, increase its score by 1,

7. If more terms remain, go to 3.

We should also at this point mention that even though we estimate that a smaller number of cuts will lead to a higher valued lower bound of the ground state energy, we again have no guarantees regarding how/if the structure (in the previously discussed sense) of the ground state energy curve is preserved under the different partitions. It may be that a partition that produces a "worse" (i.e., algebraically lower valued) lower bound preserves the structure better than a "better" (i.e., algebraically higher valued) lower bound; eq. (13) gives no guarantees in this regard. Therefore, this is to be investigated in practice.

5 On the Number of Electrons in *Ansatz*-Proposed States

One of the first steps of this work is taking the Born-Oppenheimer Hamiltonian and converting it to a second-quantized language. This sidesteps the issue of (anti-)symmetrization, and prepares the operator to be converted to a quantum-computer–friendly form via a Jordan-Wigner transform. The Hamiltonian, in this 2nd-quantized form, preserves the number of electrons:

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s \qquad\qquad \hat{n}_i = a_i^\dagger a_i$$

(Second quantized Hamiltonian)   (Number operator for orbital i)

$$\{a_i, a_j\} = \left\{a_i^\dagger, a_j^\dagger\right\} = 0 \qquad \left\{a_i, a_j^\dagger\right\} = \delta_{ij}$$

(Cannonical commutation relations for the creation/annihilation operators)

For the $h_{pq}a_p^\dagger a_q$ terms of the Hamiltonian. . .

$$\left[\hat{n}_i, \sum_{pq} a_p^\dagger a_q\right] = \sum_{pq} a_i^\dagger a_i a_p^\dagger a_q - \sum_{pq} a_p^\dagger a_q a_i^\dagger a_i$$

$$= \sum_{pq} a_i^\dagger (\delta_{ip} - a_p^\dagger a_i) a_q - \sum_{pq} a_p^\dagger a_q a_i^\dagger a_i$$

$$= \sum_q a_i^\dagger a_q - \sum_{pq} a_i^\dagger a_p^\dagger a_i a_q - \sum_{pq} a_p^\dagger (\delta_{iq} - a_i^\dagger a_q) a_i$$

$$= \sum_q a_i^\dagger a_q - \sum_p a_p^\dagger a_i - \sum_{pq} a_i^\dagger a_p^\dagger a_i a_q + \sum_{pq} a_p^\dagger a_i^\dagger a_q a_i$$

$$= \sum_q (a_i^\dagger a_q - a_q^\dagger a_i)$$

$$\left[\sum_i \hat{n}_i, \sum_{pq} a_p^\dagger a_q\right] = (i \text{ becomes a dummy index}) = 0$$

Note how the number of electrons for a particular orbital is not necessarily unchanged. Likewise for the $\sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s$ terms, the commutator with the total number operator becomes, when brought to normal ordering,

$$\left[\sum_i \hat{n}_i, \sum_{pqrs} a_p^\dagger a_q^\dagger a_r a_s\right] = \sum_{ipqrs} a_i^\dagger a_i a_p^\dagger a_q^\dagger a_r a_s - \sum_{ipqrs} a_p^\dagger a_q^\dagger a_r a_s a_i^\dagger a_i =$$

$$= \sum_{iqrs} (a_i^\dagger a_q^\dagger a_r a_s) - \sum_{iprs} (a_i^\dagger a_p^\dagger a_r a_s) + \sum_{ipqrs} (a_i^\dagger a_p^\dagger a_q^\dagger a_i a_r a_s) -$$

$$- \sum_{ipqrs} (a_p^\dagger a_q^\dagger a_r a_i) + \sum_{ipqs} (a_p^\dagger a_q^\dagger a_s a_i) - \sum_{ipqrs} (a_p^\dagger a_q^\dagger a_i^\dagger a_r a_s a_i) =$$

(Renaming dummy indexes)

$$= 0$$

Such that H and $\sum_i \hat{n}_i$ commute. This is physically expected; we expect the number of electrons to remain unchanged. However, the number of electrons being preserved does not guarantee all physical constraints; we are interested in the lowest-energy state of a given Hamiltonian *given* a fixed number of electrons, i.e., for a *specific ion*. If we accept any state of the Fock space as a candidate to the molecule's ground state, we may find *the most stable ion* of the molecule to be the most stable configuration in the Fock space, or even superpositions of states belonging to various ions!

We are, thus, interested in searching in a constrained Fock space of N electrons. This is best discussed, in the context of this work, for two situations of Q and N, where Q denotes the number of available qubits and N the number of electrons to consider.

5.1 $Q \geq N$

If $Q \equiv N$ then, per the variational technique previously described, there is no search to be made. We can only encode a simple Hartree-Fock state into the quantum computer, and evaluate the corresponding energy:

$|11\cdots1\rangle$ preparation   Observe value

Act with Hamiltonian
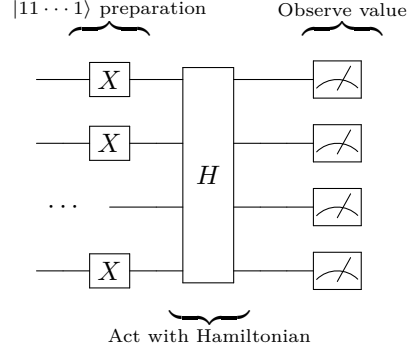
If $Q > N$, then the variational approach proposed does apply, as the "excess" qubits become available to represent excitations. In other words, there is a physically valid subspace of the Fock space to explore, which are all the states

$$\left\{ |\Phi\rangle \in \mathcal{F}^Q \;\; \Big| \;\; (\sum_i \hat{n}_i) |\Phi\rangle = N |\Phi\rangle \right\}$$

What is non-trivial is to parameterize this subspace, i.e., to find an *Ansatz* that correctly produces only such constraint-observing states. Gard et al. [19] propose precisely such an *Ansatz*. In algorithm 5 we present an algorithm to construct a general Q-qubit, N-electron *Ansatz*, per [19] (see, in particular, fig. 5 of the reference). The reasoning behind the algorithm is straightforward: we distribute the excitation gates (Pauli X) as evenly as possible in the circuit, and then connect the excited qubits with the remaining wires via A gates. This connection pattern is stored (as an ordered list of wires to place A gates on), such that it may be repeated as many times as specified by a "layers" parameter (L).
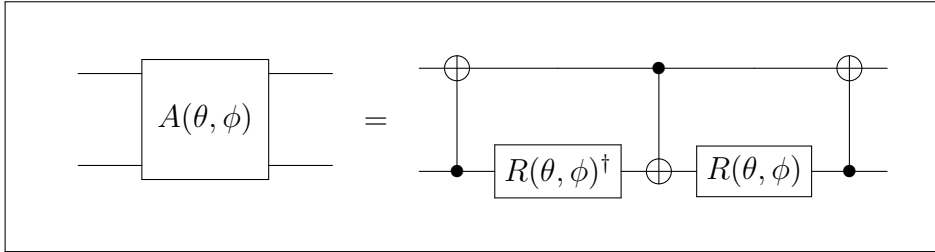
Figure 7: Definiton of the A quantum gate, per [19]. Note that the gate can be decomposed into (controlled) one-qubit gates, such that it may be simulated using QOP, but that care must be taken when parameterizing these gates, as there are only two free parameters.

With this *Ansatz* established, we may search the space of excitations for the ground state of a particular ion using the previously described quantum variational algorithm approach.

5.2 $Q < N$

For the case where a trial state of the full Hamiltonian cannot be encoded into the quantum computer, the situation is not so clear. As previously discussed, and per the technique proposed

Algorithm 5 Producing a Q-qubit, N-electron *Ansatz*. See fig. 7 for the definition of the A gate.

   function ADD(GATE, QUBIT)   ▷ Adds a gate to (the lowest affected index) qubit
Require: Q ▷ Number of qubits
Require: N ▷ Number of electrons
Require: L ▷ Number of layers

   Start
   $x \leftarrow []$
   for $i \leftarrow ($**take** N **from** $0, \lfloor \frac{Q}{N} \rfloor, 2\lfloor \frac{Q}{N} \rfloor, \cdots)$ **do**
      ADD(X, i)
      $x \leftarrow [\ldots x, i]$
   end for
   $a \leftarrow []$
   for $i \in x$ **do**
      if $i \neq q - 1$ **and** $i + 1 \notin x$ **then**
         $a \leftarrow [\ldots a, i]$
      end if
   end for
   $k \leftarrow a$
   while $k \neq []$ **do**
      $k' \leftarrow []$
      for $i \in k$ **do**
         if $i > 0$ **and** $i - 1 \notin x$ **and** $i - 1 \notin a$ **then**
            $a \leftarrow [\ldots a, i - 1]$
            $k' \leftarrow [\ldots k', i - 1]$
         end if
      end for
      for $i \in k$ **do**
         if $i < q - 2$ **and** $i + 2 \notin x$ **and** $i + 1 \notin a$ **then**
            $a \leftarrow [\ldots a, i + 1]$
            $k' \leftarrow [\ldots k', i + 1]$
         end if
      end for
      $k \leftarrow k'$
   end while
   for $l \leftarrow 1, \cdots, L$ **do**
      for $i \in a$ **do**
         ADD(A, i)
      end for
   end for
   End

and tested in this work, we may give up on calculating the ground state energy, and rather observe the relationship between the geometric (or otherwise physical) parameters of the Hamiltonian and a lower bound to the associated ground state energy. Within certain locality conditions we may obtain such a lower bound by considering the ground state energy of terms of the Hamiltonian, separately. Already at this point it is not obvious what constraints to impose on the ground states of the separate sub-Hamiltonians; consider, as an example, a 2-local Hamiltonian involving 4 qubits, which we represent as (in line with previously used notation):

$$H = h_{12} + h_{34}$$

If we are searching for the 3-electron ground state of H, then the candidate states are

$$|1110\rangle, |1101\rangle, |1011\rangle, |0111\rangle$$

or superpositions thereof. Note that in the subspace of h_{12}, this means that $|10\rangle$, $|01\rangle$, and $|11\rangle$ are all valid states of the first half of the system, i.e., may be part of a full-system state that satisfies the electron count constraint. However, they do not satisfy the constraint by themselves. Furthermore, superpositions of the 4-qubit states result in entanglement, so that no superposition may be considered in the $1, 2$ subspace without violating the number constraint. In summary, these limitations relate to the loss of cross-term relationships when considering terms separately.

The situation is even less clear once a Schrieffer-Wolff transform is applied to H: as discussed, this transform (via the successive application of "gadgets") does not map H to a matching-spectrum Hamiltonian, but rather to some different system's Hamiltonian H' that matches the energy spectrum of H only in the low energy subspace of the spectrum. At its limit, we may observe that only the ground state energy of H' matches the ground state energy of H (within bounded error). Now, it is not even clear what constraints to impose on the new system; the S.W. transform concerns itself with ground state energies, not with ground states. An auxiliary body is introduced (which cannot be directly identified with an orbital) and the other qubits of the system may no longer be regarded as straightforwardly related to orbital occupations. Regardless, once sub-terms of H' are taken separately, imposing any constraint presents the same difficulties as discussed for H.

We approach these difficulties by disregarding electron count constraints completely, which in terms of the *Ansatz* used in the QVA approach, translates into spanning as much of the Hilbert space as possible (see fig. 4). The reasoning behind this approach is that, by the variational principle, disregarding electron count constraints may only result in a lower-bound to the correct ground state energies. As we are disregarding a constraint to what would already be a lower-bound (when considering the terms of the Hamiltonian separately), we are effectively taking the lower bound of a lower bound. This decision is thus in line with the rest of the work; we seek to investigate whether, in practice, calculating such a lower bound may yield information on the geometry parameters, when to calculate the actual energy (or, in this case, enforce the electron number constraints) is not possible. We do expect, however, the technique to be insensitive to ionization, as the number of electrons stops being encoded anywhere.

6 Results

The produced C/Python library is publicly accessible, and can be found in the public 'QOP' GitHub repository[16].

---

[16]`https://github.com/mikeevmm/QOP`

We chose as testbeds for the method proposed in section 4 molecules of H_2, HLi and O_2, for the following reasons:

- All molecules are diatomic, thus having a single, well-defined, geometric parameter: bond-length;

- H_2 is a very simple and symmetric molecule, and thus requires few qubits to exactly express the Hartree-Fock Hamiltonian. This means that both the Jordan-Wigner and the Schrieffer-Wolff Hamiltonians can be exactly diagonalized, enabling a comparison between the H.F. energy and the S.W. energy, and between the S.W. energy and its lower bound;

- HLi is still a small molecule, but is asymmetric, so may highlight the effect of asymmetry;

- O_2 is a larger molecule.

For the production of the (qubit) Hamiltonian, we employed OpenFermion [32] and Psi4 [37] (using the OpenFermion-Psi4 package to interface the two [14]). The use of these packages resulted in a Hamiltonian, which was then treated with custom-made code. The code's logical flowchart is presented in fig. 8. The full code is available online at `https://github.com/mikeevmm/TeseMestrado/blob/master/lowerbound.py`.

As previously discussed, there are different possible approaches to locality-lowering of a Hamiltonian ("decomposition" techniques) and to lower bound calculation of the corresponding ground state energy ("partition" techniques). Ideally, the effect of each of these approaches would be separately studied: for each decomposition scheme used, the ground state energy of the *full* resulting Hamiltonian operator would be calculated by, for example, calculating the associated matrix representation, and then finding its lowest eigenvalue with a numerical linear algebra approach. Then, the compound effect of applying different partition schemes would be studied by applying each partition scheme to the result of each decomposition, and comparing the resulting lower bound energy to the ground state energy of the post-decomposition Hamiltonian.

However, this is not possible in practice, in particular due to the number of ancillary qubits introduced by the different decomposition schemes. Even when using a (k-)Singular Value Decomposition based scheme, the number of introduced ancillary bodies is significant, but the issue is particularly pronounced for the direct decomposition scheme. Introducing many auxiliary bodies is reasonable, and expected, as we are ultimately interested in reducing the locality of the original Hamiltonian, but representing the resulting Hamiltonian (as a whole) in matrix form becomes exponentially harder with each extra qubit introduced: the number of elements in the matrix of a q-qubit system is 2^{2q}. This meant that in general, it was not possible to separately study the results of decomposition/partition. However, in figs. 9 and 10 we present a comparison between the ground state energy of post-decomposition Hamiltonians and the energy obtained using a typical Hartree-Fock approach, for systems where the calculation of the ground state energy of the post–Schrieffer-Wolff Hamiltonian was possible. This serves as an illustration of the formal results presented by Bravyi et al. in reference [10], as well as an illustration of how the truncation of singular values (cf. section 4.1.3) can still produce a good approximation to the original Hamiltonian. In fig. 9 we merely present both curves overlapped; in fig. 10 a finer comparison of the two energies was done by using a "loss" function, defined as

$$\text{Loss}(E_{\text{Hartree-Fock}}, E_{\text{Schrieffer-Wolff}}) = 10 \log_{10}\left(\left|\frac{E_{\text{Hartree-Fock}} - E_{\text{Schrieffer-Wolff}}}{E_{\text{Hartree-Fock}}}\right|\right) \text{ (dB)} \quad (15)$$
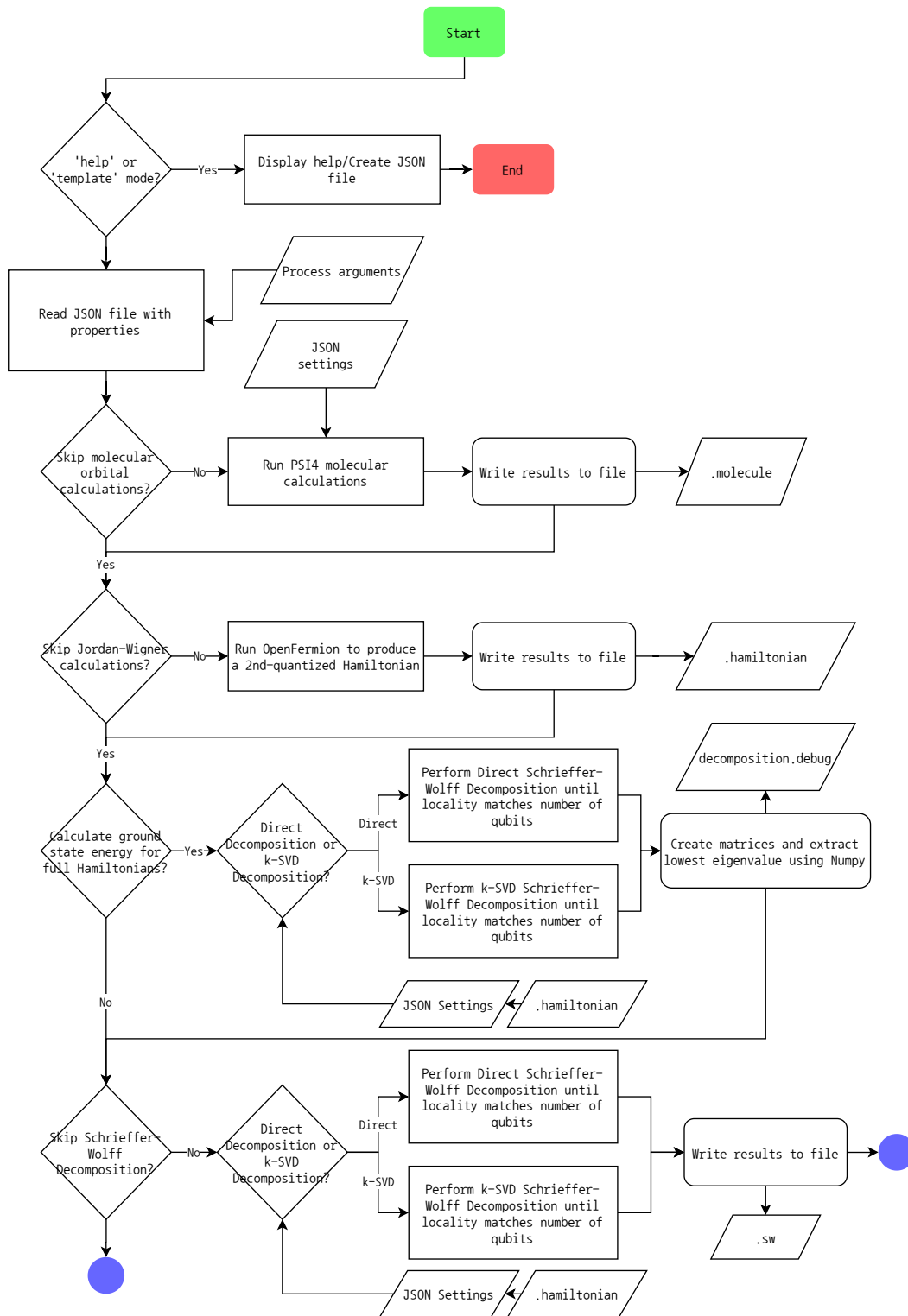
(a) Logical flowchart (part I, see fig. 8b)
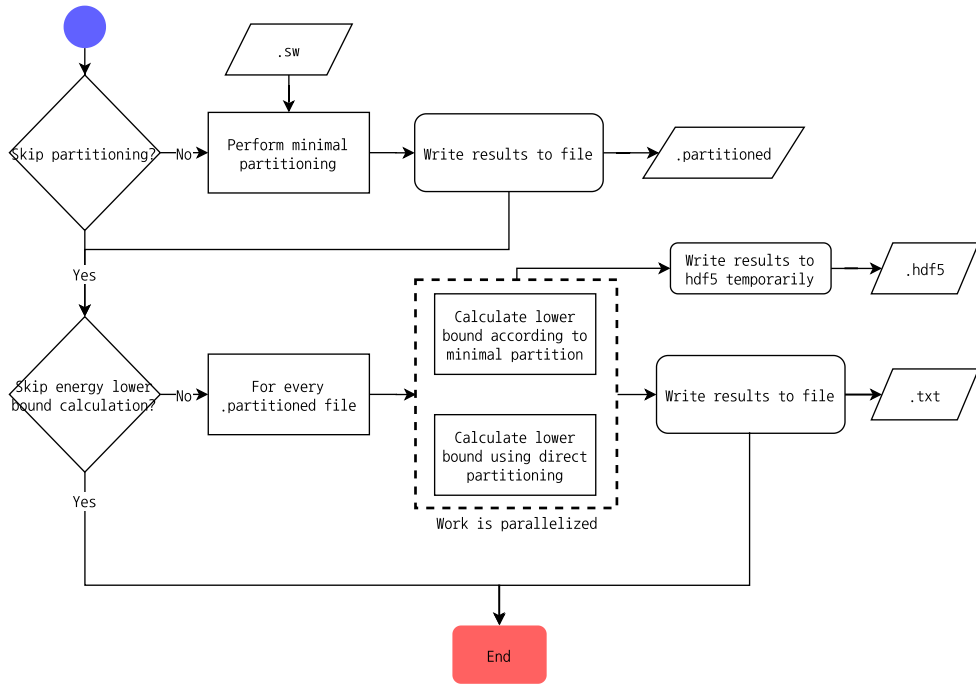
(b) Logical flowchart (part II, see fig. 8a)

Figure 8: Logical flowchart of the developed code implementing the technique described in section 4. The flowchart is split into figs. 8a and 8b due to its size; the circle indicates a continuation from the other subfigure. The respective code can be found online at `https://github.com/mikeevmm/TeseMestrado/blob/master/lowerbound.py` (Python). It is taken as convention that diamonds represent decisions, lozenges represent data, and rectangles represent actions (rounded rectangles are used for sequence-terminating actions).

For each decomposition process, the compound result of different partition schemes are presented. Finally, because calculating the lowest eigenvalue of a given Hamiltonian using a variational approach may induce some error or even silently fail (by either not converging or converging to the wrong value), we first present results that use Numpy [36] to obtain the ground state energy. This is done by converting the relevant term to a matrix form and computing the lowest eigenvalue (in either a dense or sparse scheme). Then, we present a comparison between these results and the results obtained by simulating a quantum variational approach to obtain ground state energies.

Before presenting the results of the given original technique, we first illustrate the relevance of the high-locality terms for the different considered systems, in terms of the 'at-rest' parameters. This is done by calculating the ground state energy of a new Hamiltonian, calculated for each system by discarding terms of the original Hamiltonian above the intended locality. This avoids any need for decomposition and/or partition. For systems where the high-locality terms are determinant, ignoring these terms should result in different minimizing geometric parameters, whereas for systems for which high-locality terms are of little relevance, ignoring them should still produce a reasonable approximation to the minimizing parameters. The obtained results are presented in fig. 11.

6.1 Ground State Energy From Matrix

In fig. 12, we present the obtained results for a direct decomposition scheme (cf. section 4.1.1) combined with both direct and minimal partitioning schemes (cf. section 4.2). We present results for all systems in study, i.e., H_2, HLi and O_2.

Likewise, in fig. 13 we present the obtained results for an ∞-SVD based decomposition scheme (cf. section 4.1.2) but coupled with both direct and minimal partitioning only for the H_2 and HLi systems. For the O_2 system, the post–Schrieffer-Wolff transformation Hamiltonian (using an ∞-SVD based decomposition scheme) involves ~ 300 qubits, which requires more memory than available for the calculation of the minimal partition (using our technical implementation); as such, for this system, only the direct partition result is presented.

In fig. 14 we present the obtained results for a k-SVD based decomposition scheme. Now, the previous limitation for the O_2 system does not apply, such that for every system, we present a k-SVD based decomposition coupled with both direct and minimal partitioning. For O_2 we present results for both 1-SVD and 3-SVD decompositions. Note how the resulting 'lower bound' is no long necessarily an actual lower bound to the original Hartree-Fock energy; as we have approximated the Hamiltonian at each step, it should be a lower bound to the approximated Hamiltonian's ground state energy.

6.2 Ground State Energy From Simulated Quantum Variational Algorithm

We first present a practical comparison of the different optimization algorithms considered in section 1.3; we present the evolution of an optimization over a pathological surface as guided by, respectively, ADAM and ADADELTA and L-BFGS (cf. section 1.3). The considered optimization function was a Rosenbrock surface [43], which is designed to have a minimum in the middle of a long and narrow valley, such that it may be relatively easy to converge towards the valley, but hard to converge along the valley towards the true global minimum. The Rosenbrock surface is presented in figure fig. 15, and is given by

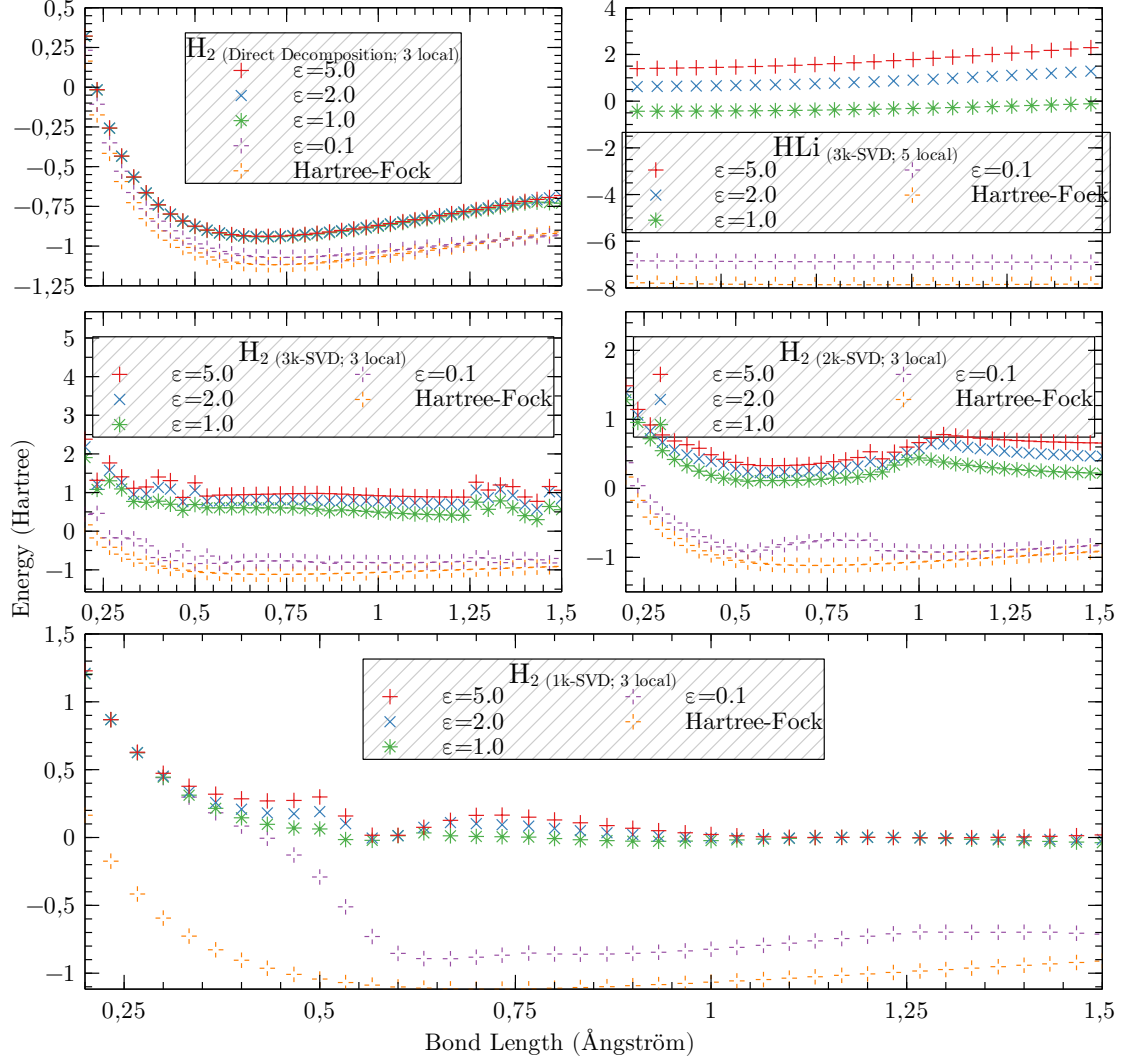$$\mathcal{R}(x, y) = (1 - x)^2 + 100 \cdot (y - x^2)^2 \tag{16}$$

Figure 9: Comparison of the Hartree-Fock ground state energies with the ground state energies associated with the (full) post–Schrieffer-Wolff (SW) transformation Hamiltonian, for different bond lengths, systems and decomposition schemes (as discussed on section 4). Various ϵ values were considered for the Schrieffer-Wolff transformation; see eqs. (9) and (11) for the relationship between ϵ and the asymptotic error between the S.W. Hamiltonian's ground state energy and the original Hamiltonian's ground state energy. Only a few cases are considered, as the requirements for calculating the post-SW Hamiltonian's ground state energy quickly become intractable; see discussion in section 6.
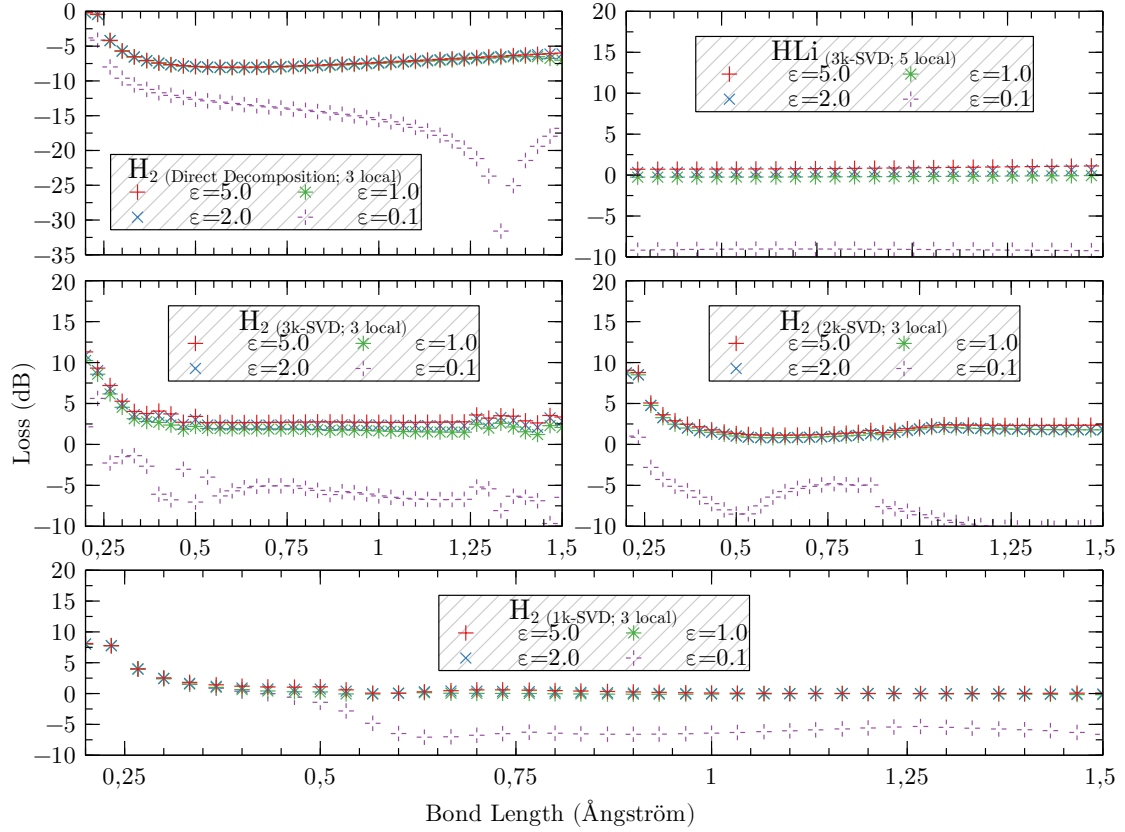
Figure 10: Comparison of the curves presented in fig. 9, in terms of the loss function defined in eq. (15). This mode of comparison is a finer measure of how well the ground state of each post–Schrieffer-Wolff Hamiltonian approximates the Hartree-Fock energy. The ϵ parameters asymptotically relate to the error between the post–Schrieffer-Wolff Hamiltonian's ground state energy and the original Hamiltonian's ground state energy; see section 1.4, eqs. (9) and (11).
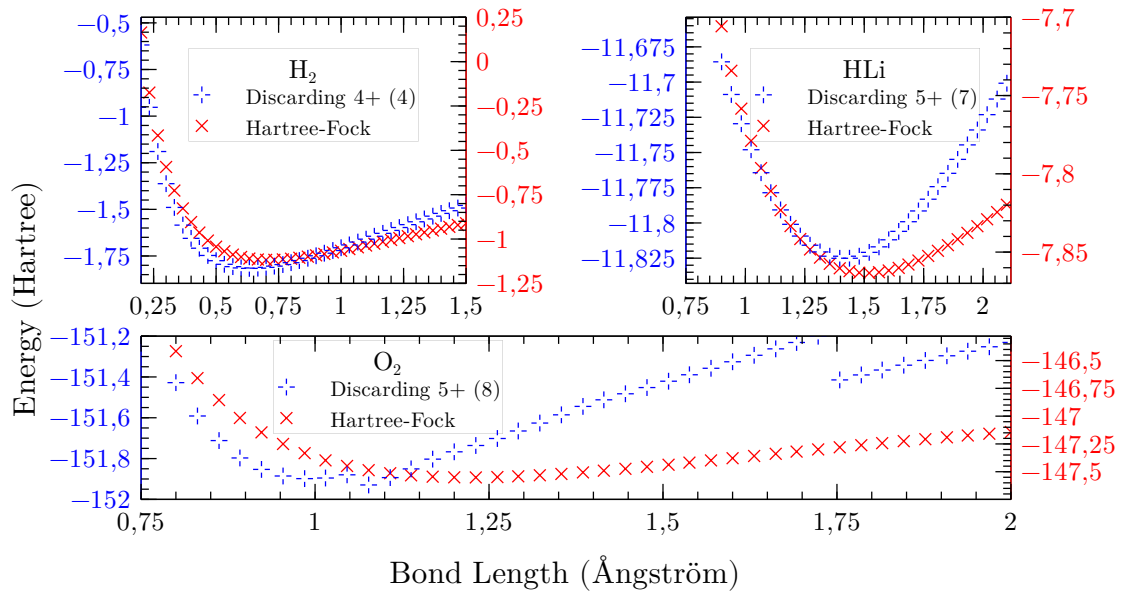
Figure 11: Hartree-Fock energies and ground state energies of the Hamiltonian with high-locality terms suppressed, for different bond lengths, for each considered system (H_2, HLi, O_2). The number of qubits in the Hamiltonian varies according to the bond length, but an approximate average is given in the key, in parenthesis. The vertical scales for the Hartree-Fock energies and the energies resulting from discarding high-locality terms are considered separately (H.F./discarding to the right/left respectively), as we are mainly interested in the minimizing bond length.
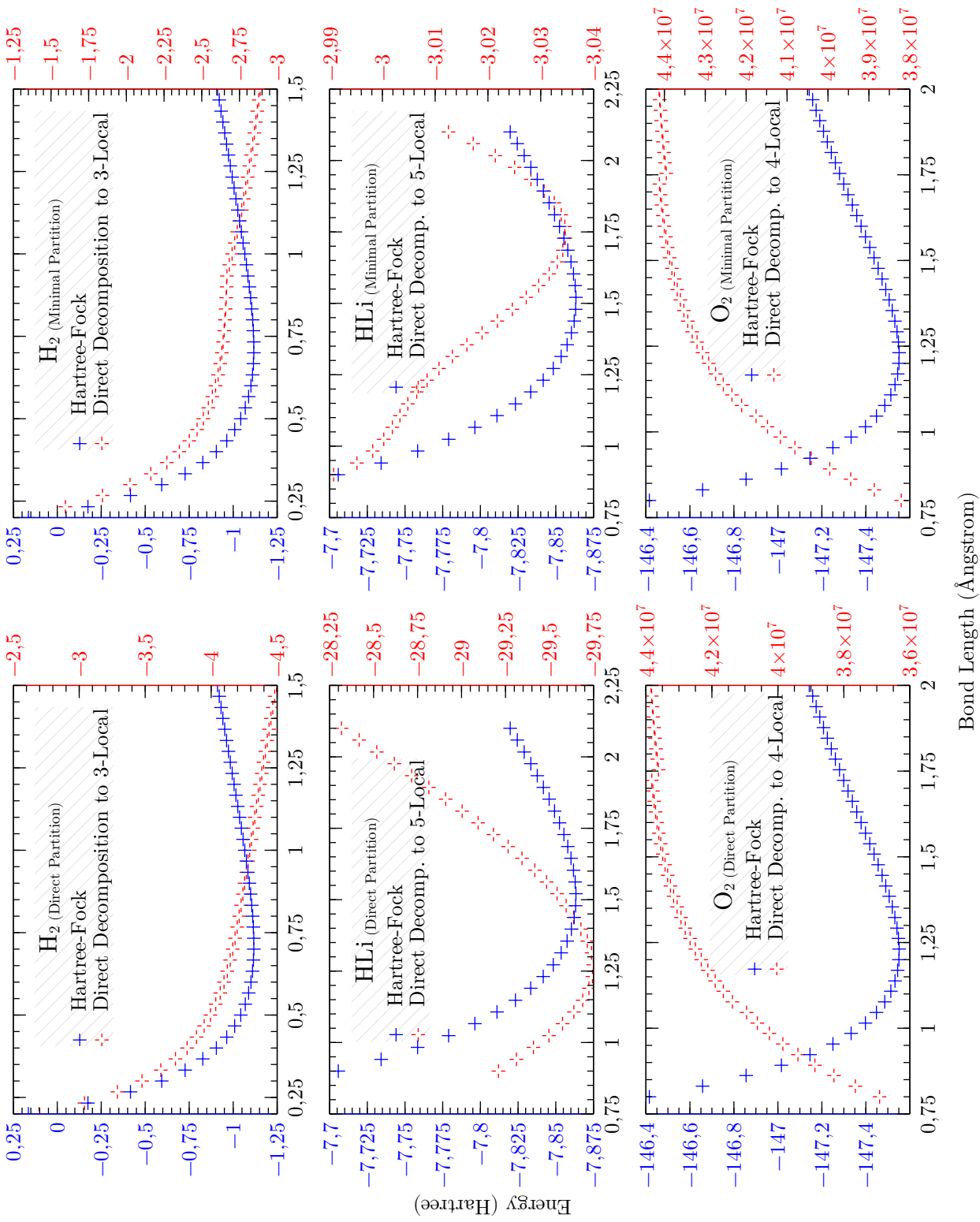
Figure 12: Results obtained using a direct decomposition scheme (cf. section 4.1.1), combined with either a minimal or direct partitioning scheme (cf. section 4.2), as compared to the Hartree-Fock energy of the system, for various bond lengths. The considered systems are H_2, HLi, O_2. The lower bound values here presented were obtained using a *non*-variational approach. The energy scales for the Hartree-Fock (HF) energy and energy lower bound are considered separately, as we are mainly interested in the minimizing bond lengths; the HF energy scale is presented to the left, the lower bound energy scale is presented to the right of each plot. The ϵ parameters, related to the Schrieffer-Wolff transformation's asymptotic error and as defined in eqs. (9) and (11), used in the calculations were: $\epsilon_{H_2} = 0.1$, $\epsilon_{HLi} = 0.5$, $\epsilon_{O_2} = 1.0$.
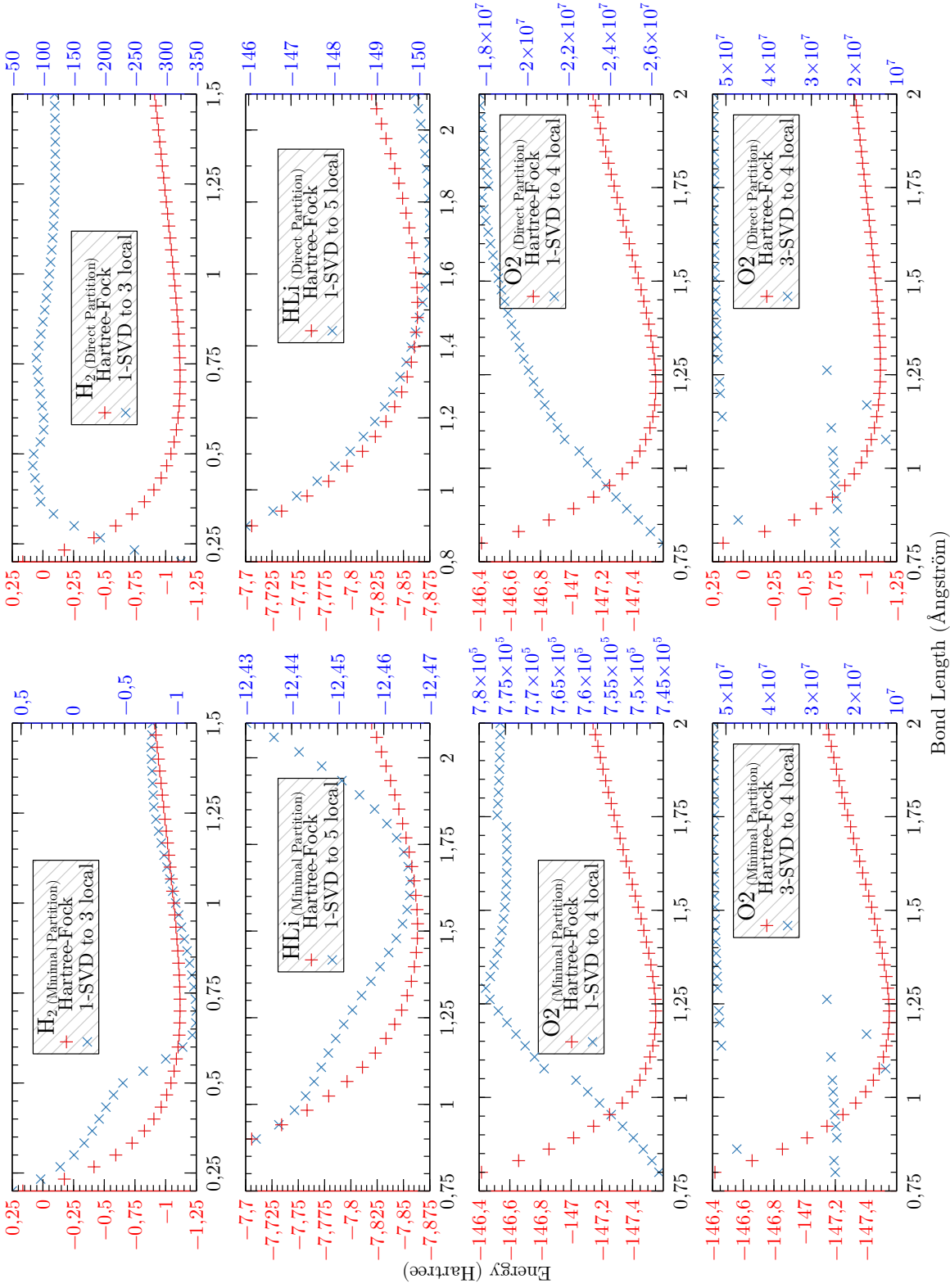
40
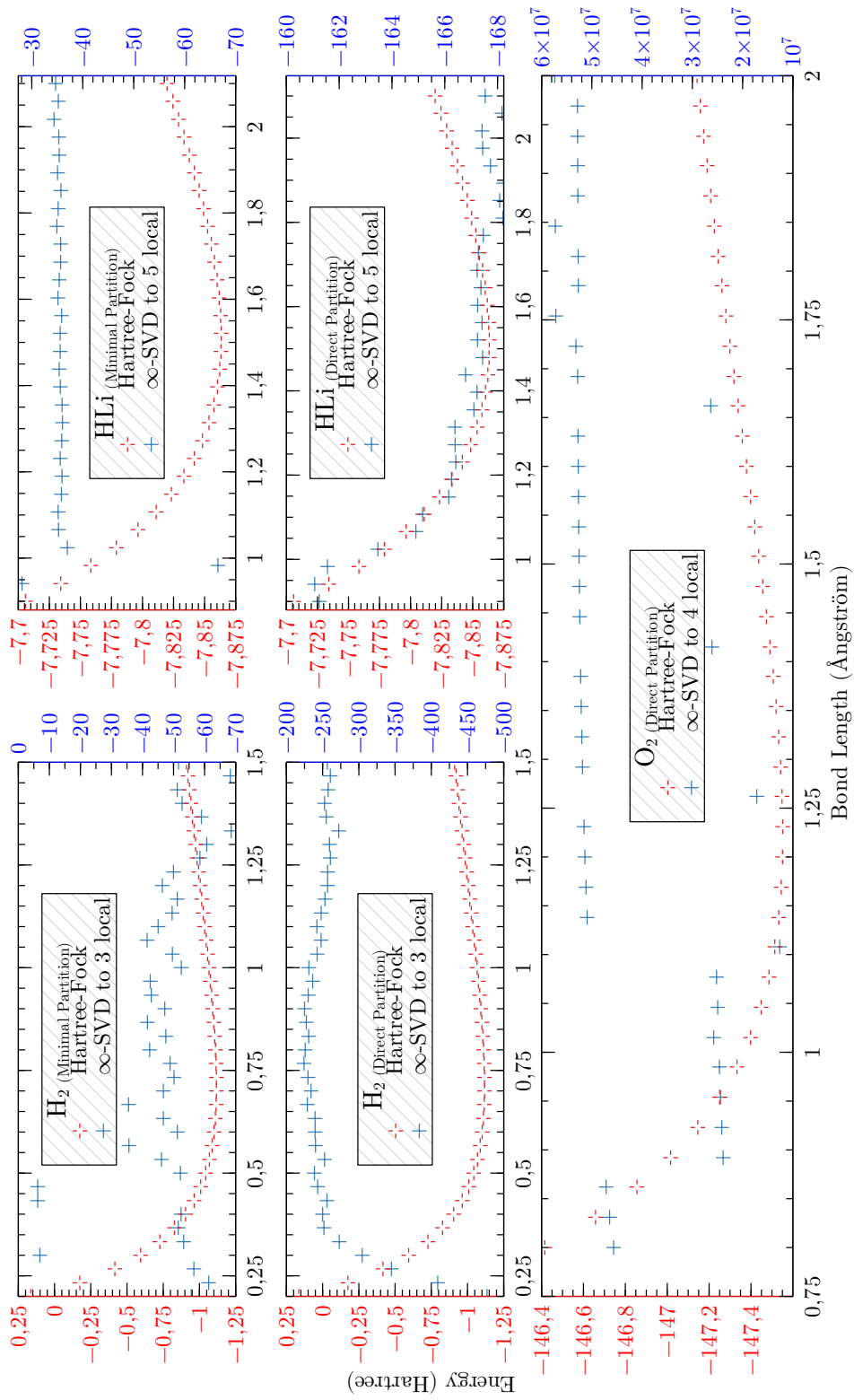
Figure 13: Results obtained using an ∞-SVD based decomposition scheme (cf. section 4.1.2). For the H_2 and HLi systems, this decomposition scheme was coupled with both direct and minimal partitioning schemes. For O_2, the memory requirements for determining the minimal partition were too high, so that only a direct partition result is presented. The lower bound values here presented were obtained using a *non*-variational approach. The scales for the reference Hartree-Fock (HF) energies and the lower bound values are once again separate, as to evidence the structure of each curve and/or minimizing bond lengths; the HF energy scale is presented to the left, the lower bound energy scale is presented to the right of each plot. The ϵ parameters, related to the asymptotic error of the Schrieffer-Wolff transformation and as defined in eqs. (9) and (11), used in the calculations were: $\epsilon_{H_2} = 0.01$, $\epsilon_{HLi} = 0.1$, $\epsilon_{O_2} = 3$.

Figure 14: Results obtained using k-SVD based decomposition schemes (cf. section 4.1.3), coupled with both direct and minimal partitioning schemes. The lower bound values here presented were obtained using a *non*-variational approach. The scales for the reference Hartree-Fock (HF) energies and the lower bound values are once again separate, as to evidence the structure of each curve and/or minimizing bond lengths; the HF energy scale is presented to the left, the lower bound energy scale is presented to the right of each plot. The ϵ parameters, defined in eqs. (9) and (11), and related to the asymptotic error on the Schrieffer-Wolff transformation, that were used were: $\epsilon_{H_2} = 0.01$, $\epsilon_{HLi} = 0.1$, $\epsilon_{O_2;1-SVD} = 0.1$, $\epsilon_{O_2;3-SVD} = 3.0$.

so that the global minimum is located at $\langle 1, 1 \rangle$.
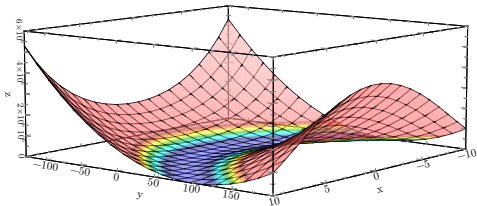
Figure 15: Rosenbrock function with global minimum at $\langle 1, 1 \rangle$. Note that the function is very steep away from the minimum, but very 'flat' in the valley that contains the minimum.

The optimization results are presented in fig. 16; because the L-BFGS algorithm converges to the minimum in much less iterations, its results are presented separately. Like previously done for error measurement, the convergence is measured in terms of a loss function, defined as

$$\text{Loss}(z) = 10 \cdot \log_{10}(1 + z) \tag{17}$$

where z is the value that the cost function (i.e., the Rosenbrock function) takes at each iteration point.

Optimization is considered converged once the cost function remains constant over 15 iterations, and was otherwise stopped after 10 000 iterations.

In fig. 17 we present a comparison between some selected results presented in section 6.1 and the corresponding results obtained using a quantum variational approach, as described in section 1.3, to obtain the lowest eigenvalues for the Hamiltonians at each bond length. For the direct decomposition scheme, the results for different numbers of layers in the *Ansatz* (cf. section 1.3) are presented, to highlight the impact of a greater number of degrees of freedom in the resulting eigenvalue. For results that were not at comparable energy scales (between the variational and non-variational curves), a second graph is presented with separate energy scales for the variational and non-variational curves, so as to highlight potential structure similarities (in the previously defined sense), despite the different scales. Note that, in obtaining these results, the parameters for the *Ansatz* were initially randomized, for each system. Afterwards, for each system, the eigenvalues for each bond length were calculated in ascending bond length order, using the converged *Ansatz* parameters of the last bond length calculated as the starting parameters for the new bond length calculation. The reasoning behind this was to induce an "annealing" effect, assuming that the ground *state* for a certain bond length should not differ greatly from the ground state of a similar bond length.

7 Discussion

We first note from fig. 11, broadly comparing with figs. 12 to 14, that simply discarding the terms of high locality apparently yields more accurate minimizing geometric parameters than the proposed original approach: all of the systems preserve the energy curve structure reasonably well when discarding higher-locality terms, with O_2 the system most 'sensible' to this process. However, one should note that OpenFermion calculates the Hartree-Fock orbitals (via Psi4) as a side-effect of calculating the Hartree-Fock energy, and uses these orbitals as the reference orbital set. Furthermore, a "frozen core" of doubly paired electrons in the lower energy orbitals is considered, so that only electrons outside that frozen core are accounted for. It is therefore expected that, up to a point and with this basis, the higher-locality terms are not absolutely determinant in the energy curve structure of the molecule (but rather produce finer corrections); recalling that in the post–Jordan-Wigner Hamiltonian the qubit state can be informally regarded
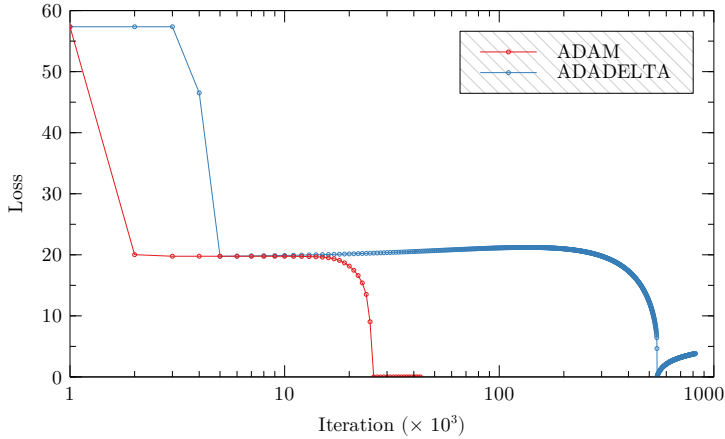
(a) Progress of loss function (eq. (17)) over the iterations of ADAM and ADADELTA optimizations along a Rosenbrock function (eq. (16)). Notice that the horizontal scale is logarithmic, and iterations are counted in thousands.
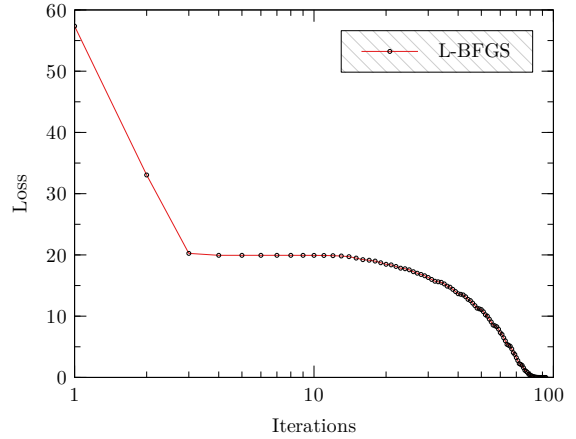
(b) Progress of loss function (eq. (17)) over the iterations of an L-BFGS optimization along a Rosenbrock function (eq. (16)). Notice that the horizontal scale is logarithmic.

Figure 16: Comparison of the performance of the different optimization algorithms chosen for implementation in the QVA simulation (cf. section 3). All algorithms were given $\langle -2.5, 80.0 \rangle$ as the starting point, and used for optimization over the function defined by eq. (16). The closeness to the global minimum is measured using the 'loss' function defined in eq. (17). For ADAM and ADADELTA, the implementations available in the Keras [13] library were used, whereas for L-BFGS, the implementation available in Scipy [47] was used. Note the severe difference in horizontal scale between fig. 16a and fig. 16b; like discussed in section 1.3, there is a tradeoff between scalability and performance. Note also that ADADELTA seems to almost converge to the global minimum, but fails to do so.
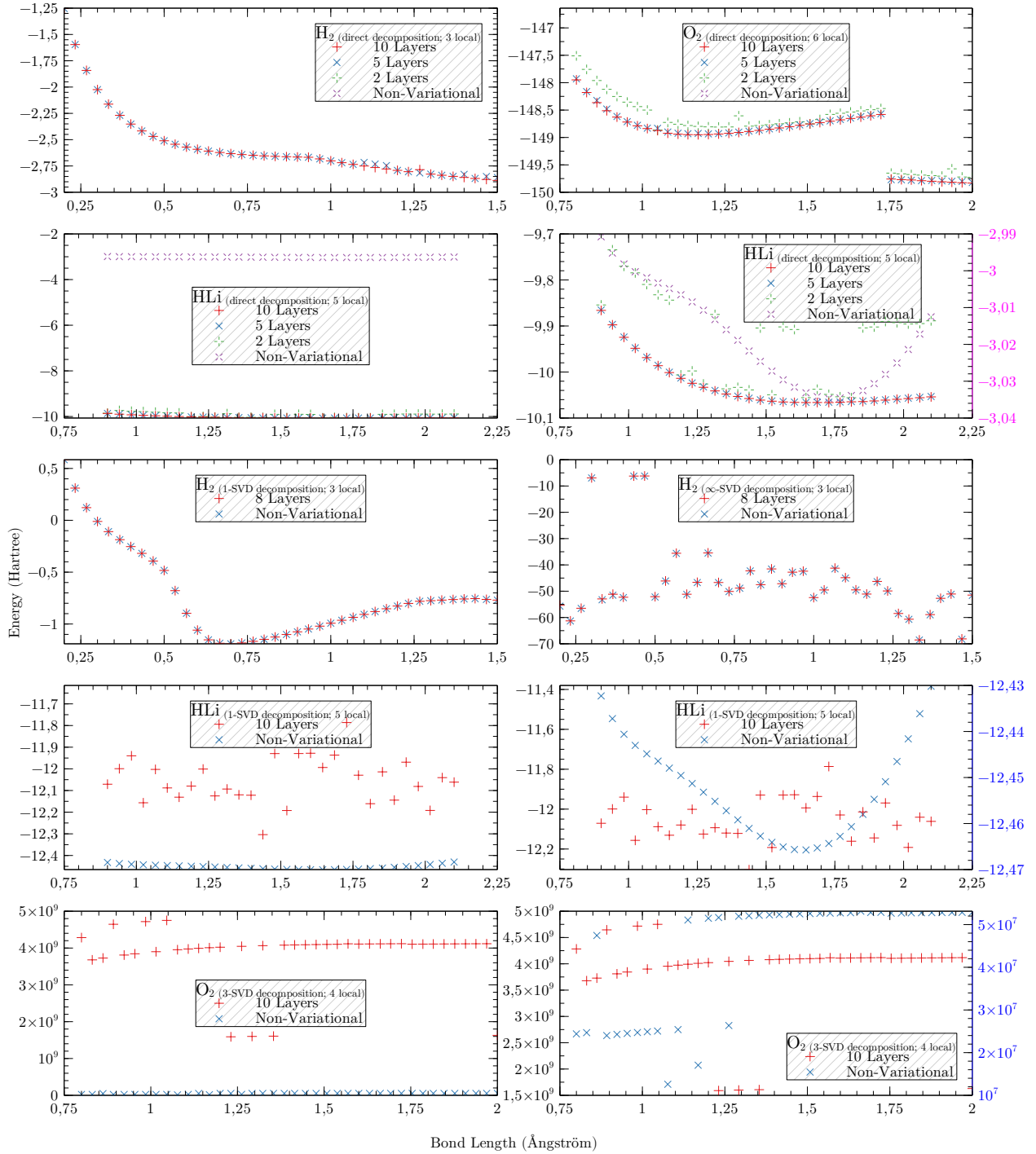
Figure 17: Comparison between the results obtained with a quantum variational approach to the calculation of the lowest eigenvalues of the Hamiltonians (cf. section 1.3) with the corresponding non-variational results presented in section 6.1. Whenever the variational and non-variational results differ greatly in the energy scale, we present also a graph with separate energy scales for the variational and non-variational results, with the scale to the left of the graph corresponding to the variational results, and the scale to the right of the graph (colored) corresponding to the non-variational results.

as the Fock space (cf. eq. (6)), and recalling also the one- and two-electron integrals (eqs. (2) and (3)), we see that these high locality terms will (in informal terms) account for interactions between multiple orbitals for the 'unfrozen' electrons.

However, such an approach implies the ability to classically perform a Hartree-Fock calculation, which would render the proposed technique moot. For sufficiently complex systems, a Hartree-Fock (H.F.) approach will not be feasible, and in general the basis with which the electron integrals (2), (3) are calculated will be arbitrary. In such a case, there would be no reason to assume that high-locality terms should relate to small corrections of the minimizing geometric parameters. The original technique here proposed does not rely on a previous H.F. calculation, as it is applicable to any qubit Hamiltonian. As such, the results presented in fig. 11 should be taken into account when analysing the results produced with the original technique, but do not necessarily reflect the general case.

Figures 9 and 10 are in line with the theoretical results of reference [10], and indicate also that a k-SVD decomposition approach (cf. section 4.1.3) is reasonable in producing a post–Schrieffer-Wolff Hamiltonian that approximates the original Hamiltonian. Note how the degree of approximation can vary with the bond length when taking a k-SVD decomposition approach (fig. 10). Note also how for all decomposition styles the structure is reasonably preserved. This is useful, as we may choose a higher value of ϵ (i.e., a greater asymptotic error on the Schrieffer-Wolff transform; cf. section 1.4, eqs. (9) and (11)) without, in principle, radically altering the minimizing parameters[17].

This explains the results obtained in fig. 12, in particular for HLi with a minimal partitioning approach: note how the values for the "lower bound" curve are actually higher than the Hartree-Fock energies. This is because the lower bound values are a lower bound to the post–Schrieffer-Wolff Hamiltonian, which can be seen to be of higher energy than the Hartree-Fock counterpart (fig. 9). However, we obtain a minimizing geometric parameter of 1.72(4)Å, which reasonably approaches the minimizing bond length obtained with a Hartree-Fock treatment (1.52(4)Å; a $(13 \pm 3)\%$ relative error). Also for HLi, the direct partitioning scheme yields a minimizing parameter of 1.27(4)Å, again not too distant from the Hartree-Fock derived value, with a $(17 \pm 3)\%$ deviation from the H.F. value. The same does not occur for H_2, where taking a lower bound (with both a minimal and direct partitioning scheme) "destroys" the structure of the energy curve, and a minimizing bond length no longer exists[18].

A similar consideration could be made for O_2 — at first sight it seems that taking a lower bound simply destroyed the curve structure — but we should note the energy scales involved: the Hartree-Fock energies are in the order of $-10^2 (E_{\mathrm{h}})$, whereas the energy lower bounds are of the order of $+10^7 (E_{\mathrm{h}})$. This seems to indicate numerical error, rather than a consequence of taking a lower bound to the energies. We relate this to the process by which the used gadgets (cf. section 1.4) operate: as can be seen from eqs. (8) and (10), the subdivision gadgets broadly consist of a two-level gapped system (with gap Δ), with ground state $|0\rangle_u \otimes |\psi\rangle$ and first excited state $|1\rangle_u \otimes |\psi\rangle$ (u denotes the ancillary qubit, and $|\psi\rangle$ describes an arbitrary state of the rest of the system). The intended ground state energy is then achieved by introducing a perturbation potential, such that in second/third perturbation order (for the subdivision/3-to-2-local gadgets, respectively; see the corresponding sections in reference [10]) the ground state energies match. Δ also relates to the bounded error in the post–Schrieffer-Wolff ground state

---

[17]The reason why we should not choose ϵ as arbitrarily small is discussed in the context of the results obtained for O_2.

[18]Strictly speaking, the upper limit of the bond length is the minimizing geometric parameter, but this does not convey physical information (as the limits of the bond length are arbitrarily set by us). Therefore, "no minimizing bond length" will be taken to mean no $\left\{ x \in (l, h) \,\middle|\, \min_x \mathbf{LB}\,[H(x)] \right\}$ exists, with $x \in [l, h]$, and $\mathbf{LB}\,[H]$ a lower bound to the ground state energy of H.

energy; in particular, the error is made to asymptotically scale proportionally to ϵ by taking $\Delta \propto \epsilon^{-2}$, for the subdivision gadget (cf. section 1.4.1), and $\Delta \propto \epsilon^{-3}$, for the 3-to-2 subdivision gadget (cf. section 1.4.2). As such, a choice of small ϵ will lead to extremely large gaps after several Schrieffer-Wolff transformations. Furthermore, the gap scales also proportionally to the largest coupling of the Hamiltonian to be transformed, so that the gap may become extremely large after several Schrieffer-Wolff transformations, even for a relatively large value of ϵ. Once such a situation is achieved, the perturbation terms (which are critical in determining the "true" ground state energy) become numerically irrelevant in comparison to the gap term, so that the resulting ground state energy "lower bound" becomes invalid.

We may observe a similar effect in the k-SVD based decomposition approach to O_2 (fig. 14). Interestingly, however, the energy curve associated to a 1-SVD/minimal partitioning scheme for O_2 seems to still convey information about the 'correct' geometric parameters: note how the lower bound curve has a maximum at less than 0.05Å from the at-rest bond length given by the Hartree-Fock curve. It is unclear why this should be: we posit that it may relate to the value energy of the 'frozen core' (cf. above), or that it may be a consequence of the entanglement structure of O_2 for the ground state energy minimizing geometric parameters, so that by only selecting the 1$^{\text{st}}$ major component of the Schmidt decomposition of the Hamiltonian at each step of decomposition (cf. section 4.1.3), the correct bond length is differentiated.

For H_2 and HLi, applying a k-SVD based decomposition together with a minimal partition scheme apparently yields good results, producing an energy-minimizing bond length for both systems: for H_2 the obtained bond is of 0.70(4)Å (exactly matching the Hartree-Fock–based result), and for HLi the lower bound curve indicates a bond length of 1.64(4)Å, which corresponds to an $(8 \pm 3)\%$ deviation over the Hartree-Fock–based bond length of 1.52(4)Å.

Direct partitioning schemes, on the other hand, yield worse results. When combining them with k-SVD decomposition schemes, no 'at-rest' bond lengths are obtained for H_2. For HLi, we obtain a bond length of 1.72(4)Å, which stands at a $(13 \pm 4)\%$ deviation from the H.F. based bond length. This seems to confirm the intuitive idea that minimizing the number of "cuts" in a partitioning scheme better preserves the structure of the energy curve.

Finally, we see that the ∞-SVD decomposition scheme is simply not effective at preserving the energy curve structure (fig. 13). This may be related to the (∞-SVD related) results presented in fig. 10: as previously discussed, the entanglement structure of the Hamiltonian may vary with bond length, and so the number of corresponding singular values also varies with bond length. This results in a 'broken' energy lower bound curve, where each point effectively relates to a separate energy scale. This prevents identifying useful minima in the curve.

The results obtained by simulating a quantum variational method (cf. section 1.3) show that the quantum variational approach can be very effective, with a relatively small number of layers in the *Ansatz*: for H_2, direct decomposition, and O_2, direct decomposition, the variational results correctly produce the non-variational ones. For O_2, the 5-layer *Ansatz* already produces results comparable to the 10-layer *Ansatz*, and even the 2-layer *Ansatz* reproduces the non-variational curve structure reasonably well.

On the other hand, the variational results do not match the non-variational results for the HLi system; but observing the variational/non-variational results at the respective energy scale, one may see that, interestingly, the variational results not only yield a minimizing parameter, 1.64(4)Å, but this minimizing parameter is closer to the Hartree-Fock–based result of 1.52(4)Å (at a $(8 \pm 5)\%$ deviation) than the non-variational yielded energy-minimizing bond length of 1.72(4)Å ($(13 \pm 3)\%$ deviation from the H.F.–based value).

The same does not happen for the SVD decomposition based results: apart from H_2, for which the variational and non-variational results match exactly, the variational results do not match the non-variational curves, nor do they present any relevant structure, when normalized to

their energy scale. This may be related to the considerations made in section 5, but it is unclear whether it is related to the convergence of the *Ansatz* parameters, or whether a greater number of layers in the *Ansatz* could produce results closer to the non-variational results.

Finally, notice that, again for O_2, even though there is no distinguishable curve structure to the lower bound produced, the variational approach produces an absolute minimum at 1.23(3)Å, exactly matching the 'at-rest' bond length obtained from the Hartree-Fock curve — this strengthens the idea that the entanglement structure of the Hamiltonian at the 'at-rest' bond length for O_2 or the frozen core energy at that bond length separate it from other possible bond lengths.

8 Conclusion

In this work we have presented an original technique for calculating the lowest energy geometric parameters of molecular systems, or, more generally, the parameters of some parameterized Hamiltonian that result in minimal energy. This technique seeks to exploit the natural applications of quantum computers to the treatment of quantum systems, while maintaining minimum quantum computational requirements, in terms of number and quality requirement of the available qubits. We have also presented a completely original implementation in C of a quantum circuit simulator, and a quantum variational algorithm simulator (which relies on the quantum circuit simulator). We have applied the technique to a few test systems, as a proof of concept: H_2, HLi and O_2. We have found that for some approaches to the H_2 and HLi systems, the technique successfully produces 'at-rest' bond lengths comparable to those obtained using a Hartree-Fock treatment, when assuming ideal convergence of the quantum variational algorithm. We have identified possible numerical errors that intrinsically arise from the application of the work of Bravyi et al., which are particularly noticeable for the O_2 system. Finally, we have shown that the quantum variational approach can be comparable to classical numerical determination of the lowest eigenvalue of an operator, but that the inability to encode particle number constraints can produce eigenvalues lower than those obtained using a non-variational approach. However, these lower values sometimes still produce 'at-rest' bond lengths, comparable to those obtained with a Hartree-Fock approach.

The results presented in this thesis should be viewed as a first step of a long term project aiming at using a quantum computer to obtain geometric parameters of molecules which cannot be simulated in a classic computer.

Appendix A: Quantum Phase Estimation

Consider some unitary operation U, with eigenvectors $\{|u\rangle\}$, and corresponding eigenvalues $\{\lambda_u\}$. The quantum circuit for the Phase Estimation Algorithm (PEA) allows one to estimate the eigenvalues $\{\lambda_u\}$, by estimating φ_u such that $\lambda_u = e^{2\pi i \varphi_u}$. The PEA circuit, as applied to an operator U, is given in fig. 1 (reproduced below).
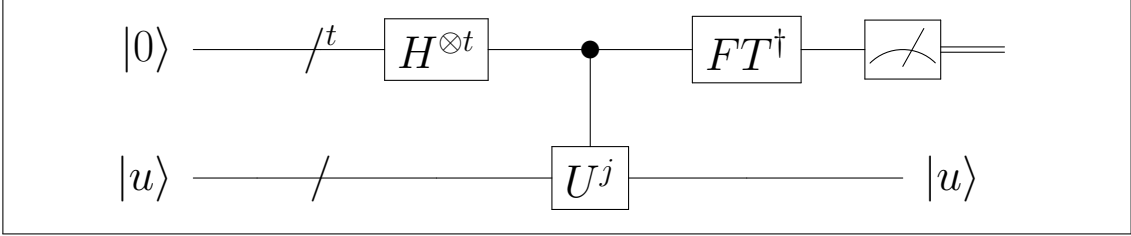

Figure 1: Circuit schematic reflecting the Quantum Phase Estimation algorithm for an operator U. The first register is a bundle of t wires, and the bottom register has as many wires as dictated by U. H denotes the Hadamard gate. U^j denotes 2^j applications of U controlled by the $(t-j)^{\text{th}}$ wire in the first register (as indexed top-to-bottom), with an implicit run of $j = 0 \ldots t-1$. $FT^\dagger$ denotes the inverse Quantum Fourier Transform. $|u\rangle$ is generally taken to be an eigenvector of U but need not be.

The discussion of the effects of this circuit is best split into three progressively general conditions on $|u\rangle$.

A.1 $|u\rangle$ as an Eigenvector With a t-bit φ_u

If $|u\rangle$ is an eigenvector of U and the corresponding eigenvalue is $\lambda_u = e^{2\pi i \varphi_u}$ such that φ_u can be expressed exactly in t bits, then the circuit
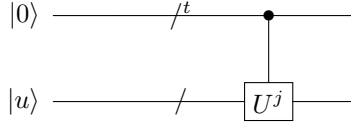
produces the state

$$\bigotimes_{j=1}^{t} \left(\frac{1}{\sqrt{2}} (|0\rangle + e^{2\pi i 2^{t-j} \varphi_u} |1\rangle) \right) \otimes |u\rangle$$

and considering that the Quantum Fourier Transform (QFT) state mapping is of the form

$$|p\rangle \mapsto \frac{1}{\sqrt{2^n}} \bigotimes_{k=1}^{n} \left(|0\rangle + e^{2\pi i p/2^k} |1\rangle \right)$$

such that the inverse operation would result in a mapping of

$$\frac{1}{\sqrt{2^t}} \bigotimes_{k=1}^{t} \left(|0\rangle + e^{2\pi i p 2^{t-k}} |1\rangle \right) \mapsto |p\rangle$$

then the inverse Quantum Fourier Transform (IQFT) in the first t wires subspace recovers the bits of φ_u exactly[19], in the form of the final state $|\varphi_u\rangle |u\rangle$.

A.2 $|u\rangle$ as any Eigenvector

In general the eigenvalue λ_u won't be exactly expressible in t bits, in which case the last statement about the effect of the IQFT does not hold, and rather we have[20]

$$
\bigotimes_{j=1}^{t} \left(\frac{1}{\sqrt{2}} (|0\rangle + e^{2\pi i 2^{t-j} \varphi_u} |1\rangle) \right) \otimes |u\rangle =
$$

$$
= \frac{1}{\sqrt{2^t}} \sum_{j=0}^{2^t-1} e^{2\pi i \varphi_u j} |j\rangle |u\rangle \xmapsto{\text{QFT}^\dagger} \frac{1}{2^t} \sum_{k,l=0}^{2^t-1} e^{-2\pi i k l / 2^t} e^{2\pi i \varphi_u k} |l\rangle
$$

With this knowledge of the amplitudes, and noticing that a t-qubit circuit will be able to provide at best an approximation to the tth binary digit of φ, we can[21] calculate a bound for the probability of obtaining a measurement that deviates from the smallest t-bit approximation by at least e:

$$
p(\text{“deviation”}) \leq \frac{1}{2(e-1)}
$$

and so conversely a bound for the probability of observing a measurement of the smallest t-bit approximation within some deviation e. We then demand that this deviation is at most in the nth bit, i.e., give margin for error "up to" that bit. This is done by setting $e = 2^{t-n} - 1$. Plugging this value into the "no deviation" probability, we obtain

$$
p(\text{“no deviation”}) \leq 1 - \frac{1}{2(2^p - 2)} \qquad \text{where} \qquad p \overset{\text{def}}{=} t - n
$$

So a probability of no deviation greater or equal to $1 - \epsilon$ requires

$$
2^p \geq \frac{1}{2\epsilon} + 2
$$

satisfied by

$$
p = \left\lceil \log_2 \left(\frac{1}{2\epsilon} + 2 \right) \right\rceil
$$

meaning a total number of qubits

$$
t = n + \left\lceil \log_2 \left(\frac{1}{2\epsilon} + 2 \right) \right\rceil
$$

Inverting the logical relationship, we have that a circuit with this many qubits in its state register will result in a readout of an n-bit approximation of φ with probability of at least $1 - \epsilon$.

---

[19]with reversed bit ordering; notice that if $p = p_0 2^0 + \cdots + p_n 2^{n-1}$ then $e^{2\pi 2^{-k} p} = e^{2\pi (p_0 2^{-k} + \cdots + p_{k-1} 2^{-1}) \times e^{2\pi n}}$, resulting in $e^{2\pi i 0 . p_0 \cdots p_{k-1}} \xmapsto{\text{QFT}^\dagger} p' = p_{k-1} 2^0 + \cdots + p_0 2^{k-1}$

[20]To obtain the IQFT mapping observe that $(\text{QFT})_{uv}$ is a unitary matrix, and as such $\text{QFT}^\dagger_{uw} = \text{QFT}^\dagger_{uw} = (\text{QFT})^*_{wu}$. Making use of a completeness relation, $\text{QFT}^\dagger |u\rangle = \sum_{w=0}^{2^n-1} |w\rangle \langle w| \text{QFT}^\dagger |u\rangle = \frac{1}{\sqrt{2^n}} \sum_{w=0}^{2^n-1} e^{-2\pi i u w / 2^n} |w\rangle$
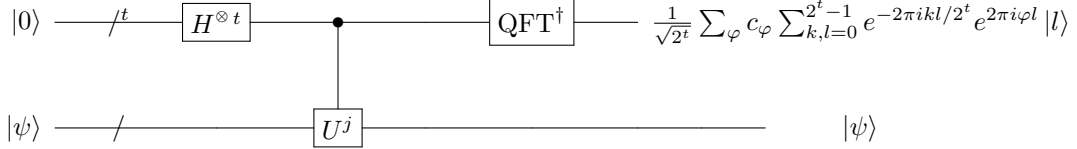
[21]See [35, pp. 223, 224] for a more complete picture.

A.3 $|u\rangle$ as a Random State

Consider now that the second register is instead initialized as some vector $|\psi\rangle$. We can expand this state in the U operator's eigenvectors;

$$|\psi\rangle = \sum_\varphi c_\varphi |\varphi\rangle$$

from where the previous discussion can be carried for each of the terms, as weighed by c_φ:

Therefore, like in appendix A.2, performing a measurement of the first register yields a t-bit approximation of the eigenvalue, except the measured value is approximately φ_i with probability $|c_{\varphi_i}|^2$ for some eigenvalue $\varphi_i \in \{\varphi_1, \ldots, \varphi_n\}$.

A.4 Efficiency and Considerations

QPA is not a full quantum algorithm, but rather a quantum subroutine. This is because it makes use of a quantum oracle, namely the "black box" implementation of the U^j gate, and presupposes the encoding of the candidate state into the circuit. However, a few considerations about the efficiency of this routine (as opposed to the classical counterpart) can be made.

First, notice that an n-qubit (quantum) computer supports operators with a matrix representation of size 2^n. This in itself provides an exponential advantage, which is intrinsic to quantum computing. On the other hand, the QFT has a bound of $\Theta(n^2)$ in the number of operations required, and relaxing the requirements to polynomial precision we can obtain a polynomial bound for a polynomial accuracy [35, pp. 220, 221]. Therefore, if the oracle operation U^j can be implemented in $\text{Poly}(n)$, the whole routine is polynomially efficient, while providing an exponential advantage in memory over the classical counterpart. Likewise, efficient encoding of the input state does not affect the efficiency of the QPE routine.

However, in practice, such an algorithm requires both full coherence of the qubits throughout the protocol, and possibly a large number of qubits (especially in light of the discussion of appendix A.2). The existence of errors is also not considered, for which error correction schemes would be necessary, further increasing the qubit count requirement. These requirements are currently intractable [42, 33] in the Noise Intermediate Scale Quantum (NISQ) era, in which a moderate amount of qubits, though noisy, are available.

Appendix B: "Tainted Numbers"

Presented by Gidney in [23], "tainted numbers" are a way to algorithmically derive the matrix associated to a controlled operation, when the matrix form of the operation is known. For this, the number μ is introduced, such that ($\otimes$ denotes the Kronecker product operation)

$$\mu \otimes U = \mu \cdot I_n$$
$$U \otimes \mu = I_n \cdot \mu$$

It is from the "absorbing" nature of μ that the name 'tainted' is given. In [23], no formalization of μ's algebra is given, with an analysis of practical consequences of its use (and comparison with expected results) done instead.

An attempt to formalize μ into an algebra with arbitrary-size matrices quickly shows that a group algebra is not possible:

$$G = \{\mu^{-1}, \mu\} \cup \mathbb{C}^m$$
$$\mu^{-1} \cdot \mu = 1$$

for any $m \times m$ matrix A,

$$\mu^{-1} \cdot \mu \cdot A = \begin{cases} (\mu^{-1} \cdot \mu) \cdot A & = 1 \cdot A \\ \mu^{-1} \cdot (\mu \cdot A) & = \mu^{-1} \cdot I_m \end{cases}$$

As such μ should be treated with some caution; Gidney himself reports some cases where the use of μ produces the wrong matrix [22]. We thus avoid this characterization and simply show that for an "unobstructed" controlled gate (i.e., no other gates acting on the qubits between the gate and the control), an informal use of μ produces the expected result.

Consider a controlled single-qubit gate Q with corresponding tensor form $Q^\alpha{}_\beta$. It may be controlled by the qubit $m + 1$ positions above or below it. In an unobstructed case then, the controlled operation $\tilde{Q}$ is respectively given by either

$$\tilde{Q} = |0\rangle\langle 0| \otimes I^{\otimes m} \otimes I + |1\rangle\langle 1| \otimes I^{\otimes m} \otimes Q$$

or

$$\tilde{Q} = I \otimes I^{\otimes m} \otimes |0\rangle\langle 0| + Q \otimes I^{\otimes m} \otimes |1\rangle\langle 1|$$

(where I denotes the identity operation)

If we consider then the 'control matrix' C with tensor representation $C^\chi{}_\nu = \delta^0{}_\nu \mu + \delta^1{}_\nu$,

$$\begin{aligned} = C^\mu{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi &= (\delta^0{}_\nu \mu + \delta^1{}_\nu) \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi \\ &= \delta^0{}_\nu \mu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi + \delta^1{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi \end{aligned}$$

$\delta^\alpha{}_\beta = I$ and from $\mu \otimes A = \mu \otimes I$ it is intuitive that $\mu \delta^\alpha{}_\beta = \delta^\alpha{}_\beta \mu$, so that

$$\delta^0{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots \mu U^\epsilon{}_\pi + \delta^1{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi = \delta^0{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots \delta^\epsilon{}_\pi + \delta^1{}_\nu \delta^\alpha{}_\beta \delta^\gamma{}_\lambda \cdots U^\epsilon{}_\pi$$

$$= |0\rangle\langle0| \otimes I^m \otimes \mu \cdot I + |1\rangle\langle1| \otimes I^m \otimes U$$

which matches the expected expression for a controlled gate, up to a factor of μ (which disappears when the matrix is "cleaned"). The procedure for a control m wires below the gate is completely analogous.

References

[1] Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. `https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf`. Not published, but introduced in a "Coursera" course for the first time by Hinton et al. Accessed 2019-11-16.

[2] S. Andradottir. A review of simulation optimization techniques. In *1998 Winter Simulation Conference. Proceedings (Cat. No.98CH36274)*. IEEE.

[3] T. A. Arias, M. C. Payne, and J. D. Joannopoulos. Ab initiomolecular-dynamics techniques extended to large-length-scale systems. *Physical Review B*, 45(4):1538–1549, Jan. 1992.

[4] Y. B. Band and Y. Avishai. Many-body theory. In *Quantum Mechanics with Applications to Nanotechnology and Information Science*, pages 825–870. Elsevier, 2013.

[5] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter. Elementary gates for quantum computation. *Physical Review A*, 52(5):3457–3467, Nov. 1995.

[6] C. D. Batista and G. Ortiz. Generalized jordan-wigner transformations. *Physical Review Letters*, 86(6):1082–1085, Feb. 2001.

[7] C.-E. Bichot and P. Siarry, editors. *Graph Partitioning*. John Wiley & Sons, Inc., Feb. 2013.

[8] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[9] C. Bravo-Prieto, R. LaRose, M. Cerezo, Y. Subasi, L. Cincio, and P. J. Coles. Variational quantum linear solver: A hybrid algorithm for linear systems, 2019.

[10] S. Bravyi, D. P. DiVincenzo, D. Loss, and B. M. Terhal. Quantum simulation of many-body hamiltonians using perturbation theory with bounded-strength interactions. *Physical Review Letters*, 101(7), Aug. 2008.

[11] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal of Scientific Computing*, 16:1190–1208, Sept. 1995.

[12] P. J. Chase. Algorithm 382: combinations of m out of n objects [g6]. *Communications of the ACM*, 13(6):368, June 1970.

[13] F. Chollet et al. Keras. `https://keras.io`, 2015.

[14] T. O. Developers. Openfermion-psi4. `https://github.com/quantumlib/OpenFermion-Psi4`, 2017.

[15] D. P. DiVincenzo. Two-bit gates are universal for quantum computation. *Physical Review A*, 51(2):1015–1022, Feb. 1995.

[16] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, Sept. 1936.

[17] A. E. Feldmann and L. Foschini. Balanced partitions of trees and applications. *Algorithmica*, 71(2):354–376, July 2013.

[18] T. P. S. Foundation. The python language reference, 2019. Accessed on 2020-04-21.

[19] B. T. Gard, L. Zhu, G. S. Barron, N. J. Mayhall, S. E. Economou, and E. Barnes. Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm. *npj Quantum Information*, 6(1), Jan. 2020.

[20] M. R. Garey and D. S. Johnson. *Computers and intractability*, volume 174. freeman San Francisco, 1979.

[21] J. P. B. George P. Wadsworth. *Introduction to Probability and Random Variables*. McGraw-Hill Book Company, Inc., 1960.

[22] C. Gidney. private communication.

[23] C. Gidney. Impractical Experiments #3: Treating Controls Like Values. `http://web.archive.org/web/20190109155033/https://algassert.com/impractical-experiments/2015/05/17/Treating-Controls-like-Values.html`, 2015. Accessed: 2019-01-09 at 15:50:33.

[24] R. Graham. *Concrete Mathematics: A Foundation for Computer Science.* Pearson Education, 1994.

[25] B. S. Institution. *The C standard: incorporating Technical Corrigendum 1 : BS ISO/IEC 9899/1999.* John Wiley, 2003.

[26] E. Jeřábek. Lower bound for sum of binomial coefficients? MathOverflow. `https://mathoverflow.net/q/55617` (version: 2011-02-16).

[27] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, Sept. 2017.

[28] A. Katwala. Inside big tech's high-stakes race for quantum supremacy. `https://www.wired.co.uk/article/quantum-supremacy-google-microsoft-ibm`, 2020. Retrieved 2020-06-28.

[29] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2014.

[30] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe. Experimental comparison of two quantum computing architectures. *Proceedings of the National Academy of Sciences*, 114(13):3305–3310, Mar. 2017.

[31] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, Feb. 2016.

[32] J. R. McClean, K. J. Sung, I. D. Kivlichan, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, P. Gokhale, T. Häner, T. Hardikar, V. Havlíček, O. Higgott, C. Huang, J. Izaac, Z. Jiang, X. Liu, S. McArdle, M. Neeley, T. O'Brien, B. O'Gorman, I. Ozfidan, M. D. Radin, J. Romero, N. Rubin, N. P. D. Sawaya, K. Setia, S. Sim, D. S. Steiger, M. Steudtner, Q. Sun, W. Sun, D. Wang, F. Zhang, and R. Babbush. Openfermion: The electronic structure package for quantum computers, 2017.

[33] H. Mohammadbagherpoor, Y.-H. Oh, P. Dreher, A. Singh, X. Yu, and A. J. Rindos. An improved implementation approach for quantum phase estimation on quantum computers, 2019.

[34] H. Mohammadbagherpoor, Y.-H. Oh, A. Singh, X. Yu, and A. J. Rindos. Experimental challenges of implementing quantum phase estimation algorithms on ibm quantum computer, 2019.

[35] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2009.

[36] T. E. Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[37] R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. D. Remigio, R. M. Richard, J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, M. Saitow, X. Wang, B. P. Pritchard, P. Verma, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. Evangelista, J. M. Turney, T. D. Crawford, and C. D. Sherrill. Psi4 1.1: An open-source electronic structure program emphasizing automation, advanced libraries, and interoperability. *Journal of Chemical Theory and Computation*, 13(7):3185–3197, June 2017.

[38] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre. Data re-uploading

for a universal quantum classifier. *Quantum*, 4:226, Feb. 2020.

[39] J. Perla. Notes on AdaGrad. https://static.miraheze.org/yoavfreundwiki/6/6a/Adagrad.pdf, 2014. Accessed: 2019-11-20.

[40] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1), July 2014.

[41] J. Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, Aug 2018.

[42] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer. Defining and detecting quantum speedup. 2014.

[43] H. H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, Mar. 1960.

[44] A. Spörl, T. Schulte-Herbrüggen, S. J. Glaser, V. Bergholm, M. J. Storcz, J. Ferber, and F. K. Wilhelm. Optimal control of coupled josephson qubits. *Physical Review A*, 75(1), Jan. 2007.

[45] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Review*, 35(4):551–566, 1993.

[46] T. Tilma and E. C. G. Sudarshan. Generalized euler angle parametrization forSU(n). *Journal of Physics A: Mathematical and General*, 35(48):10467–10501, Nov. 2002.

[47] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H.

Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[48] S. Wallis. Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics*, 20(3):178–208, July 2013.

[49] T. Worsch. Lower and upper bounds for (sums of) binomial coefficients. 1994.

[50] J.-F. Yang and C.-L. Lu. Combined techniques of singular value decomposition and vector quantization for image coding. *IEEE Transactions on Image Processing*, 4(8):1141–1146, 1995.

[51] M. D. Zeiler. Adadelta: An adaptive learning rate method, 2012. Unpublished; developed at Google Inc. USA and New York University, USA.